

# Charting the endocrine cell system in health and diabetes using single-cell transcriptomics

Sophie A. Tritschler, M.Sc.

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität  
München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Markus List

Prüfende der Dissertation:

1. Prof. Dr. Dr. Fabian J. Theis
2. Prof. Dr. Heiko Lickert

Die Dissertation wurde am 26.04.2024 bei der Technischen Universität München eingereicht  
und durch die TUM School of Life Sciences am 01.08.2024 angenommen.

# Abstract

Diabetes mellitus is a complex, chronic metabolic disease with high unmet medical need. In diabetes, hormonal regulation of blood glucose is impaired because the endocrine cell system in the pancreas and intestine becomes progressively dysfunctional. Developing regenerative treatments that restore functioning endocrine cells to regain an endogenous endocrine control is therefore an attractive approach to halt disease progression. To be successful in any such endeavors it is necessary to first better understand the molecular determinants of cell identity, normal cell function and dysfunction in disease as well as how the biology of preclinical models translates to humans. Since endocrine cells are functionally heterogeneous and plastic, this also requires delineating the spectrum of states endocrine cells acquire. Through recent technological and computational advances, single-cell transcriptomics has become a powerful tool with the necessary depth and resolution to study cellular trajectories, composition, identities and states in such complex, multicellular systems. In this cumulative thesis I use single-cell transcriptomics to chart specific aspects of the endocrine cellular landscape of the pancreas and intestine, with the goal of aiding our understanding of the cells and molecular processes that form the endocrine system during critical windows of life and disease and could be targeted by therapeutics to regenerate a normal functioning endocrine system.

I first reviewed the current literature and discussed the impact of single-cell transcriptomics and computational approaches on learning cellular trajectories and studying heterogeneity in the pancreatic endocrine cell system. I then analyzed single-cell transcriptomic data zooming into different parts of the developing, healthy and diabetic endocrine pancreas and small intestine. First, I established a comprehensive reference cell map of the healthy human, pig and mouse endocrine pancreas as a framework to systematically study physiological cell states and to link findings between humans and two clinically relevant animal models. I then reconstructed a lineage model of endocrine differentiation in the developing mouse pancreas, which revealed markers that distinguish differentiation states and predicted regulators that shape cell identity. To study how these change in disease, I next created a map of cells and their relations to delineate how hyperglycemia and treatment affects the endocrine pancreas in diabetic mice. This suggested that re-differentiation of dysfunctional pancreatic  $\beta$ -cells is a viable regenerative treatment option for diabetes and revealed targetable molecular pathways to do so. Finally, I characterized the cellular repertoire of the small intestinal epithelium to describe how an obesogenic diet and pre-diabetes affects lineage allocation and endocrine composition of the gut. This indicated mechanisms of intestinal maladaptation and enteroendocrine dysfunction, which may contribute to drive progression of diabetes.

Through nuanced descriptions of specific endocrine cell populations this thesis generated novel insights into inherent endocrine biology and cellular drivers of diabetes in the pancreas and intestine: it described cellular and molecular mechanisms of endocrine cell identity and (dys-) function; identified biomarkers of (patho-)physiological cell states; predicted potentially targetable molecular pathways; and created a framework to assess the translational value of two preclinical models. The novel findings together with the accessible cellular maps represent a rich resource to empower diverse efforts towards regenerative therapies for diabetes.

# Zusammenfassung

Diabetes mellitus ist eine komplexe, chronische Stoffwechselerkrankung mit einem hohen ungedeckten medizinischen Bedarf. Bei Diabetes ist die hormonelle Regulierung des Blutzuckerspiegels gestört, weil das endokrine Zellsystem im Pankreas und im Darm zunehmend dysfunktional wird. Die Entwicklung regenerativer Behandlungen, die funktionierende endokrine Zellen wiederherstellen, um eine endogene endokrine Kontrolle wiederzuerlangen, ist daher ein attraktiver Ansatz, um das Fortschreiten der Krankheit aufzuhalten. Um bei solchen Bemühungen erfolgreich zu sein, müssen zunächst die molekularen Determinanten der Zellidentität, der normalen Zellfunktion und der Funktionsstörung bei Krankheiten besser verstanden werden. Zudem muss geklärt werden, wie sich die Biologie präklinischer Modelle auf den Menschen übertragen lässt. Da endokrine Zellen funktionell heterogen und plastisch sind, muss auch das Spektrum der Zustände, das endokrine Zellen annehmen, beschrieben werden. Durch die jüngsten technologischen und computergestützten Fortschritte ist die Einzelzelltranskriptomik zu einem leistungsfähigen Instrument geworden, das über die notwendige Tiefe und Auflösungsschärfe verfügt, um die zellulären Trajektorien, die Zusammensetzung, die Identitäten und die Zustände in solch komplexen, multizellulären Systemen zu untersuchen. In dieser kumulativen Dissertation erfasse ich mit der Einzelzelltranskriptomik spezifische Aspekte der endokrinen Zelllandschaft des Pankreas und des Darms mit dem Ziel unser Verständnis der Zellen und molekularen Prozesse, die das endokrine System während kritischer Lebens- und Krankheitsperioden bilden, zu verbessern und auf die Therapeutika abzielen könnten, um ein normal funktionierendes endokrines System zu regenerieren.

Zunächst gebe ich einen Überblick über die aktuelle Literatur und diskutiere die Bedeutung der Einzelzell-Transkriptomik und computergestützter Ansätze zur Erforschung der zellulären Trajektorien und der Heterogenität im endokrinen Zellsystem des Pankreas. Ich analysiere dann transkriptomische Einzelzelldaten bestimmter Bereiche des sich entwickelnden, gesunden und diabetischen endokrinen Pankreas und Dünndarms. Mit einer umfassenden Referenzzellkarte des gesunden endokrinen Pankreas von Mensch, Schwein und Maus, die ich erstellt habe, untersuche ich systematisch die physiologischer Zellzustände und verknüpfe die Ergebnisse von Patienten mit zwei klinisch relevanten Tiermodellen. Anschließend rekonstruiere ich ein Abstammungsmodell der endokrinen Differenzierung in dem sich entwickelnden Pankreas der Maus, das Marker zur Unterscheidung von Differenzierungszuständen aufzeigt und Regulatoren vorhersagt, die die Zellidentität prägen. Um zu untersuchen, wie sich diese im Krankheitsfall verändern, erstelle ich folgend eine Karte der Zellen und ihrer Abstammungen, die die Auswirkung der Hyperglykämie und Behandlung

auf den endokrinen Pankreas von diabetischen Mäusen beschreibt. Dies deutet darauf hin, dass die Redifferenzierung von dysfunktionalen  $\beta$ -Zellen des Pankreas eine praktikable regenerative Behandlungsoption für Diabetes darstellt, und es zeigt zudem behandelbare molekulare Prozesse auf. Schließlich beschreibe ich das zelluläre Repertoire des Dünndarmepithels, um den Einfluss einer adipogenen Ernährung und Prädiabetes auf die Abstammungszuordnung und die endokrine Zusammensetzung des Darms herauszustellen. Es zeigt sich, dass die Mechanismen der Fehlanpassung des Darms und der enteroendokrinen Dysfunktion zum Fortschreiten von Diabetes beitragen können.

Durch differenzierte Beschreibungen bestimmter endokriner Zellpopulationen liefert diese Arbeit neue Erkenntnisse über die inhärente endokrine Biologie und die zellulären Triebkräfte von Diabetes in Pankreas und Darm: Sie beschreibt zelluläre und molekulare Mechanismen der endokrinen Zellidentität und (Dys-)Funktion, identifiziert Biomarker für (patho-)physiologische Zellzustände, sagt potenziell behandelbare molekulare Signalwege voraus und schafft einen Rahmen zur Bewertung des translationalen Werts von zwei präklinischen Modellen. Die neuen Erkenntnisse sind zusammen mit den verfügbaren zellulären Karten eine reichhaltige Ressource, um die diversen Anstrengungen in Richtung regenerativer Diabetes-Therapien zu unterstützen.

# Acknowledgement

First and foremost, I thank Fabian Theis and Heiko Lickert for creating a fantastic, thriving and inspiring research environment. Thank you for your trust, support and mentorship, for the many exciting projects I got to work on, and for pushing and challenging me throughout my PhD with new opportunities to grow and learn.

The ICB and IDR has been a fun place to work. The many scientific discussions, collaborations, retreats, lunch runs and Wiesn' visits together with the amazing lab members were a blast. Thank you Anna, Marianne, Elisabeth and Sabine for all your help and running the show in the back; Stephan, Aimée, Lexi, Maren, Moritz, David, Leander, Philipp, Volker and Marius for the great collaborations; Mostafa and Anika for your leadership in many projects; Maren for your warm welcome, your guidance and being a great office mate for four years; Malte for always having an open door and discussing science late into the night; Alex for teaching me efficiency and craziness; Dominik and Johanna for keeping me in shape; Olle and Mo for the memorable kick-off meeting; and Niklas for the fun and chats.

I thank my family and the Suola WG for cheering along the way.

And finally, thank you Simon for all you are. I am lucky that the coin was tails.

## List of contributed articles

This thesis is a publication-based thesis and is built on the following core publications as a main author. I, Sophie Tritschler, am one of the main authors of the below listed manuscripts<sup>1-6</sup>. Summaries of these articles and descriptions of my contributions can be found in *chapter 3*.

### Publications in the context of my doctoral thesis as main author

(i) **Sophie Tritschler**, Moritz Thomas, Anika Böttcher, Barbara Ludwig, Janine Schmid, Undine Schubert, Elisabeth Kemter, Eckhard Wolf, Heiko Lickert, Fabian J. Theis, **A transcriptional cross species map of pancreatic islet cells**, *Molecular Metabolism*, Volume 66, 101595 (2022)

(See also publication<sup>1</sup> in the bibliography and in Appendix A)

(ii) Aimée Bastidas-Ponce\*, **Sophie Tritschler\***, Leander Dony, Katharina Scheibner, Marta Tarquis-Medina, Ciro Salinno, Silvia Schirge, Ingo Burtscher, Anika Böttcher, Fabian J. Theis, Heiko Lickert, Mostafa Bakhti; **Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis**, *Development* 15; 146 (12): dev173849 (2019)

(See also publication<sup>2</sup> in the bibliography)

(iii) Stephan Sachs\*, Aimée Bastidas-Ponce\*, **Sophie Tritschler\***, Mostafa Bakhti, Anika Böttcher, Miguel A Sánchez-Garrido, Marta Tarquis-Medina, Maximilian Kleinert, Katrin Fischer, Sigrid Jall, Alexandra Harger, Erik Bader, Sara Roscioni, Siegfried Ussar, Annette Feuchtinger, Burcak Yesildag, Aparna Neelakandhan, Christine B Jensen, Marion Cornu, Bin Yang, Brian Finan, Richard D DiMarchi, Matthias H Tschöp, Fabian J Theis, Susanna M Hofmann, Timo D Müller, Heiko Lickert, **Targeted pharmacological therapy restores  $\beta$ -cell function for diabetes remission**, *Nature Metabolism*, 2, 192–209 (2020)

(See also publication<sup>3</sup> in the bibliography and in Appendix B)

(iv) Alexandra Aliluev\*, **Sophie Tritschler\***, Michael Sterr, Lena Oppenländer, Julia Hinterdobler, Tobias Greisle, Martin Irmeler, Johannes Beckers, Na Sun, Axel Walch, Kerstin Stemmer, Alida Kindt, Jan Krumsiek, Matthias H Tschöp, Malte D Luecken, Fabian J Theis, Heiko Lickert, Anika Böttcher, **Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice**, *Nature Metabolism*, 3, 1202–1216 (2021)

(See also publication<sup>4</sup> in the bibliography and in Appendix C)

(v) **Sophie Tritschler**, Fabian J. Theis, Heiko Lickert, Anika Böttcher, **Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas**, *Molecular Metabolism*, Volume 6, Issue 9, Pages 974-990 (2017)

(See also publication<sup>5</sup> in the bibliography and in Appendix D)

(vi) **Sophie Tritschler\***, Maren Büttner\*, David S. Fischer, Marius Lange, Volker Bergen, Heiko Lickert, Fabian J. Theis, **Concepts and limitations for learning developmental trajectories from single cell genomics**, *Development* 15; 146 (12): dev170506 (2019)

(See also publication<sup>6</sup> in the bibliography)

\* *these authors contributed equally.*

## Further publications not included in this thesis

I also contributed to the following publications during my PhD not included in this thesis.

(vii) Anika Böttcher, Maren Büttner, **Sophie Tritschler**, Michael Sterr, Alexandra Aliluev, Lena Oppenländer, Ingo Burtscher, Steffen Sass, Martin Irmeler, Johannes Beckers, Christoph Ziegenhain, Wolfgang Enard, Andrea C Schamberger, Fien M Verhamme, Oliver Eickelberg, Fabian J Theis, Heiko Lickert, **Non-canonical Wnt/PCP signalling regulates intestinal stem cell lineage priming towards enteroendocrine and Paneth cell fates**, *Nature Cell Biology* 23, 23–31 (2021)

(See also publication <sup>7</sup> in the bibliography)

(viii) Philipp Angerer, Lukas Simon, **Sophie Tritschler**, F. Alexander Wolf, David Fischer, Fabian J. Theis, **Single cells make big data: New challenges and opportunities in transcriptomics**, *Current Opinion in Systems Biology*, Volume 4, Pages 85-91 (2017)

(See also publication <sup>8</sup> in the bibliography)

(ix) Mostafa Bakhti\*, Katharina Scheibner\*, **Sophie Tritschler**, Aimée Bastidas-Ponce, Marta Tarquis-Medina, Fabian J Theis, Heiko Lickert, **Establishment of a high-resolution 3D modeling system for studying pancreatic epithelial cell biology in vitro**, *Molecular Metabolism*, Volume 30, Pages 16-29 (2019)

(See also publication <sup>9</sup> in the bibliography)

(x) David S Fischer\*, Leander Dony\*, Martin König, Abdul Moeed, Luke Zappia, Lukas Heumos, **Sophie Tritschler**, Olle Holmberg, Hananeh Aliee, Fabian J Theis, **Sfaira accelerates data and model reuse in single cell genomics**, *Genome Biology* 22, 248 (2021)

(See also publication <sup>10</sup> in the bibliography)

(xi) Ciro Salinno, Maren Büttner, Perla Cota, **Sophie Tritschler**, Marta Tarquis-Medina, Aimée Bastidas-Ponce, Katharina Scheibner, Ingo Burtscher, Anika Böttcher, Fabian J Theis, Mostafa Bakhti, Heiko Lickert, **CD81 marks immature and dedifferentiated pancreatic  $\beta$ -cells**, *Molecular Metabolism*, Volume 49, 101188 (2021)

(See also publication <sup>11</sup> in the bibliography)

(xii) Mostafa Bakhti\*, Aimée Bastidas-Ponce\*, **Sophie Tritschler**, Oliver Czarnecki, Marta Tarquis-Medina, Eva Nedvedova, Jessica Jaki, Stefanie J Willmann, Katharina Scheibner, Perla Cota, Ciro Salinno, Karsten Boldt, Nicola Horn, Marius Ueffing, Ingo Burtscher, Fabian J Theis, Ünal Coskun, Heiko Lickert, **Synaptotagmin-13 orchestrates pancreatic endocrine cell egression and islet morphogenesis**, *Nature Communications* 13, 4540 (2022).

(See also publication <sup>12</sup> in the bibliography)

(xiii) Corinne Benakis, Alba Simats, **Sophie Tritschler**, Steffanie Heindl, Simon Besson-Girard, Gemma Llovera, Kelsey Pinkham, Anna Kolz, Alessio Ricci, Fabian J Theis, Stefan Bittner, Özgün Gökce, Anneli Peters, Arthur Liesz, **T cells modulate the microglial response to brain ischemia**, *eLife* 11:e82031 (2022)

(See also publication <sup>13</sup> in the bibliography)

\* these authors contributed equally



# Content

<b>1.</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>1.1.</b>	<b>Restoring endocrine function to treat diabetes</b>	<b>1</b>
1.1.1.	Restoring functional mass of pancreatic insulin-producing $\beta$ -cells	2
1.1.2.	Restoring homeostasis in the endocrine intestine	4
<b>1.2.</b>	<b>Modeling the endocrine system at single-cell resolution</b>	<b>6</b>
1.2.1.	Endocrine heterogeneity and plasticity	6
1.2.2.	Elucidating cellular composition, identity and states in tissues	6
1.2.3.	Computational methods empower single-cell genomics	8
1.2.4.	Characterizing cell populations of the endocrine pancreas	8
1.2.5.	Charting the endocrine system of the intestinal epithelium	9
1.2.6.	Animal models to study endocrine systems	10
<b>1.3.</b>	<b>Goals of this thesis</b>	<b>10</b>
<b>1.4.</b>	<b>Summary of results</b>	<b>12</b>
<b>2.</b>	<b>METHODS</b>	<b>14</b>
<b>2.1.</b>	<b>Single-cell RNA sequencing technology</b>	<b>15</b>
<b>2.2.</b>	<b>Preprocessing of single-cell RNA sequencing data</b>	<b>16</b>
2.2.1.	Read alignment, count matrix generation and quality control	17
2.2.2.	Normalization and batch correction	17
2.2.3.	Feature selection, dimensionality reduction and visualization	18
<b>2.3.</b>	<b>Analysis to identify cellular structure</b>	<b>19</b>
2.3.1.	Cell clustering	19
2.3.2.	Cell annotation approaches	20
2.3.3.	Identification of doublet clusters	20
2.3.4.	Reference mapping for cross-species comparison	21
2.3.5.	Trajectory inference and gene dynamics in continuous data	22
2.3.6.	Inference of cellular dynamics	23
<b>2.4.</b>	<b>Analysis to reveal mechanisms</b>	<b>23</b>
2.4.1.	Differential expression analysis	24
2.4.2.	Prior knowledge contextualization through gene set enrichment	24
2.4.3.	Population composition analysis	24
<b>3.</b>	<b>SUMMARY OF CONTRIBUTED ARTICLES</b>	<b>26</b>
<b>3.1.</b>	<b>Research Articles</b>	<b>26</b>
3.1.1.	A transcriptional cross species map of pancreatic islet cells	26

3.1.2. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis .....	28
3.1.3. Targeted pharmacological therapy restores $\beta$ -cell function for diabetes remission.....	30
3.1.4. Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice .....	32
<b>3.2. Review Articles .....</b>	<b>34</b>
3.2.1. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas .....	34
3.2.2. Concepts and limitations for learning developmental trajectories from single cell genomics .....	36
<b>4. DISCUSSION AND OUTLOOK.....</b>	<b>37</b>
4.1. Summary.....	37
4.2. Advancing our understanding of endocrine disease biology .....	38
4.2.1. Towards multimodal models of cellular function and heterogeneity .....	38
4.2.2. From genes to mechanisms.....	41
4.3. Advancing disease relevance with large-scale atlases.....	41
4.4. Towards unified foundational models of biology .....	42
<b>BIBLIOGRAPHY .....</b>	<b>44</b>
<b>APPENDIX.....</b>	<b>61</b>
Appendix A.....	62
Appendix B.....	97
Appendix C.....	131
Appendix D.....	167

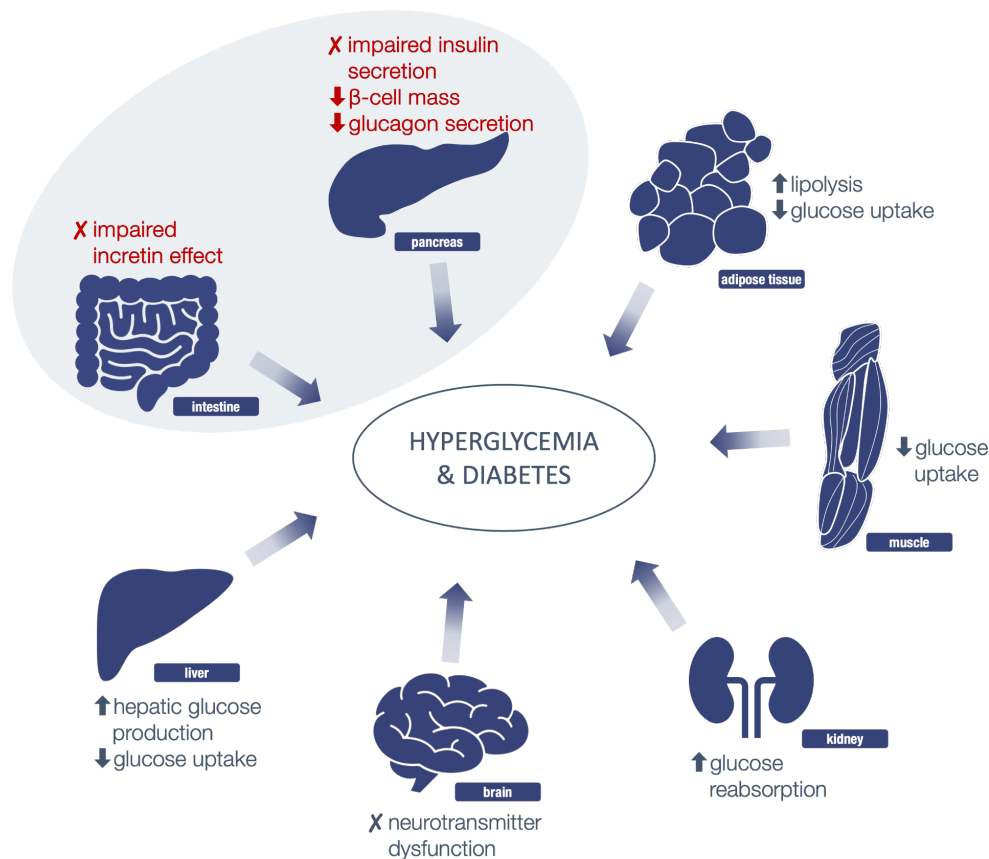
# 1. Introduction

Diabetes mellitus has become a global epidemic and heavily burdens the human health and socioeconomic system<sup>14</sup>. Diabetes is a complex and multifactorial metabolic disease caused by a pathological combination of environmental, medical, genetic and epigenetic factors of which many are still incompletely understood. Diabetic patients suffer from chronically elevated blood glucose levels because of defective glucose and lipid metabolism and hormonal dysregulation in multiple organs (**Figure 1**). A hallmark common to all types of diabetes is insulin deficiency. Pancreatic  $\beta$ -cells secrete insufficient amounts of insulin to stimulate glucose uptake in peripheral organs, because of progressive loss of functional  $\beta$ -cell mass or insulin resistance or both. If uncontrolled, the persistent high glucose levels can lead to devastating complications in many organs including enteropathy, retinopathy and nephropathy. Modern pharmacological treatments, including the new “GLP1” blockbuster, are significantly more efficacious in glucose control through added benefits on multiple organs and have improved the quality of life in patients<sup>15,16</sup>. However, they still only target the symptomatic hormone deficiencies, are applicable to only a subset of patients and it is unclear whether they will achieve sustained glycemic control and diabetes remission<sup>17,18</sup>. In addition, they often require life-long compliance and are accompanied by side effects<sup>19</sup>. The only treatments today that can put diabetes into long-term remission are bariatric surgery or transplantation<sup>20</sup>. To circumvent the need for invasive and costly treatment and improve patient outcomes in a wide range of patients, there is therefore still an unmet need for treatment options that are disease modifying, more efficacious, improve glucose control through other mechanisms and or target multiple aspects of the disease.

## 1.1. Restoring endocrine function to treat diabetes

Regenerative approaches aim at recovering the endogenous control of glucose through repairing, replacing or regenerating cells and tissues. Endocrine cells are an attractive target for such an approach because hormone dysregulation by the endocrine system and loss of functional endocrine cells are key drivers of the development and progression of diabetes (**Figure 1**). Restoring homeostasis, normal function and mass of endocrine cells holds promise to regain a fine-tuned, multi-dimensional and sustained systemic hormone regulation, which cannot be achieved through current hormone therapies that simply supplement key hormones. For a long time most regenerative treatment strategies for diabetes were focused on restoring pancreatic endocrine function, specifically regenerating insulin-producing  $\beta$ -cell mass and

function<sup>21–23</sup>. However, with the recent treatment success of bariatric surgery and therapies, which substitute or stimulate production of intestinal hormones (e.g., GLP1 or GIP incretins), targeting endocrine dysregulation in the intestine has also gained attention<sup>16,24–26</sup> (**Figure 1**). In the following sections I present potential regenerative treatment strategies for diabetes that target the pancreatic or intestinal endocrine cell system and highlight some of the remaining gaps for successful development.



**Figure 1.** The complex, multifaceted pathophysiology of diabetes. Defects in multiple organs contribute to hyperglycemia and diabetes and deregulated glucose metabolism. Main dysfunctions of the endocrine pancreas and intestine are highlighted.

### 1.1.1. Restoring functional mass of pancreatic insulin-producing $\beta$ -cells

Overcoming insulin deficiency through insulin treatment has for decades been the only available therapy for diabetic patients and is still today an important component of diabetes management. However, it is associated with risks including hypoglycemia, weight gain and developing insulin resistance. A regenerative treatment could overcome these and in addition have the potential for a disease modifying effect or even disease reversal. Insulin-producing  $\beta$ -cells are located in pancreatic islets and regulate systemic blood glucose levels together with the other islet endocrine cells - glucagon-producing alpha cells, somatostatin-producing

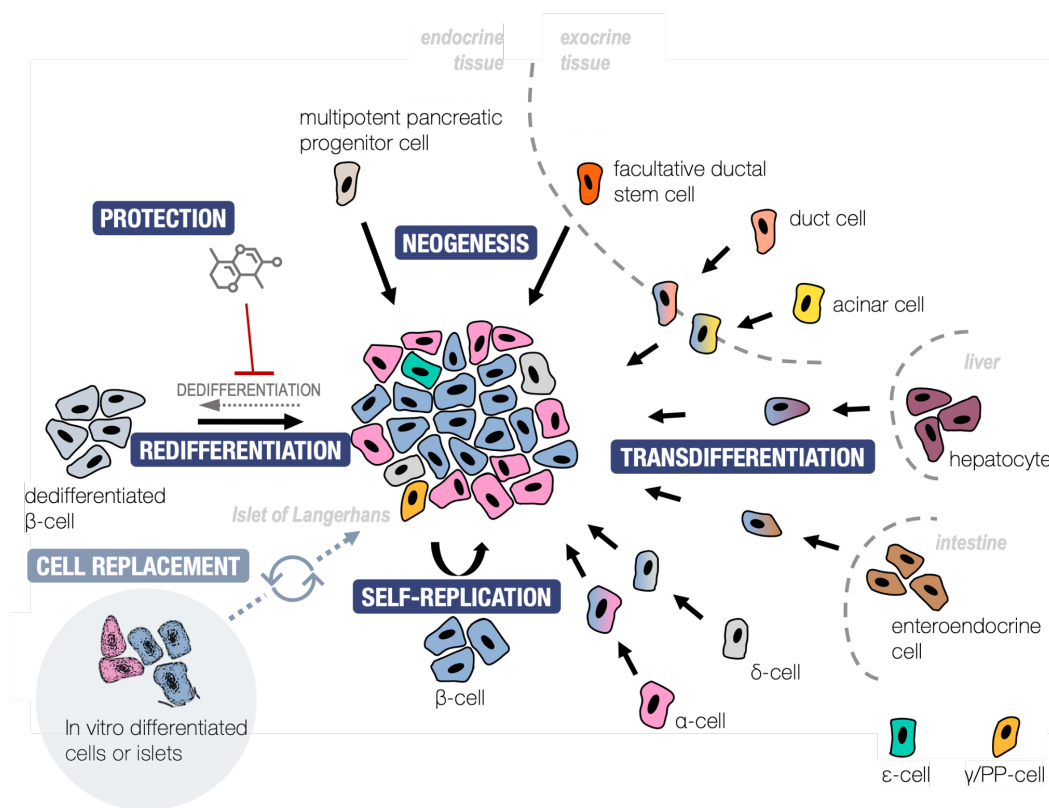
delta cells, pancreatic polypeptide-producing PP cells and ghrelin-producing epsilon cells. In the predominant forms of diabetes, type 1 (T1D) and type 2 (T2D),  $\beta$ -cells are gradually lost or become dysfunctional, due to autoimmune destruction (T1D) or the persistent exposure to metabolic stress and exhaustion due to insulin hypersecretion as peripheral organs are increasingly insulin resistant (T2D). There is no consensus yet on whether insulin resistance or hyper-responsiveness of  $\beta$ -cells to environmental factors and subsequent insulin hypersecretion is the primary driver in T2D pathogenesis<sup>27</sup>, and many of the underlying molecular pathomechanisms remain elusive. Still, it is clear that progressive  $\beta$ -cell loss and dysfunction impairs controlled insulin secretion and leads to overt diabetes. Thus, regardless of the primary defect, it is key for any regenerative therapy to stabilize functional  $\beta$ -cell mass<sup>21</sup>. Today, islet transplantation is the only established regenerative therapy for  $\beta$ -cells. It is used for a small subset of T1D patients and can normalize glucose levels but is currently not an option for most patients. Suitable donors are scarce and there are potential adverse effects of the life-long immunosuppression, which is required to prevent rejection and recurrent autoimmune destruction<sup>28,29</sup>. Xenotransplantation, for example of porcine islets, is a promising alternative to overcome donor shortage, but its clinical relevance is still limited. It is only efficacious when a large amount of islets are transplanted, immune reactions persist and outcomes of trials in non-human primates were inconsistent<sup>30–33</sup>.

Two alternative strategies to restore functional islets have gained attraction: i) cell-replacement therapy with *in vitro* stem cell-derived islet cells and ii) pharmacotherapy to protect and or regenerate  $\beta$ -cells *in situ* (**Figure 2**).

Recent progress showed that physiologically relevant  $\beta$ -cells and islet-like clusters can be generated *in vitro* from human embryonic stem cells and induced pluripotent stem cells derived from patients<sup>34–40</sup>. The *in vitro* generated  $\beta$ -cells secrete insulin in response to glucose similar to primary  $\beta$ -cells and restore normoglycemia in animal models of diabetes<sup>34,37,39–41</sup>. However, there are still barriers for stem-cell derived  $\beta$ -cells to become relevant in the clinic. The *in vitro* generated cells are heterogeneous and only a subset of cells is glucose-responsive<sup>38,39</sup>. Moreover, the cells still lack certain metabolic and transcriptional features of primary cells and do not functionally mature<sup>39,42</sup>. To differentiate cells more efficiently and improve their functional activity *in vitro*, the field would benefit from a better understanding of the factors and gene regulatory events that direct islet cells to differentiate and mature, and from biomarkers that label cells as they progress through different stages of development.

Anti-diabetic drugs that protect or restore functional  $\beta$ -cell mass *in situ* could bypass some of the hurdles associated with exogenous cell sources<sup>5,43–45</sup>.  $\beta$ -cell death is a key characteristic of T1D and T2D and increasing evidence suggests that  $\beta$ -cell stress and failure precedes T1D and T2D onset<sup>45</sup>. Thus, protecting  $\beta$ -cells from the different etiological stress factors in T1D and T2D may naturally prevent or delay disease onset. Recent findings in animal models

showed that when  $\beta$ -cells were protected from cellular stress functional  $\beta$ -cell mass could be preserved<sup>45</sup>. For example, after selectively removing senescent  $\beta$ -cells with senolytic drugs in a mouse model of T1D, more  $\beta$ -cells survived and insulin secretion was preserved<sup>46</sup>. Besides lowering stress to prevent  $\beta$ -cell failure and death, functional  $\beta$ -cell mass could be regenerated from endogenous sources in patients. The main options to restore functional  $\beta$ -cell mass *in vivo* are to redifferentiate dysfunctional, dedifferentiated  $\beta$ -cells, to expand remaining  $\beta$ -cells by triggering self-replication, to induce neogenesis from putative progenitors and to transdifferentiate other cell types into insulin-producing cells<sup>5,44</sup>. Like for *in vitro* differentiation, success of such approaches is defined by whether and how efficiently functional  $\beta$ -cells can be regenerated or protected. Thus, they rely on our understanding of the molecular programs and regulators that shape  $\beta$ -cell identity, maintain its function and determine whether candidate cell sources can be reprogrammed.



**Figure 2.** Potential approaches to protect or regenerate  $\beta$ -cells *in situ* through pharmacotherapies or cell replacement therapy. Figure was adopted from *Tritschler et al*<sup>6</sup> and further extended.

### 1.1.2. Restoring homeostasis in the endocrine intestine

Besides loss of pancreatic endocrine function, pathological factors deregulate other parts of the endocrine system. In many T2D patients, relieving insulin resistance at time of diagnosis would result in a near-normal glucose control, since they still have sufficient  $\beta$ -cell mass. A

strong risk factor for T2D and insulin resistance is obesity, and weight loss can delay progression of T2D. Sustained weight reduction is however difficult to achieve through lifestyle changes alone<sup>47</sup>. Thus, targeting aspects of the endocrine system to control excess weight in addition to lowering glucose is an attractive option, especially early in disease and in obese patients.

Multiple interventions that affect the endocrine response of the intestine have been established in the clinic to do so<sup>26</sup>. The intestine is the body's primary site of nutrient absorption and digestion and releases hormones that among other things can promote satiety, digestion and the endocrine response of the pancreas<sup>48,49</sup>. Impaired intestinal hormone secretion contributes to systemic energy and glucose imbalance, changes eating behaviors and predisposes patients to developing obesity<sup>50</sup>. Intestinal hormones are secreted by enteroendocrine cells that together with absorptive enterocytes and other specialized cells line the intestinal epithelium. When diet or nutrition change the intestinal epithelium rapidly responds to adapt its hormone signaling and more efficiently absorb and digest available nutrients<sup>48,49,51-53</sup>. This rapid response is possible because cells are constantly renewed from a pool of multipotent stem cells and because differentiated cells remain plastic. However, constant overnutrition can lead to a maladaptive response and chronically changes gut morphology, function and hormone release<sup>53,54</sup>. This is associated with two hallmarks of obesity and diabetes: food intake is excessive and incretins, gut hormones that stimulate insulin secretion upon a meal, are reduced<sup>55</sup>. Current treatment options try to intervene with both aspects.

Bariatric surgery is a potent and sustainable treatment of T2D and rapidly attenuates hyperglycemia and reduces weight<sup>20</sup>. Surgery leads to intestinal remodeling and changes in intestinal hormone and incretin secretion, which impact glucose homeostasis,  $\beta$ -cell function, appetite and food intake<sup>56</sup>. Despite its benefits, it is not economically feasible to apply bariatric surgery to manage the T2D epidemic. Therefore, minimally invasive treatments that try to mimic the effects of bariatric surgery have been developed<sup>16,24</sup>. Pharmacological approaches, which supplement production of incretins, have quickly become new blockbuster drugs for T2D and obesity as they are particularly effective in reducing hyperglycemia and weight<sup>15,17,57</sup>. However, they currently still recapitulate only one arm of the multi-dimensional response to surgery and thus are less efficacious<sup>25</sup>. In addition, it is unclear whether their effect on weight reduction,  $\beta$ -cell function, and diabetes remission are sustained long-term or when treatment is discontinued<sup>18</sup>. Thus, it will likely be more efficacious and potentially disease modifying when normal food intake and hormone regulation is regained through restoring a balanced composition and normal function of enteroendocrine and other intestinal cell types. Especially if the intervention is early and sufficient  $\beta$ -cell mass remains, this could prevent or delay the onset and progression of diabetes in obese patients. Given the high cell turnover and adaptive capacity of the gut it seems feasible to reverse maladaptation and enteroendocrine

dysregulation by directly targeting epithelial cells. To successfully develop such an approach, we still need a better understanding of the molecular and cellular features regulating intestinal lineage allocation, cell specification and plasticity in different dietary conditions and healthy, obese or T2D patients.

## 1.2. Modeling the endocrine system at single-cell resolution

While conceptually different, all regenerative strategies to restore  $\beta$ -cells or the enteroendocrine response require that sufficient truly functional endocrine cells are (re)established as a clinical endpoint to be efficacious. This depends on the ability to efficiently differentiate cells towards mature cells and induce and maintain their function.

### 1.2.1. Endocrine heterogeneity and plasticity

In both the pancreas and intestine multiple endocrine cell types exist, which together contribute to hormonal regulation. All endocrine cell types differentiate from a common progenitor, and a combination of different regulators specify their cell fate and maintain their identity. Even when fully differentiated, endocrine cells remain functionally heterogeneous and plastic. In the pancreas, islet cells differ even in homeostasis in their maturity, responsiveness to nutrient cues, proliferative capacity, hormone secretion, stress state and other phenotypes<sup>58-60</sup>. Similarly, mature enteroendocrine cells have been described to be spatially compartmentalized, remain plastic and or even polyhormonal<sup>61-64</sup>. Emerging evidence suggests that this heterogeneity allows to further fine-tune the response to physiological changes. Hence, to be able to efficiently differentiate endocrine cells and or restore their function it is key to expand our understanding of the inherent biology and heterogeneity of endocrine cell systems: how endocrine cells and their cellular ecosystem acquire their identity; how their composition changes under different physiological conditions; how they maintain function and how they respond to pathological factors; and to identify cellular sources with regenerative potential.

### 1.2.2. Elucidating cellular composition, identity and states in tissues

The importance of studying individual cells, their trajectories and molecular programs, as well as their interactions in tissues has been appreciated for many decades but limited by technical challenges. Traditional methods to study the diverse cell states and their composition in tissues use imaging or fluorescence-activated cell sorting (FACS). Although they can resolve single cells, they depend on a limited number of known biomarkers to classify different cell states. These classification schemes often cannot resolve the continuous states of transitioning cells, are difficult to apply globally, and importantly do not capture novel, non-



intuitive cell states. In recent years, single-cell genomics has become an indispensable tool to study heterogeneous, multicellular tissues like endocrine pancreas or intestine.

By now, often designated as a new area in cell biology, it has started with a bold vision. An international consortium outlined the goal to map every cell type in the human body to create a 3-dimensional Human Cell Atlas (HCA) - an atlas that will lead to a richer understanding of one of the fundamental units of living organisms: the cell - and with that accelerate both basic and clinical research<sup>65</sup>. Major technological breakthroughs have enabled it to probe cells and multicellular ecosystems at scale and with high resolution<sup>8,65</sup>. Single-cell omics technologies precisely measure genomic, epigenomic, transcriptomic or proteomic profiles of individual cells, and thus provide rich molecular cell-by-cell descriptions of developing, homeostatic, diseased or perturbed systems. While initially costly, still limited in throughput and depth and prone to introduce technical biases, the different technologies have now scaled to the point that it is possible to generate reproducible datasets of ten thousands to millions of cells at a cost affordable for most research labs. With such single-cell datasets it has now become possible to study multicellular tissues at an entirely new resolution, opening the door to addressing novel biological questions<sup>66-68</sup>.

Firstly, single-cell resolution provides us an unbiased but comprehensive view of cellular systems. Cellular populations can be unbiasedly extracted and annotated from the high-dimensional data. We can identify cell types and states by clustering similar cells, which does not rely on prior definitions, hypotheses or biomarkers and can reveal rare or novel populations that are masked in bulk measurements. Moreover, continuous processes - such as temporal axes (e.g., differentiation) or spatial axes (e.g., tissue zonation) - as well as gene expression dynamics can be accurately reconstructed when asynchronously progressing cells are sampled together.

Secondly, omics data provides a holistic view of a cell's molecular programs and is less biased towards prior hypotheses. The rich molecular profiles can be contextualized with prior knowledge or orthogonal data to decode the cell types or states in which disease genes act, to derive insights in cell function and disease mechanisms, and to discover novel therapeutic targets and biomarkers. Further, omics profiling allows us to link individual genes to context-specific gene sets and programs in which they participate. Thus, translation between patients, species, models or even related diseases can be assessed at the level of gene programs, and thus shared function and mechanisms identified even when underlying causal genes differ. Moreover, this enables us to explore the many genes or proteins that still lack comprehensive functional annotation and thereby overcome the pitfall of being restricted to the space of what has been extensively reported already<sup>69,70</sup>.

Lastly, single-cell data captures changes in both cell-intrinsic expression and cellular composition. It allows us to decipher which cell types or states respond to external cues and

how composition of a multicellular environment contributes to the response or these cues themselves, which are often confounded in bulk analyses. This can expand our understanding of how multicellular systems evolve, of the role individual cells and factors play in tissue-wide pathogenesis and enable us to predict treatment efficacy and resistance.

Since the start of the HCA initiative more than 50 million cells of 16 organs and tissues have been sequenced<sup>71</sup>. These single-cell maps and atlases of the HCA and many other studies have started to transform our understanding of biology and human disease<sup>66,67</sup>.

### 1.2.3. Computational methods empower single-cell genomics

With the increased complexity and volume of single-cell data, the ability to transform the wealth of information into biological insights ultimately depends on the quality of the computational and statistical analysis. Therefore, together with the technological and data boom, there was an equally impressive burst of novel computational methods to analyze single-cell data<sup>8,72,73</sup>. The new modes of data analysis build on methods at the intersection of data mining, machine learning and statistics to extract and discover patterns in the high-dimensional data. As volume and accuracy of the data grew, computational developments enabled researchers to extract increasingly complex aspects of fundamental biology, from modeling cellular dynamics<sup>74,75</sup> and gene regulation<sup>76</sup>, to predictions of cell fates<sup>77</sup> and perturbational responses<sup>78</sup> or inference of cell-cell communication networks<sup>79</sup>. At the same time, the rapid growth has not only forced a continuous reevaluation of the underlying statistical assumptions, but also an adaptation of methods to handle the increasing data volume and enable scalable analysis and integration of data from various sources<sup>8,80</sup>. Multiple deep learning approaches have been introduced, which integrate datasets using dimension-reduced latent spaces to build atlases not only across studies but even across data modalities<sup>81-83</sup>. With such atlases more comprehensive, robust and harmonized estimates of cellular states can be made and they bring us one step closer to modeling population-wide variability. To help navigate the sheer amount of analysis options today, many efforts have been made to benchmark methods<sup>84-88</sup>, establish best practices<sup>68,89-91</sup> and build scalable toolkits for analysis<sup>92-94</sup>. Ultimately, computation empowers the application of single-cell genomics and as methods advance paves the way for discoveries and exploring other aspects of biology.

### 1.2.4. Characterizing cell populations of the endocrine pancreas

A series of studies have created mostly transcriptomic, single-cell resolved maps of healthy and diabetic islets from human donors<sup>95-105</sup> and mice<sup>46,106,107</sup>. Cellular maps of the developing embryonic pancreas have provided insights into the series of events that define endocrine cell

fate decisions and identity<sup>108,109</sup>. Similarly, a set of single-cell studies describe the transcriptional programs pancreatic islet cells acquire during postnatal maturation and have detected novel genes that mark stages of  $\beta$ -cell maturation in mice<sup>110,111</sup>. Healthy maps of adult islet cells have been used to systematically assess islet cell composition, characterize the transcriptional programs of islet cell types including rare cell types, derive novel cell type markers, and dissect cellular heterogeneity<sup>95–97,102,103</sup>. Further, by characterizing cells of diabetic islets, aberrant gene expression could be linked to islet cell types and suggested novel genes and processes associated with pathological dysfunction, dedifferentiation or enhanced survival<sup>101,102,104</sup>. Moreover, multiple studies reported that  $\beta$ -cell subtype composition is shifted with disease<sup>97,104,112,113</sup>. A mouse islet atlas built from multiple datasets described differences in  $\beta$ -cell states and markers across disease models and life stages<sup>114</sup>. The atlas also indicates the extent of  $\beta$ -cell plasticity and response range to different stress factors. Finally, single-cell transcriptomics has been proved valuable to assess *in vitro*  $\beta$ -cell differentiation protocols. It was leveraged to evaluate differentiation efficiency as well as benchmark maturation and function against primary islets<sup>38,39</sup>. The findings of these studies have indicated the potential of single-cell transcriptomics to advance our understanding of islet cell differentiation and elucidate tissue-wide cellular heterogeneity and function<sup>5</sup>. However, the impact of early studies is still limited because of high donor-donor variation, low cell numbers or sequencing depth, and technical biases introduced by isolation and sequencing technologies as well as lack of cross-study harmonization. This not only reduced the ability to derive consistent insights across studies but also to resolve heterogeneity within cell types or infer cell state transitions. In addition, most studies depend on cell maps of a single data point, and they lack or have limited complementary biochemical or functional data. Therefore, the identified molecular  $\beta$ -cell states, their origin and their biological and functional significance remain yet to be validated.

### 1.2.5. Charting the endocrine system of the intestinal epithelium

Like for pancreatic islets, single-cell transcriptomic studies have provided nuanced descriptions of the cell types that line the intestinal epithelium in physiological and pathological conditions<sup>7,61,64,115–120</sup>. Cellular maps of healthy intestinal epithelia revealed considerable cellular heterogeneity and plasticity of enteroendocrine and other cell types even in homeostasis<sup>61,62,121–123</sup>. Composition and gene expression vary along different spatial axes highlighting how intestinal function is compartmentalized<sup>61,64,117,124</sup>. Enteroendocrine cells express one or multiple hormones, which suggests that more subtypes than captured in traditional lineage nomenclatures exist<sup>61,62,121–123</sup>. Multiple studies combined single-cell transcriptomics with a lineage reporter model to reconstruct the enteroendocrine lineage

hierarchy and describe changes in gene expression when progenitor cells differentiate and become specified as hormone-secreting cells<sup>7,62</sup>. Lastly, single-cell data indicated that even mature enteroendocrine cells are surprisingly plastic and can switch hormone expression within specified sublineages<sup>61,62</sup>. Although focused on homeostasis, these studies show the power of single-cell transcriptomics to describe the composition, lineage hierarchy, and plasticity of the enteroendocrine system indicating it will be equally powerful to elucidate mechanisms underlying intestinal maladaptation and enteroendocrine dysregulation.

### 1.2.6. Animal models to study endocrine systems

Despite rapid technological advances, increase in data volume and unprecedented insights into cells and tissues, the progression of promising preclinical findings to therapeutic success remains a bottleneck. Today, animal models and *in vitro* systems are commonly used for basic and preclinical research. It is clear that there is not a single ideal surrogate to model systemic diseases and complex cellular populations such as diabetes and endocrine cells<sup>125</sup>. Therefore, to guide selection, correctly interpret findings, identify and validate targets, or screen drugs, it is important to understand which aspects of human endocrine heterogeneity and biology are conserved in such surrogate systems.

Diabetes and endocrine cells are mostly studied in rodents. However, since endocrine development, whole-body anatomy and physiology differs between humans and rodents, it is variable how well rodent animal models can predict human physiology and positive clinical outcomes<sup>126</sup>. Pigs are an alternative to rodents with higher translational potential<sup>127</sup>. Pigs are a large-animal model, which more closely resemble human physiology, anatomy and disease<sup>128,129</sup>. Moreover, pigs have proven to be a promising source for pancreatic islet xenotransplantation, and ethical concerns are smaller for animal studies in pigs than for non-human primates<sup>130,131</sup>. Still, it is unclear whether the translational value of pigs is higher than for rodents for modeling the diverse human endocrine cell states, and their composition<sup>132</sup>. So far, no systematic computational framework or datasets have been established that enable us to evaluate transcriptional endocrine heterogeneity in rodents and pigs, and assess how relevant findings in these models are for human endocrine biology.

## 1.3. Goals of this thesis

Diabetes mellitus is a complex and heterogeneous, chronic metabolic disease with high unmet need. Diabetes is characterized by abnormally high glucose levels. Blood glucose is regulated by a multifaceted system that includes insulin and other hormones, which are secreted by endocrine cells of the pancreas and intestine. In diabetes, aspects of this hormone regulation are impaired, which ultimately leads to progressive loss of insulin-producing  $\beta$ -cells and

insufficient insulin secretion and action. Regenerative approaches that restore an endogenous fine-tuned endocrine control and function are attractive options to complement current pharmacological treatments. They are disease-modifying and thus have the potential for sustained diabetes remission or reversal. Such strategies include i) restoring or preserving functional insulin-producing  $\beta$ -cell mass through cell-replacement therapy of *in vitro* differentiated cells or pharmacotherapy to protect and or regenerate  $\beta$ -cell cells *in situ*, and ii) pharmacotherapy to regain a balanced intestinal hormone regulation through intestinal remodeling in obese patients. The development of any such approach requires restoring enough truly functional endocrine cells. Thus, it relies on an improved understanding of how differentiating cells acquire endocrine cell identity and what determines normal function and dysfunction in disease (**Figure 3**). In addition, as most pre-clinical research still relies on animal models, it depends on whether findings in animal models can be translated to humans. Since endocrine cell populations are heterogeneous and plastic, creating such understanding requires characterizing both the composition and molecular programs of individual cells in the tissue. Recent years have shown the transformative potential of single-cell transcriptomics and computational methods to study such complex, plastic multicellular endocrine tissues. Single-cell data provides the necessary cellular and molecular resolution and throughput, emerging analytic methods can extract increasingly complex facets of cell biology.

This cumulative thesis uses single-cell transcriptomics to characterize specific aspects of the inherent biology of endocrine cells in health and disease, with the goal of revealing cellular and molecular mechanisms that could ultimately be induced for regenerative treatments.

I first review the literature to address the following research questions:

- 1) How can single-cell genomics advance our understanding of the cellular heterogeneity and plasticity in the pancreas to develop regenerative therapies?
- 2) How can computational methods be used to learn cellular trajectories in development or disease from single-cell genomic data?

Then I analyze single-cell transcriptomic data to address the following research questions:

- 3) What are the cellular and molecular features of normal endocrine function in human pancreatic islets, and can they be modeled by two clinically relevant species?
- 4) What are the gene programs that shape  $\beta$ -cell identity during pancreatic development in mice?
- 5) Can single-cell transcriptomics combined with pharmacotherapy elucidate features of  $\beta$ -cell dysfunction and regeneration upon long-term hyperglycemia in mice?

- 6) How does composition and function of the intestinal epithelium and enteroendocrine system change in obese mice and what are mechanisms to revert these changes?

## 1.4. Summary of results

The series of research articles I present provide novel insights into cellular and molecular mechanisms of endocrine identity, function and dysfunction in disease (**Figure 3**).

In two review articles I summarize and discuss the literature and current understanding of the impact of single-cell transcriptomic data to i) elucidate the heterogeneity and plasticity of the pancreas<sup>5</sup>, and ii) computationally learn cellular trajectories in development or disease<sup>6</sup>.

I then analyzed single-cell transcriptomic data zooming into aspects of the developing, healthy and diseased endocrine and intestinal pancreas. To validate biological and functional significance of the findings and add mechanistic insights, this was complemented by my collaborators with lineage tracing, other omics, biochemical and functional data.

First, I created a comprehensive reference map of healthy human, mouse, and, for the first time, pig pancreatic islet cells to identify evolutionary conserved and distinct molecular features of islet cell biology<sup>1</sup>. The map confirmed that human alpha- and  $\beta$ -cells are transcriptionally heterogeneous, and captured alpha- and  $\beta$ -cell states associated with mature, functional cells as well as distinct cellular stress responses. I established an analysis framework that facilitates cross-study and cross-species comparisons through prior knowledge- and data-driven gene sets instead of individual genes. I validated the identified human cell states in 9 publicly available studies. Cross-species analysis indicated that pig and mouse cells mirror a subset of the human states, and it revealed aspects of gene expression that are conserved or diverge within these states. A subset of functional and regulatory human features was better reflected in pigs than mice, which demonstrates the potential of pigs as a model for studying human islet cell biology. Together, this is a valuable framework and resource for the field that established a healthy baseline to study disease and will facilitate modeling endocrine cell heterogeneity and function across species.

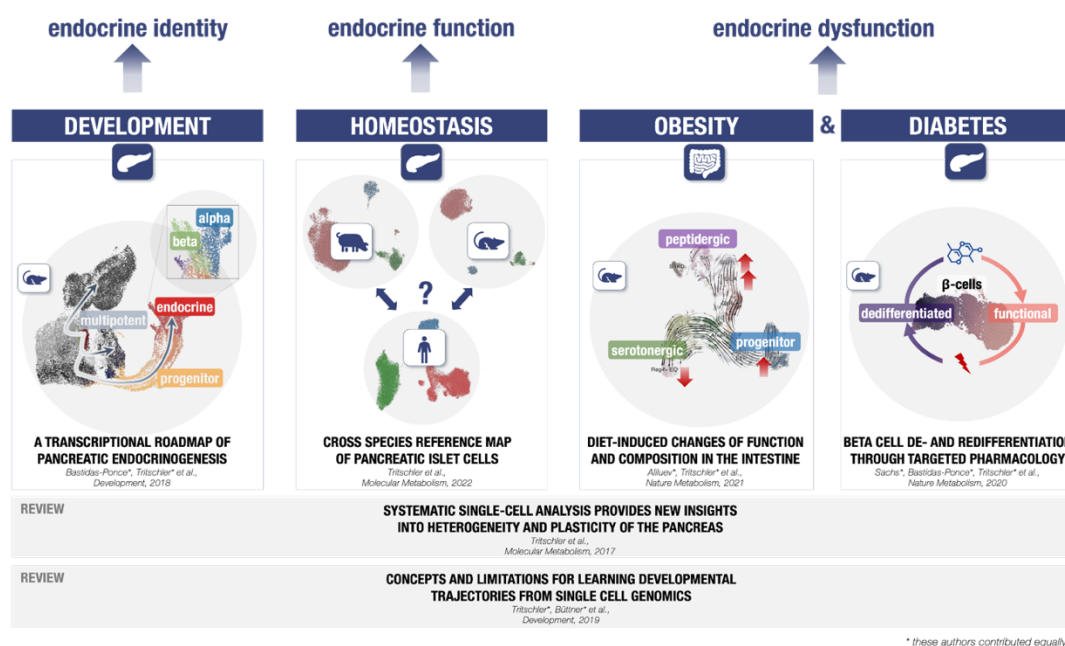
Next, I reconstructed a lineage model and the differentiation trajectories from endocrine lineage cells, which were isolated from developing embryos at multiple time points<sup>2</sup>. This revealed that transient endocrine precursors are heterogeneous and express a molecular signature that is stage-dependent and defines endocrine cell allocation. These novel signatures are a blueprint of how cells become stepwise lineage restricted and acquire their identity during endocrinogenesis. Hence, the model provides a guide on how to direct differentiation towards a specific endocrine cell fate and label differentiation stages, which is important to improve protocols to differentiate  $\beta$ -cells *in vitro* for cell replacement, and,

likewise, to re-differentiate dedifferentiated  $\beta$ -cells or reprogram non  $\beta$ -cells to  $\beta$ -cells to regenerate functional  $\beta$ -cell mass *in vivo*.

Further, I created a transcriptional map of pancreatic islet cells from healthy, diabetic and treated mice to characterize how  $\beta$ -cells respond to a long-term hyperglycemic environment and pharmacological treatment<sup>3</sup>. It revealed that in diabetic mice a subset of  $\beta$ -cells survived, but dedifferentiated to a dysfunctional, immature-like  $\beta$ -cell state, and that they responded to pharmacological treatment that restored  $\beta$ -cell function. Trajectory inference suggested that  $\beta$ -cells de- and redifferentiate upon treatment and uncovered molecular processes governing these cellular transitions and treatment effects. Overall, the study positions  $\beta$ -cell re-differentiation as a viable new treatment option to halt or revert diabetes. It also substantiates that single-cell data combined with perturbations is a powerful approach to uncover molecular mechanisms as well as pharmacological entry points.

Finally, I used single-cell transcriptomic data to chart the landscape of small intestinal crypt cells, their lineage relationships and how these adapt to an obesogenic diet<sup>4</sup>. We found that an obesogenic diet changed cellular composition, which altered enteroendocrine hormone secretion, shifted nutrient absorption and led to morphological changes of the gut epithelium including prolonged villi and enlargement of specific regional zones. The molecular profiles suggested that this dysfunction emerges already at the level of stem cells and early progenitors. The study identified factors that may contribute to intestinal maladaptation and enteroendocrine dysfunction in obese mice and therefore are potential therapeutic entry points to prevent progression to diabetes.

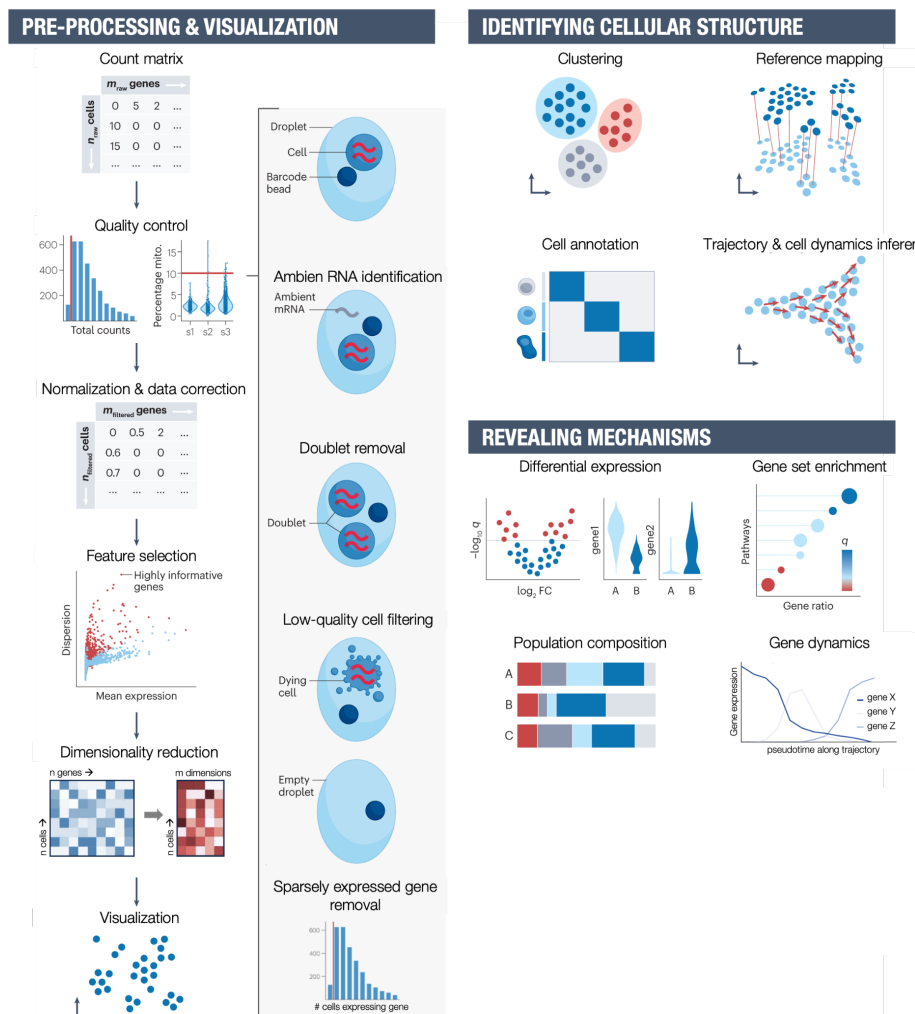
Summaries of the presented articles and my contributions can be found in *chapter 3*.



**Figure 3.** Overview of research questions and contributions of this cumulative doctoral thesis. The figure was generated using images from<sup>1-4</sup>.

## 2. Methods

Single-cell transcriptomics has revolutionized the way we explore multicellular systems. With the rapidly growing volume and complexity of single-cell transcriptomic data, the number of computational methods to analyze such data has grown equally fast. Since the ability to extract meaningful biological insights ultimately depends on the quality of computational analysis, analysis workflows are tailored to research questions, technology and experimental design. While many routes of analysis exist, methods can be broadly bucketed to address three tasks: data processing and visualization, identifying cellular structures and revealing gene or tissue-level mechanisms<sup>90</sup> (**Figure 4**). Benchmarking efforts<sup>84–88</sup>, best practices<sup>89,90</sup> and scalable analysis tool-kits<sup>92–94</sup> aim at introducing some standards in the field and help researchers navigate the large number of analysis options.



**Figure 4.** Overview of processing and analysis steps of single-cell transcriptomic data. The figure was adopted from Heumos et al.<sup>90</sup> and extended to reflect the workflow used in this thesis. Reproduced with permission from Springer Nature (License Number 5774150778617).



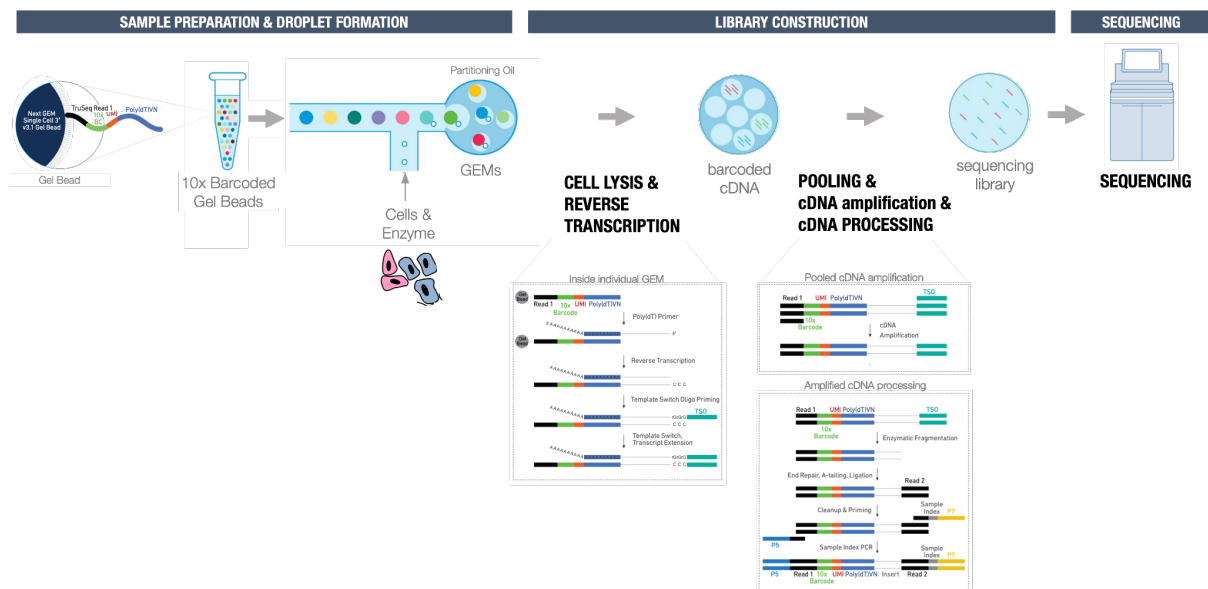
In this chapter, I summarize the single-cell RNA sequencing (scRNA-seq) technology and the computational concepts and methods used in this thesis. First, I describe the technology that was used to generate scRNA-seq data (*section 2.1*). Next, I summarize how raw sequencing data was processed and quality controlled (QCed) (*section 2.2*). Finally, I describe the main methods and underlying concepts I used to analyze, explore and interpret data to identify cellular structure (*section 2.3*) and to elucidate underlying mechanisms (*section 2.4*). As the field rapidly develops, new methods outperform older approaches or address new questions. While the overall analysis likely remains valid, tool recommendations, applications and best practices can change. I focus this overview on methods used in the articles of this thesis and available at time of publication and only briefly highlight development since then. For all steps, I followed best practices and used methods as recommended by the single-cell field<sup>89,90</sup>. Today, most tools are easily accessible through rich analysis platforms<sup>92-94</sup>. Unless stated otherwise, I used the Scanpy python toolkit for analysis<sup>92,133</sup>. For details on QC decisions specific to each dataset, marker genes for cell annotation and other parameter choices I refer to each manuscript<sup>1-4</sup>. To derive biological conclusions, we interpreted and contextualized all analysis results with other omics, biochemical and functional data that was generated by my collaborators as outlined in each manuscript.

## 2.1. Single-cell RNA sequencing technology

ScRNA-seq technologies measure messenger ribonucleic acids (mRNAs) of individual cells to generate gene expression profiles of every cell in a sample. The workflow of scRNA-seq involves five steps: i) isolating and capturing individual cells, ii) lysing cells to extract their RNA content, iii) reverse transcription to convert RNA into complementary DNA (cDNA), iv) amplification of the cDNA to generate a sequencing library, and v) sequencing in a next-generation sequencing platform (**Figure 5**).

Various methods have been developed and continue to evolve and optimize different steps in the workflow to suit different research needs. The most commonly used scRNA-seq technologies mainly differ in their approach of cell isolation and capture, barcoding strategies, and library preparation, which affect throughput, sensitivity, transcript coverage, scalability or cost of a method<sup>134</sup>. Here we used the 10x Chromium system<sup>135</sup>, which uses advanced microfluidics to facilitate parallel processing and efficient profiling of hundreds to thousands of cells, and thus enables scalable and cost-effective profiling of large cell populations generate a sequencing library, and v) sequencing in a next-generation sequencing platform (**Figure 5**). The 10x protocol version used selectively captures and sequences the poly-adenylated 3' end of the transcripts<sup>135</sup> (*Chromium Next GEM Single Cell 3' Reagent Kits v3.1, Document Number*

CG000204 Rev D, 10x Genomics, (2022, July 13)). This 3' enrichment is more cost-effective and increases gene coverage per cell compared to full-length methods as sequencing is focused on a small part of the transcript and therefore overall sequencing depth is reduced. Moreover, the 10x system tags each RNA molecule with a unique molecular barcode (UMI) to reduce biases introduced when cDNA is amplified and quantify gene expression more accurately.



**Figure 5.** The 10x Chromium system uses microfluidics to encapsulate individual cells and uniquely barcoded gel beads into droplets. Within the droplets, cells are lysed to extract their RNA content. Then the gel beads prime reverse transcription to convert RNA into molecule- and cell-specific barcoded cDNA. Finally, cDNA is amplified, and the library prepared for downstream sequencing with a next-generation sequencing platform. The figure was generated using images from the 10x Genomics user guide (*Chromium Next GEM Single Cell 3' Reagent Kits v3.1, Document Number CG000204 Rev D, 10x Genomics, (2022, July 13)*).

## 2.2. Preprocessing of single-cell RNA sequencing data

Preprocessing represents the start of an analysis workflow and filters and transforms raw data in an attempt to remove unwanted sources of variation and improve signal-to-noise ratio. Systematic and random noise inherent to scRNA-seq data, technical artefacts or biological confounders such as cell cycle can bias analyses and distort biological findings. To remove such unwanted effects, it is important to apply appropriate analytical and statistical methods that account for underlying structure and complexity of the data. Preprocessing steps usually involve read alignment, quality control, data normalization and correction as well as feature selection and dimensionality reduction (**Figure 4**).

### 2.2.1. Read alignment, count matrix generation and quality control

Processing and QC steps are required to generate a cell-by-gene matrix of gene expression from the raw output of the sequencing protocol. I preprocessed raw sequencing data of each sample using the Cell Ranger v.2.0.0 analysis pipeline provided by 10x Genomics<sup>135</sup>. The pipeline performs barcode demultiplexing to assign each read to a sample, droplet & UMI, aligns reads to the corresponding reference genome, QCs and filters reads, and quantifies UMIs to create a matrix of raw mRNA counts per droplet-barcode. For the 10x protocol a droplet-barcode can tag a single cell, a doublet (multiple cells in one droplet) or an empty droplet. The pipeline calls non-empty droplet-barcodes based on the overall distribution of total UMI counts per barcode. Doublets were identified in a downstream step (*section 2.3.3*).

Subsequent QC of the raw count matrix further guaranteed sufficient data quality for downstream analysis (**Figure 4**). For each sample, QC steps included (1) to remove sparsely expressed genes and (2) exclude low quality or outlier cells with high fraction of mitochondria-encoded counts (>20%) or total UMI or gene count at the tails of their distribution. For every step, I defined QC thresholds through visual inspection of distributions. I set permissive QC thresholds to avoid excluding viable cell populations and revisited the thresholds if affecting downstream analysis performance poorly<sup>89</sup>.

### 2.2.2. Normalization and batch correction

After QC, data is normalized to account for the variability in count depth (total number of counts), which is present due to inherent sampling effects in mRNA capture, reverse transcription and sequencing (**Figure 4**). Multiple approaches to normalize scRNA-seq data have been developed over recent years, and their performance varies between datasets<sup>136,137</sup>. Linear, global scaling methods scale count data cell-wise to calculate relative count abundances in a cell and allow for better comparison between cells with different count depth. I used the following linear methods to calculate size factors for scaling.

- 1) Total counts of a cell.
- 2) Total count normalization after excluding genes contributing to more than 5% of the total counts in a cell as suggested<sup>138</sup>. This extension accounts for variability in total counts introduced by few highly expressed genes (e.g., hormone-encoding genes).
- 3) SCRAN algorithm<sup>139</sup>, which estimates cell-wise size factors from pools of cells to account for heterogeneity in cell size and reduce the effect of technical drop-out.

Normalized data is then log-transformed:  $\log(\text{count} + 1)$ . Log-transformation aims at representing differences in gene expression as log-fold changes, and at reducing both the observed dependence of the variance on the mean and data skewness to better approximate

normally distributed data. Although scRNA-seq data is not log-normally distributed, log transformation is useful for these purposes. Lastly, to further reduce unwanted technical variability between batches I performed linear data correction using the ComBat algorithm<sup>140</sup>. I only performed batch correction if warranted necessary after visual inspection of how well batches are distributed between clusters and how well they overlap in low-dimensional representation of the data. For all normalization and correction steps I fixed zero values to avoid spurious sample-to-sample differences around zero. I then used the QCed, normalized and corrected count matrices as input to analyses unless indicated.

### 2.2.3. Feature selection, dimensionality reduction and visualization

Single-cell data is high dimensional, complex, noisy and sparse, which poses a challenge to analyze, visualize and interpret the data, and computational efficiency. Two key steps implemented to overcome this are feature selection and dimensionality reduction (**Figure 4**). They aim at finding a lower-dimensional representation of the data, which retains the most informative features and preserves relevant structure, while it filters noise and reduces complexity and size of the data. This assumes that not all genes are needed to sufficiently describe informative and relevant biological variation, i.e., cellular expression profiles and heterogeneity. As the first step, the gene feature space is normally reduced to top 2000-4000 highly variable genes (HVGs). HVGs are the genes that contribute most to the observed variation in the data and were here defined as the genes with the highest variance-to-mean ratio after binning all genes by mean expression<sup>94</sup>. The number of HVGs selected depended on the task and complexity of the input data. Alternatively, in specific cases I defined the gene feature space based on prior knowledge to focus the analysis on specific biological processes of interest and facilitate integration across datasets. Next, dimensionality is further reduced through summarizing the data in a lower-dimensional space, in which features represent a linear or non-linear combination of selected genes. For single-cell data often multiple dimensionality reduction techniques are applied, e.g., linear data transformation methods such as principal component analysis (PCA) precede non-linear transformations. The dimensionality reduction techniques used in this thesis included:

- I. PCA, a linear dimensionality reduction method that transforms the data to a low-dimensional space by maximizing the variance explained by each feature (principal component). PCA is a pre-processing step to many non-linear dimensionality reductions and downstream analysis steps.
- II. Diffusion maps, a non-linear dimensionality reduction method that aims at preserving local similarities and diffusion processes, and thus is useful to capture continuous patterns such as cellular differentiation and transitions<sup>141,142</sup> (see also *section 2.3.5*).

- III. Correlation based-gene sets<sup>143</sup>, which summarize the data through sets of highly correlated- and anti-correlated genes. They enable identifying context-specific gene sets to describe cellular states independent of curated markers or pathways.
- IV. UMAP algorithm<sup>144</sup>, which is a non-linear, graph-based method that aims at preserving local and global structures in the data. UMAP is used to compute a 2D embedding to visualize scRNA-seq data and was shown to better scale when compared to other visualization techniques.

## 2.3. Analysis to identify cellular structure

Once data is preprocessed, downstream analyses aim at extracting biological insights from the data. Here, I will consider these analyses from two viewpoints: cell-level approaches to identify different cellular populations (*this section*), and investigations to describe mechanisms of cell or tissue-wide phenotypes (*section 2.4*). Cell annotations explain cellular heterogeneity in the data, and many downstream analyses build on these descriptions of cellular structures in the high-dimensional data. Cell annotations can be obtained from either identifying discrete states (*section 2.3.1-2.3.3*), through integration and label transfer from a reference (*section 2.3.4*) or by modeling continuous processes (*section 2.3.5 and 2.3.6*).

### 2.3.1. Cell clustering

Clustering is usually the first step in downstream analyses to determine the identity of a cell and the cellular composition of a sample (**Figure 4**). Clustering methods group cells with a similar gene expression profile, typically represented by the low-dimensional embedding. I employed an unbiased graph-based approach for clustering, which aims to detect cellular communities in a graph-representation of the data<sup>145</sup> and has been shown to perform best for analyzing single-cell data<sup>88,89</sup>. First, a cell-cell graph is built using a k-nearest neighbor (kNN) approach, which connects each cell to its k most similar cells<sup>144</sup>. The similarity between cells is determined by the Euclidean distance in the low dimensional PC space. Next, the community detection algorithm identifies regions in the graph, in which cells are more densely connected than expected. For community (cluster) detection in the kNN-graph I used the Louvain algorithm<sup>145</sup> as implemented in the *louvain-igraph* package<sup>146</sup>. To account for regions with differences in cell densities and or biological variability beyond cell type identity (e.g., cell cycle), clustering resolution can be adapted. I iteratively explored different cluster resolutions in different parts of the graph to control the number of clusters, and merged clusters that expressed the same genes of interest.

### 2.3.2. Cell annotation approaches

Cell annotation approaches assign each cluster or cell a biologically meaningful label. They can broadly be grouped into manual and automated approaches. Manual approaches leverage the gene expression profile of a cluster and contextualize it with prior knowledge. To do so, the profile of one cluster is compared to the profiles of other clusters in a dataset. Here, I derived cluster profiles through a differential expression test between the cells in one cluster versus the cells of all other cells (t-test with overestimated variance or the Wilcoxon rank sum test). I then annotated each cluster based on enriched literature-curated marker genes, which are known to be expressed by the diverse cell types within a tissue. To increase the confidence in these annotations I confirmed that not only individual markers, but also expected biological processes are enriched in the cluster profiles (*section 2.4.2*).

In addition to a gene-level characterization, cells and clusters can also be annotated based on the activation of a gene set derived from prior knowledge (e.g., a pathway) or the data (e.g., correlated genes). Using gene sets for annotation is less prone to noise and or drop out in the data, which is important especially when comparing different conditions or datasets. Gene set activations are determined per cell, and thus can also be used to annotate cells independently of clustering. Here, I computed gene set activation in a cell as the mean expression of the genes in the gene set subtracted with the mean expression of a gene set of the same size randomly sampled from a background of genes with similar expression values<sup>94</sup>. I used prior knowledge or literature-derived annotation for cell cycle classification<sup>147</sup>, functional annotation with hallmark gene sets<sup>148</sup>, and inference of spatial location using zoned or regional markers of the intestinal epithelium<sup>124</sup>. Examples of data-driven, context-specific genes sets included inferred cell type markers, differentially expressed genes, and correlation based-gene sets (*section 2.2.3*), which facilitated the annotation of cell states across datasets and systems. Automated approaches for cell annotation include pre-trained cell type classifiers<sup>149</sup> or reference mapping (*section 2.3.4*). The quality of these automated annotations depends on the quality and suitability of the training data as well as the model used.

### 2.3.3. Identification of doublet clusters

Doublets arise from droplets containing more than one cell. If a doublet is formed by two different cell types, this can lead to wrong annotations and conclusions. Doublets can be identified as cells simultaneously expressing profiles of two cell types or clusters, and thus can be simulated artificially by randomly sampling pairs of cells in a dataset. Multiple algorithms have been developed for doublet detection, which compare such simulated doublets to measured cells<sup>150</sup>. However, it can be a challenge to distinguish doublets from transitioning or “in between phenotype” cells such as polyhormonal cells and detection accuracy can vary

between datasets<sup>151,152</sup>. To increase accuracy, I therefore combined a doublet detection method with the following criteria to identify and exclude doublet clusters.

- 1) High doublet scores as inferred by the doublet detection algorithm Scrublet<sup>153</sup>.
- 2) No uniquely expressed genes are detected in a doublet cluster.
- 3) Doublets express marker genes at similar levels as doublet “contributor” cells.
- 4) Doublets are not previously reported or experimentally validated polyhormonal cell.
- 5) Observed fraction of a doublet cluster matched expected fractions of singlet clusters.

#### 2.3.4. Reference mapping for cross-species comparison

One of the main goals when comparing multiple systems, e.g., different species, is to identify shared and differential cell states and assess feature similarity within them. This requires defining a feature space and approach that allows to reliably match and compare cells between the systems. Gene-level comparisons can be affected by noise or sparsity of the single-cell data, and specifically for cross species data they can also be limited because reference genomes differ in quality, genes lack functional annotation or mapping of gene orthologs is poor. Dimensionality reduction to summarize the data (*section 2.2.3*) can help to address these challenges as features are defined by groups of genes and therefore feature activation is less sensitive to individual genes. In this thesis, the aim was to assess to which extent human cell states and their gene expression profiles are conserved in other species. This is conceptually similar to reference mapping approaches, which first build a low-dimensional embedding for the reference (human), and then map queries (other species) to this or an updated joint space for cell annotation or other more advanced applications of mappings between reference and query<sup>83,154–156</sup>. Many reference mapping approaches are designed to build references from complex and heterogeneous tissues and use machine learning or deep models to do so, however the features of the latent space are not directly interpretable, and thus cannot also be used to characterize cells. Therefore, I used here a simpler, but interpretable approach for reference mapping. To represent human data, I used correlation-based gene sets<sup>143</sup> as a context-specific and interpretable feature space that well described the human cell states (*section 2.2.3*). Then, pig and mouse cells were matched to the human reference states by first embedding the pig and mouse data in that feature space and then predicting most similar cell state labels (**Figure 4**). For mapping the embedding the Scanpy ingest-functionality<sup>92</sup> was used, which includes a simple kNN-classifier for label transfer and projecting the query data to the reference. With this reference mapping approach each pig and mouse cell was assigned its closest human cell state label. This allowed to

quantify conservation of human biology considering both cellular composition and similarity of gene expression within cell states.

### 2.3.5. Trajectory inference and gene dynamics in continuous data

While clustering partitions the data into discrete groups of cells, trajectory inference (TI) aims at modeling continuous structures in the data (**Figure 4**). TI approaches can be used to model lineage relationships and developmental paths of cells; the temporal or spatial order of cells; as well as gene expression dynamics regulating these continuous processes. Since single-cell data usually only provides a snapshot of such a continuous process this is a challenge and relies on computational inference. TI approaches use gene expression similarity to infer cluster or cell relationships and reconstruct cellular trajectories from the relative ordering of cells based on these similarity measures. Many TI methods exist and show variable performance for different tasks<sup>157</sup>. I used the following methods for TI:

- I. To infer lineage relationships between clusters I used partition-based graph abstraction (PAGA)<sup>158</sup>. PAGA summarizes the structure and relationships within the high-dimensional data by creating a coarse abstracted graph representation from the single-cell kNN graph, in which nodes represent clusters of cells and edge weights the connectivity between clusters. Clusters are determined through clustering or other cell annotations. The PAGA algorithm computes connectivity from the likelihood that the fraction of connections in the kNN graph between two groups exceeds the fraction of connections between randomly assigned groups. Through that PAGA identifies the connected and disconnected parts of the data and discards spurious connections in the noisy single-cell kNN graph.
- II. I used the diffusion pseudotime method<sup>159</sup> to reconstruct hypothetical axes (cellular trajectories) in the data that represent continuous processes, e.g., a high-confidence path in the PAGA graph. Diffusion pseudotime uses the concept of diffusion to compute a cell-to-cell distance. It simulates a diffusion process through random walks in the single-cell kNN graph starting from a predefined source cell and assigns each cell a pseudotime value representing its relative position to the source. By ordering the cells based on the pseudotime value a one-dimensional cellular trajectory from the source cell is then reconstructed. I visualized pseudotime and the gene expression dynamics along the inferred trajectories to identify branching events and gene expression dynamics. When multiple “orthogonal” continuous processes were present in the data (e.g., spatial and temporal trajectories), I focused pseudotime inference on processes of interest through prior feature selection.



### 2.3.6. Inference of cellular dynamics

TI approaches provide only a static, undirected representation of cellular trajectories. To infer cellular dynamics, i.e., to predict the fate of cells and direction of cellular transitions, RNA velocity inference<sup>75</sup> can be used. RNA velocity inference leverages the splicing status of transcripts in cells to estimate the speed and direction of gene expression changes. Velocity inference is based on the principle that the rate of transcription and splicing can be approximated through the ratio of unspliced to spliced transcripts in cells. In addition to providing insights into cellular dynamics, RNA velocity analysis also facilitates identification of genes that drive these cellular transitions. Putative driver genes of cell fate decisions can be identified through examining the velocity vectors of individual genes. Here, I used the scVelo method<sup>74</sup> for velocity inference, which aims at solving full transcriptional dynamics to generalize the concept of RNA velocity. scVelo estimates RNA velocities by fitting a likelihood-based dynamical model of transcription for each gene. To predict the velocity of potential cell movements in a manifold, the RNA velocity vector of individual cells is then projected into the low-dimensional UMAP embedding. First, a velocity graph is computed by assigning each edge in the single-cell kNN graph (possible transitions) a transition probability computed from the correlation of the possible and predicted transition given by the velocity vector of a cell. The projected directional movement or flow of cells in the UMAP is then the expected mean direction in the velocity graph. Building on these concepts, *CellRank* is a method that combines similarity-based TI with RNA velocity vector fields to detect start and end population in a trajectory, predict cell fate probabilities in multi-lineage trees, and visualize gene expression dynamics<sup>77</sup>. In addition, *CellRank* is built modular and supports other approaches to compute cellular transitions. For example, it leverages the concept of optimal transport (OT) to include experimental time and infer directionality<sup>160,161</sup>. OT can be leveraged to map a cell to its most likely progenies in adjacent time points. It does so by comparing and matching the probability distributions of cells at different time points to identify the most efficient way to “transport” cells from one time point to the next and through that predict transition probabilities.

## 2.4. Analysis to reveal mechanisms

Discrete or continuous cell annotations facilitate analyses to describe regulatory elements of cell or tissue-wide biological processes. Such analyses aim at deciphering mechanisms underlying cell decision-making and phenotypic changes and include detecting differences in gene expression and regulation and cell composition.

### 2.4.1. Differential expression analysis

Differential expression (DE) analysis is used to reveal differences in gene expression between two experimental conditions such as healthy and disease tissue (**Figure 4**). DE analysis is usually performed within a cell type and separated from compositional analysis to understand how a cell type responds to different biological conditions. Here, the limma-trend method<sup>162,163</sup> was used for DE, which was originally developed for bulk RNA-seq data. Limma-trend and other bulk methods have been shown to perform well for single-cell data in a large-scale benchmarking<sup>84</sup>. To facilitate interpretation, differential gene profiles can be contextualized with prior knowledge (*section 2.4.2*).

Notably, differences in background mRNA levels between samples can pose a problem in DE analysis as it leads to mistakenly called differential genes<sup>90</sup>. Background or ambient mRNA arises from mRNA of lysed cells that freely floats in the single-cell solution and is incorporated into every droplet during sample preparation. Here, I identified differential ambient genes as being ubiquitously expressed in all cell types or inferred them from the DE profiles of empty droplets. I excluded highly ambient genes before interpreting DE results.

### 2.4.2. Prior knowledge contextualization through gene set enrichment

To facilitate interpretation of a gene set or signature, e.g., differentially expressed genes, they can be contextualized with prior knowledge (**Figure 4**). For example, gene set enrichment analysis can be performed to assess whether specific gene sets from a background list of gene sets are overrepresented in the query signature. These background gene sets are derived from databases of literature curated gene sets that represent biological pathways, transcription factor targets, disease or drug perturbation responses or similar. For gene set enrichment I used the EnrichR framework<sup>164</sup>, which employs the hypergeometric test or the Fisher's exact test (if a background list is provided) to assess whether an input list of genes is significantly overrepresented in the analyzed gene set.

### 2.4.3. Population composition analysis

Compositional data analysis is used to investigate shifts in cell type or state composition between two conditions (**Figure 4**). Two common challenges need to be accounted for when analyzing compositional shifts. Firstly, most single-cell datasets have a low number of replicates. Secondly, single-cell data is proportional in nature. Cell type counts are relative measures and not independent of each other because only a subset of the total cells in a sample is captured. This means that cell type abundances in a sample are fractions, which always sum up to one and therefore are negatively correlated. To account for this proportional

nature of the data when comparing cell type compositions, I used a Dirichlet-Multinomial model as implemented by my collaborator B. Schubert. I only applied compositional analysis if cell numbers were sufficient to accurately determine the fraction of cells in a cluster, and sample numbers were at least 3 to ensure sufficient power to detect shifts despite expected variations in compositions due to sampling or other technical effects. scCODA is a tool that builds on a similar statistical approach and tackles compositional analysis and low replicate number with a Bayesian model<sup>165</sup>.

## 3. Summary of contributed articles

This chapter summarizes my contributed articles in the context of the goals of this thesis.

### 3.1. Research Articles

#### 3.1.1. A transcriptional cross species map of pancreatic islet cells

**Sophie Tritschler**, Moritz Thomas, Anika Böttcher, Barbara Ludwig, Janine Schmid, Undine Schubert, Elisabeth Kemter, Eckhard Wolf, Heiko Lickert, Fabian J. Theis, **A transcriptional cross species map of pancreatic islet cells**, *Molecular Metabolism*, Volume 66, 101595 (2022)

(See also publication <sup>1</sup> in the bibliography and in Appendix A)

#### **Summary**

Disease can only be fully understood in the context of a healthy reference, and vice versa, therefore mapping the cellular landscape of healthy tissue is key to study diseased tissue. Similarly, it is important to understand conservation of the healthy baseline in a model system to assess its translational value for disease. Here, we leveraged scRNA-seq to aid in our understanding of healthy pancreatic islet cells in humans, and their conservation in two commonly used species models, pig and mouse. We generated a comprehensive healthy reference dataset of >50'000 cells isolated from human, pig and mouse islets and made it available as a queryable and *Findable, Accessible, Interoperable, and Re-usable* (FAIR) data resource<sup>166</sup>. Our cross-species map provides high cellular (number of cells per donor) and molecular (sequencing depth) resolution when compared to previous studies of human pancreatic islets, and for the first time, single-cell transcriptomic data of pig islets. We delineated human alpha- and  $\beta$ -cell heterogeneity and described cellular states that model mature, functional cells as well as distinct cellular stress responses. The cell states were consistently identified across multiple studies. Moreover, we systematically compared gene expression across species and mapped mouse and pig cells to the human reference map, which suggests that pigs can be a model of gene expression relevant for human islet cell biology. Together, this is a valuable and easily accessible resource that benefits future studies as a comprehensive reference and framework to investigate the transcriptional programs of mature, functional or stressed human islet cells, and their conservation in two clinically relevant animal models.

#### **My contribution**

- I. I conceptualized the study together with Heiko Lickert and Fabian J. Theis.

- II. All data was generated by co-authors. I contributed to the generation of the scRNA-seq data (10x protocol).
- III. I preprocessed all raw data of the manuscript as described in *section 2.2*. Moritz Thomas contributed to the quality control and normalization steps of the human and pig data, gene mapping and performed early exploratory analyses.
- IV. I analyzed all data of the manuscript, reviewed the literature and interpreted the results.
  - A. I characterized 50'000 endocrine cells isolated from healthy human, pig and mouse pancreatic islets.
  - B. To evaluate conservation of human endocrine cell type signatures in mouse and pig, I established a framework to systematically compare and assess similarity of gene expression profiles across species both globally and focusing on specific features (literature markers, data-driven cell type enriched marker genes, transcription factors). Such a cross-species analysis framework has not been previously described for single-cell transcriptomic data. This is also the first study using single-cell transcriptomics for pig cells.
  - C. To characterize human alpha and beta cell heterogeneity and cell states I used data-driven correlation-based gene sets as well as known biological processes of beta and alpha cell function, cellular stress and hormone signaling. This thorough characterization revealed a continuum of cellular states of varying maturity and function linked to differential stressors, pathway signaling and hormone receptor expression. To validate the functional and biological relevance of these states I curated and annotated 10 publicly available single-cell studies of human adult pancreatic islets (>55 donors).
  - D. I identified novel potential beta cell maturation factors through RNA velocity analyses. I validated the biological relevance of these factors in two publicly available single-cell studies of human pancreatic development.
  - E. To evaluate conservation of the human alpha and beta cell states I used a reference mapping approach to annotate pig and mouse cells with human state labels, which uses correlation-based gene sets as features and although simple has as such not been previously described. I then systematically compare composition and expression profiles as described in B) of these states between species.
- V. I wrote and edited the manuscript and designed all the figures. All co-authors contributed to the discussion of the results and reviewing of the final manuscript.

### 3.1.2. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis

Aimée Bastidas-Ponce\*, **Sophie Tritschler\***, Leander Dony, Katharina Scheibner, Marta Tarquis-Medina, Ciro Salinno, Silvia Schirge, Ingo Burtscher, Anika Böttcher, Fabian J. Theis, Heiko Lickert, Mostafa Bakhti; **Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis**, *Development* 15; 146 (12): dev173849 (2019)

(See also publication <sup>2</sup> in the bibliography)

#### Summary

Cell replacement and *in vivo* regeneration are promising disease-modifying therapies to restore functional  $\beta$ -cell mass in diabetic patients. They both build on differentiating  $\beta$ -cells from endocrine lineage cells, and thus benefit from a better understanding of the factors that regulate endocrinogenesis during development. Single-cell transcriptomics is a powerful tool to study development. From the snapshot of asynchronously developing cells, cellular trajectories can be reconstructed *in silico* to chart and explore cellular dynamics and lineage differentiation events. Here, we used scRNA-seq to profile >35'000 endocrine lineage cells and create a molecular roadmap of the secondary stage of endocrine cell generation in pancreatic development in mice (E12.5-15.5). To enrich endocrine progenitors, which are marked by short and transient expression of the transcription factor Neurogenin 3 (Ngn3), we used a novel (Ngn3)-Venus fusion (NVF) reporter mouse line. We systematically delineated stage-dependent heterogeneity in endocrine differentiation steps and lineage relationships to describe gene expression dynamics along endocrine differentiation paths and the molecular programs that define lineage restriction towards specific endocrine subtypes. In our transcriptional map, we identified *Ngn3*<sup>low</sup> endocrine progenitors within multipotent progenitor cell, tip, trunk and ductal clusters that differentiate into *Ngn3*<sup>high</sup> endocrine precursors, endocrine lineage and finally hormone<sup>+</sup> endocrine subtypes at unprecedented resolution. We uncovered putative markers and gene programs that change along the differentiation trajectories and thus likely govern endocrine cell induction, specification and lineage allocation. This included 58 signature genes that were transiently expressed in endocrine progenitors similar to Ngn3. Moreover, we found that endocrine precursor states are heterogenous and show stage-dependent gene expression, which we predict to define lineage trajectories towards different endocrine cell fates.

Taken together, we provide a high-resolution transcriptional map of endocrinogenesis in developing mice, which is made available for future studies. This comprehensive map enabled us to untangle the stepwise lineage restriction from progenitors to hormone<sup>+</sup> endocrine subtypes and its underlying gene expression programs. It is also a resource for future efforts that focus on endocrine differentiation pathways or aim at developing methods to reliably infer

cellular trajectories and fate and their regulatory factors from time-resolved single-cell data of a developing tissue.

### **My contribution**

- I. All data was generated by co-authors.
- II. I preprocessed all raw scRNA-seq data of the manuscript as described in section 2.2.
- III. I analyzed all scRNA-seq data of the manuscript with the support of Leander Dony.
  - A. To describe the different cellular populations established during pancreatic development I annotated and characterized >35'000 pancreatic epithelial cells isolated from four subsequent days of development, which span the spectrum from multi and bipotent progenitors to differentiated endocrine, acinar and ductal cells.
  - B. To describe regulators and timing of cell lineage restriction and fate allocation I built a lineage model inferring cell relationships using PAGA and compared composition and signatures of the populations over time.
  - C. To characterize transient endocrine progenitor and precursor states and identify their molecular signatures and markers I performed trajectory inference and differential expression analysis.
  - D. To investigate the timing of endocrine specification and whether endocrine precursors are heterogenous and primed towards a specific fate, I used RNA velocity inference to predict cellular fates. I further validated the suggested stepwise specification by characterizing the profiles of early and late precursors and their relationships with unipotent cells differentiating towards specific endocrine cell types. This was one of the first applications of the scVelo tool in a publication.
- IV. I interpreted the results, wrote and edited the manuscript and designed the figures for all parts that include the scRNA-seq data together with Mostafa Bakhti. All other co-authors contributed to data generation, analysis and writing for other parts of the manuscript and to the discussion and reviewing of the final manuscript.

### 3.1.3. Targeted pharmacological therapy restores $\beta$ -cell function for diabetes remission

Stephan Sachs\*, Aimée Bastidas-Ponce\*, **Sophie Tritschler\***, Mostafa Bakhti, Anika Böttcher, Miguel A Sánchez-Garrido, Marta Tarquis-Medina, Maximilian Kleinert, Katrin Fischer, Sigrid Jall, Alexandra Harger, Erik Bader, Sara Roscioni, Siegfried Ussar, Annette Feuchtinger, Burcak Yesildag, Aparna Neelakandhan, Christine B Jensen, Marion Cornu, Bin Yang, Brian Finan, Richard D DiMarchi, Matthias H Tschöp, Fabian J Theis, Susanna M Hofmann, Timo D Müller, Heiko Lickert, **Targeted pharmacological therapy restores  $\beta$ -cell function for diabetes remission**. *Nature Metabolism*, 2, 192–209 (2020)

(See also publication <sup>3</sup> in the bibliography and in Appendix B)

#### Summary

Loss and dysfunction of insulin-secreting  $\beta$ -cells are hallmarks of diabetes. It has been proposed that  $\beta$ -cells become dysfunctional because they dedifferentiate to an immature-like state. Currently, it remains unclear whether pharmacological intervention can effectively target these dedifferentiated  $\beta$ -cells to halt disease progression or promote diabetes remission. In this study, we leveraged single-cell transcriptomics to delineate  $\beta$ -cell failure and response to pharmacological treatment in pancreatic islets of mice. In a model of chemical induced diabetes (multiple low doses Streptozotocin (mSTZ) induced  $\beta$ -cell ablation) we observed that upon long-term hyperglycemia  $\beta$ -cells persist but dedifferentiated to a dysfunctional, immature-like state. Insulin treatment as well as targeted delivery of estrogen by Glucagon-like peptide-1 (GLP-1-estrogen conjugate) restored  $\beta$ -cell maturation and function. When both treatments were combined, the required daily insulin was reduced by 60% and more  $\beta$ -cells survived and regained function. The single-cell profiles of healthy, dedifferentiated and treated  $\beta$ -cells revealed novel markers and molecular processes of  $\beta$ -cell dedifferentiation and indicated mechanisms of drug action: Insulin activates the insulin pathway in  $\beta$ -cells, while estrogen is delivered selectively to  $\beta$ -cells, activates the endoplasmic-reticulum-associated protein degradation system and further protects  $\beta$ -cells from cytokine-induced dysfunction. Moreover, the *in silico* reconstructed trajectories and cellular dynamics predict that  $\beta$ -cells de- and re-differentiated in response to pathological factors and treatment and other sources such as other islet cells or putative endocrine progenitors do not or only minimally contribute to the  $\beta$ -cell populations in our diabetic model. This was, in part, further supported as the endocrine progenitor marker Ngn3 was not induced in islets of mSTZ-diabetic and compound treated mice. Together, this study suggested that  $\beta$ -cells dedifferentiate in hyperglycemic conditions and that functional  $\beta$ -cell mass can be regenerated from dedifferentiated  $\beta$ -cells in mice, and it revealed targetable molecular pathways as well as pharmacological entry points to do so. This positions dedifferentiated  $\beta$ -cells as a viable target to halt or revert diabetes.



## My contribution

- I. All data was generated by co-authors.
- II. I preprocessed all raw data of the manuscript as described in *section 2.2*.
- III. I analyzed all scRNA-seq data of the manuscript.
  - A. To investigate the effect of long-term hyperglycemia and beta-cell ablation (mSTZ-treatment) on islet cells I characterized expression profiles and composition of islet cells of healthy and mSTZ-treated mice. Using trajectory inference, I found that cells transition from mature functional to immature to dysfunctional beta-cells suggesting beta cells dedifferentiate. I confirmed that the described maturation trajectory aligns with postnatal beta cell maturation using a public dataset.
  - B. To describe properties and identify novel markers of dysfunctional, dedifferentiated beta cells as well as other diabetic islet cell types I performed differential expression analysis.
  - C. To identify mechanisms underlying pharmacological treatment I characterized and annotated islet cells of different treatment groups. I used composition analysis, trajectory inference and differential expression analysis to describe beta cell redifferentiation states and trajectories, and potential molecular programs induced by treatments.
  - D. To further confirm the origin and fate redifferentiated beta cells I used PAGA analysis and RNA velocity inference and predicted no or only minor contribution of other cell types or progenitors.
- IV. I interpreted the results, wrote and edited the manuscript and designed the figures for all parts that include the scRNA-seq data together with Stephan Sachs and Aimée Bastidas-Ponce as well as overall conclusions. All other co-authors contributed to data generation and analysis for parts of the manuscript and to the discussion and reviewing of the final manuscript.

### 3.1.4. Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice

Alexandra Aliluev\*, **Sophie Tritschler\***, Michael Sterr, Lena Oppenländer, Julia Hinterdobler, Tobias Greisle, Martin Irmeler, Johannes Beckers, Na Sun, Axel Walch, Kerstin Stemmer, Alida Kindt, Jan Krumsiek, Matthias H Tschöp, Malte D Luecken, Fabian J Theis, Heiko Lickert, Anika Böttcher, **Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice**. *Nature Metabolism*, 3, 1202–1216 (2021)

(See also publication <sup>4</sup> in the bibliography and in Appendix C)

#### Summary

Dysregulated nutrient uptake and hormone secretion in the gut contribute to an energy imbalance that leads to obesity and increased susceptibility to type 2 diabetes and colorectal cancer. It has been proposed that an obesogenic diet affects intestinal stem cells (ISCs) and alters their identity and fate, which contributes to this functional maladaptation of the gut. However, it remains unclear which mechanisms underlie ISC alteration and how this affects the composition and function of differentiated cells in the gut. In this study, we used single-cell transcriptomics, bulk metabolomic profiling, and lineage labeling and tracing of the small intestinal epithelium in a model of diet-induced obesity and pre-diabetes to investigate how the endocrine system of the gut adapts early in disease and contributes to disease progression. We confirmed that cellular composition, gene expression programs and hormone secretion change in enteroendocrine cells and potentially contribute to systemic deregulation and compensatory hyperinsulinemia in pre-diabetic mice. Besides the endocrine compartment we found that the gut remodels and villi are longer in high-fat diet mice to increase nutrient absorption and accumulate fat. ISCs and early progenitor cells are hyperproliferative and change their regional identity, which alters function and zonation of enterocytes towards a proximal phenotype of increased carbohydrate and fat uptake, and changes cell type composition in the villi. These changes were linked to increased fatty acid synthesis, Ppar signaling and the *Insr-Igf1r-Akt* pathway. Taken together, this study described cellular and molecular processes, which are putative targets to counteract enteroendocrine dysfunction and increased intestinal food absorption. These may provide novel therapeutic options to halt obesity and prevent progression to diabetes or other associated complications.

#### My contribution

- I. All data was generated by co-authors. I contributed to the generation of the scRNA-seq data (10x protocol).
- II. I preprocessed all raw data of the manuscript as described in *section 2.2*.
- III. I analyzed all scRNA-seq and single-cell qPCR data of the manuscript.

- A. To investigate the effect of an obesogenic diet on intestinal lineage allocation I annotated and characterized epithelial cells of the small intestine crypts of healthy and obese mice and used differential expression, compositional analysis and PAGA to describe differences between the two conditions. I annotated cell types as well as cell cycle phases and regional location of the cells.
  - B. To identify changes in enteroendocrine lineage formation and composition in obese mice I thoroughly characterized the profiles of the different endocrine cell states, described transient and specific markers, performed compositional analysis and built a lineage model using PAGA and RNA velocity inference.
  - C. To identify mechanisms underlying altered lineage allocation I described differences in gene expression and proliferation in stem cells and progenitor populations.
  - D. To confirm identified pathways through a more sensitive method I analyzed single-cell qPCR data of crypt stem cells and progenitors.
  - E. To confirm that alterations in the crypt translate to the villi I annotated and characterized single cells of the villus.
- IV. I reviewed the literature, interpreted the results, wrote and edited the manuscript and designed the figures together with Alexandra Aliluev and Anika Böttcher for all parts that include the scRNA-seq data as well as the abstract, general introduction and discussion. All other co-authors contributed to data generation, analysis and writing for parts of the manuscript and to the discussion and reviewing of the final manuscript

## 3.2. Review Articles

Further, I contributed to writing two articles, which review the insights of early single-cell studies into pancreatic heterogeneity and concepts and limitations of trajectory inference from single-cell data.

### 3.2.1. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas

**Sophie Tritschler**, Fabian J. Theis, Heiko Lickert, Anika Böttcher,  
**Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas**, *Molecular Metabolism*, Volume 6, Issue 9, Pages 974-990 (2017)

(See also publication <sup>5</sup> in the bibliography and in Appendix D)

#### Summary

Lost or dysfunctional  $\beta$ -cells have become the primary focus for regenerative therapies to halt or revert diabetes. To advance these efforts multiple aspects are crucial. First, it requires us to systematically characterize  $\beta$ -cells as well as their potential cellular sources in healthy and diabetic patients to identify cells with regenerative potential. Second, we need to delineate the molecular programs that define and regulate  $\beta$ -cell differentiation, maturation, maintenance and dysfunction to be able to (re-)generate functional  $\beta$ -cells. Lastly, we must be able to translate findings from model systems to humans. Single-cell omics provide promising approaches to decipher the heterogeneity and plasticity of the pancreas. Here, we reviewed the literature and findings from emerging single-cell technologies including single-cell RNA-seq, single-cell mass cytometry, and flow cytometry that may facilitate aspects of endogenous  $\beta$ -cell regeneration. We summarized and discussed how single-cell data of human and mouse models aided our knowledge of how  $\beta$ -cells proliferate and mature postnatally, of endocrine heterogeneity and rare subtypes, and their function and impact on homeostasis and disease as well as of potential cell sources and molecular targets for regeneration. We concluded that single-cell data offers unprecedented resolution to shed light on these complex biological processes and may reveal cellular and molecular targets to regenerate functional  $\beta$ -cells *in vivo*, however most analyses and insights remain descriptive. To elucidate mechanisms, we suggest that we should, as a next step, aim at integrating the different layers of single-cell omics, adding temporal and spatial information and complementing these descriptive analyses with functional and mechanistic studies.

#### My contribution

- I conceptualized the review together with Anika Böttcher and Heiko Lickert.

- II. I reviewed the literature, wrote and edited the manuscript together with Anika Böttcher, and I designed all the figures. All co-authors contributed to the discussion and reviewing of the final manuscript.

### 3.2.2. Concepts and limitations for learning developmental trajectories from single cell genomics

**Sophie Tritschler\***, Maren Büttner\*, David S. Fischer, Marius Lange, Volker Bergen, Heiko Lickert, Fabian J. Theis, **Concepts and limitations for learning developmental trajectories from single cell genomics**. *Development* 15; 146 (12): dev170506 (2019)

(See also publication <sup>6</sup> in the bibliography)

#### **Summary**

Single-cell transcriptomic data is widely used to study developmental processes and cellular lineage relationships. Single-cell technologies continuously evolve to improve through-put and lower costs, and with that the scale of the data and number of cells that are routinely sequenced increases fast. However, it remains a challenge to analyze and interpret these large-scale datasets to reliably derive valuable biological insights. Typically, trajectory and tree models are used to describe differentiation processes and lineage hierarchies and elucidate their underlying mechanisms *in silico*. Many methods have been developed to infer such linear or branching trajectories, but to correctly apply them, it is required to understand their assumptions and limitations. Here, we described the mathematical concepts of trajectory inference from single-cell transcriptomic data and how it can be applied to study continuous processes in developmental systems or disease. We summarized the potential and key limitations of simple, similarity-based trajectory inference methods, which reconstruct a static, undirected view of a continuous process from gene expression profiles alone. We then review how additional information can be used to better constrain these models and predict cellular dynamics. Lastly, we discussed examples of how such continuous embeddings and lineage trees can be used to model disease. We concluded that trajectory models of single-cell transcriptomic data are a versatile tool that enables us to chart cell lineages of tissues and even whole, complex animals. While today these models are mostly descriptive and phenomenological, we envisioned that with increasing data size, the integration of perturbation data and other orthogonal data types such as spatial information or population size dynamics they will become even more powerful. They will allow us to predict cellular dynamics, and their underlying gene regulatory events, and infer how other layers like the cellular environment contribute to the regulation of developmental processes.

#### **My contribution**

- I. I conceptualized the review together with David S. Fischer and Fabian J. Theis.
- II. I reviewed the literature, designed the figures and wrote and edited the manuscript together with Maren Büttner. All the other co-authors contributed on writing one paragraph as well as the discussion and reviewing of the final manuscript.

## 4. Discussion and outlook

The development of regenerative treatments to regain endogenous endocrine control of blood glucose levels in diabetic patients depends on our understanding of mechanisms of cell differentiation, function and dysfunction in endocrine tissue. Single-cell transcriptomic data provides us unprecedented insights into such complex biological systems. It allows us to map and characterize heterogeneous cell populations and their dynamics, and with that extract cellular trajectories that predict a cell's origin and fate, the gene programs that define a cell's phenotype, and the cellular composition that determines tissue function. Such descriptions promise to reveal mechanisms that can be targeted by therapeutics to efficiently differentiate and or regenerate functional endocrine cell populations.

### 4.1. Summary

This cumulative thesis presents different applications of single-cell transcriptomics and computational analysis that combined with functional and biochemical data describe how endocrine cell populations are established and behave in health and diabetes. My key scientific contributions are twofold. Firstly, I discussed in two review articles how single-cell transcriptomics and computation has empowered research on endocrine heterogeneity and plasticity in the pancreas and, in general, to learn cellular trajectories and lineage hierarchies in developing or diseased tissues. Secondly, I thoroughly analyzed single-cell transcriptomic data in the context of four studies describing endocrine cell populations during specific windows of life of the endocrine pancreas and intestine. I built a cross-species map of the adult, healthy endocrine pancreas and an analytical framework to compare cellular states and their gene programs. This represents a comprehensive resource to bridge insights across human and two clinically relevant species and positioned pigs as a promising model to study human endocrine cell heterogeneity and function. Moreover, I reconstructed a lineage model for endocrine cells during pancreatic development in mice, which described stepwise cellular differentiation trajectories and predicted regulators of cell identity. Finally, to study endocrine dysfunction, I thoroughly charted the endocrine cell populations of two disease models. This revealed that in diabetic mice pancreatic b-cells dedifferentiate to a dysfunctional state, which can be re-differentiated upon pharmacological treatment. In addition, it proposed mechanisms of intestinal maladaptation and enteroendocrine dysfunction in obese mice, which may promote progression to diabetes.

Together, these studies shed light on the control and molecular basis of endocrine cell identity and (dys)function in multiple species: they describe biomarkers and gene programs of differentiating, functional and dysfunctional states and how cells transition between such

states; and they propose mechanisms to (re-)differentiate and or (re-)establish functional endocrine cells. Moreover, the single-cell maps and source code were made easily accessible, which facilitates other researchers to combine these resources with their own data and further build on this work. The novel findings together with the readily available cellular maps are valuable assets that may ultimately contribute to developing regenerative therapeutics that improve endocrine function in obese or diabetic patients.

## 4.2. Advancing our understanding of endocrine disease biology

Although our single-cell transcriptomic analyses thoroughly describe cellular and transcriptional components of endocrine systems, they only capture aspects of endocrine disease biology. Firstly, we only consider a single molecular layer (gene expression), which is not enough to fully characterize a molecular cell state, catalog cellular phenotypes and elucidate heterogeneity and plasticity. Secondly, our findings were centered on “isolated” cells and cell types of a single time point and ignore how cell types act collectively with their local environment to direct tissue function over time. Lastly, the analyses were focused on identifying statistical patterns and were therefore still mostly descriptive with minimal mechanistic and predictive insights. Technological and computational advancements and integration of other data types have since opened remarkable opportunities to advance single to multimodal cell state definitions, cellular to tissue-wide models of cell function, and descriptive to mechanistic roles of genes, programs or cell states in biological systems (**Figure 6**).

### 4.2.1. Towards multimodal models of cellular function and heterogeneity

Different approaches can help to improve our models of endocrine cell populations and cellular (dys-)function, if we i) catalog cells combining all molecular layers, and ii) consider that multiple factors combine to create a cell’s molecular identity: environmental stimuli (e.g., pathological factors or nutrient exposure), temporal processes (e.g., treatment or developmental history), and spatial context (e.g., cellular neighbors or morphogen gradients)<sup>167</sup>.

#### **Combining transcriptomics with other molecular layers**

While single-cell transcriptomics has long been at the forefront, the application of other single-cell omics<sup>168–171</sup>, multi-omics or spatial technologies<sup>172</sup> is rapidly growing and provides additional insights on islet biology<sup>173</sup>. Epigenomic and proteomic profiles add other layers to cell state definitions, and their potential to respond to environmental stimuli (**Figure 6**). For example, integration of epigenetic states may help to identify the most upstream causal factors that direct transcriptionally similar endocrine progenitors to differentiate into different



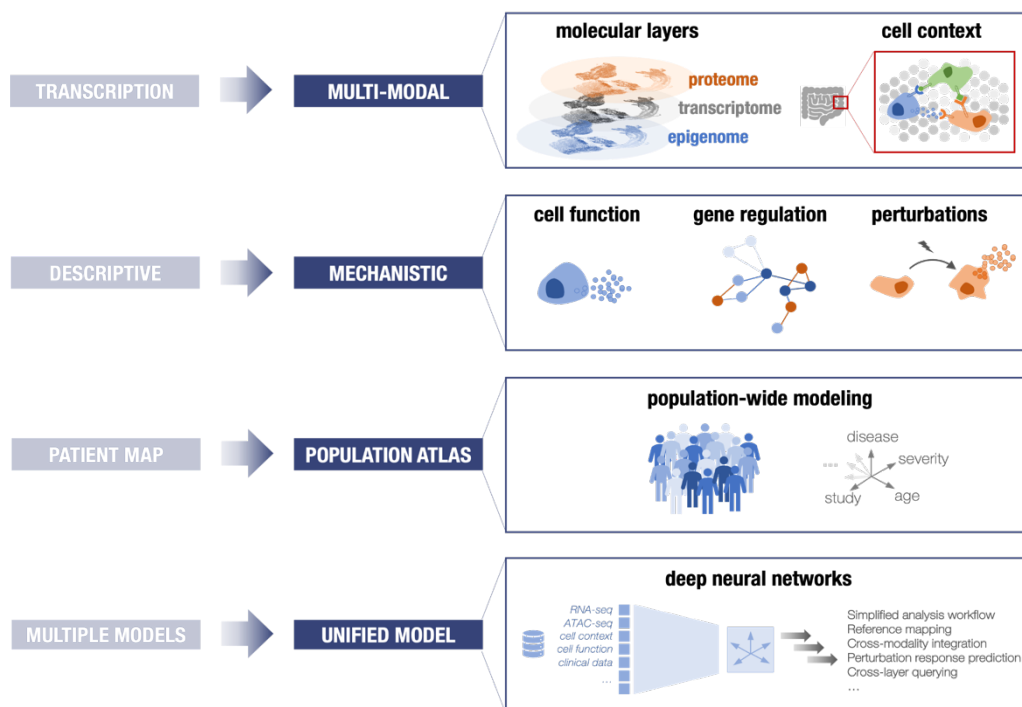
lineages<sup>174,175</sup>. The epigenome also hints at cellular plasticity, and thus may indicate how susceptible endocrine cells are to respond to stress, reprogramming or treatment<sup>176–178</sup>. Proteomic approaches can elucidate post-translational modifications, which are important components of signaling. A simple relevant example of using proteomics would be to measure the conversion of prohormones to hormones and its deregulation in disease states. Moreover, as for epigenomics, cell states can be characterized in more detail when protein expression is included<sup>169</sup>. Specifically, receptor genes, which are important therapeutic targets, are often only lowly expressed and thus not well captured in transcriptional data. Although single-cell proteomic approaches are currently still limited to a small set of proteins or low in resolution, we envision that with further technological advancements proteomics will be a powerful complement to other single-cell omics to confirm and improve cell catalogs, as recently exemplified in large studies of pancreatic islets of T1D<sup>104</sup> or TD2<sup>179</sup> patients.

### **Superimposing additional data to validate cell trajectories**

To reconstruct temporal processes and predict fate and origin of endocrine cell states, we learned cellular trajectories and lineage relationships from transcriptional similarity of cells (pseudotemporal ordering) and predicted directionality and future states from inferred gene expression dynamics (RNA velocity)<sup>6</sup>. These analyses allowed us to track cells and gene expression during cellular transitions and identify putative markers of temporal states (for example progenitors, precursors, immature and mature states). However, these findings remain largely observational, and generate hypotheses but cannot infer causal links<sup>6</sup>. Incorporating experimental time, i.e., generating time-resolved data along disease progression, or using single-cell lineage tracing approaches can further increase the power of trajectory inference frameworks and validate predictions from transcriptional snapshots alone<sup>6,180,181</sup>. For example, time-resolved data allowed us to link the identified subpopulations in endocrine precursors to cell fates as the populations emerged in a stage-dependent manner and preferentially differentiate into alpha or  $\beta$ -cells<sup>2</sup>. Similarly, time course data could help to better resolve early and late events in  $\beta$ -cell failure and differentiate protective from regenerative drug action in our diabetes model<sup>3</sup>. Genetic labeling or lineage tracing of single cells to reconstruct lineage trees or clonal relationships are required to validate that  $\beta$ -cells de- and redifferentiate. In addition, such experiments could indicate whether subpopulations with enhanced survival or regenerative properties exist in humans as it was described in T1D mouse models<sup>46,182</sup>. Finally, it could improve our understanding of enteroendocrine plasticity and how much stem or progenitor cell identity and priming contributes to the functional heterogeneity of enteroendocrine cells in disease.

## Adding cellular and spatial context

Lastly, novel technologies and computational methods have been developed to integrate spatial context into models of cell function<sup>172,183,184</sup> (**Figure 6**). Spatial information connects a cell's molecular phenotype to its neighboring cells and the environment. This elucidates how cells are organized in a multicellular tissue and how their neighboring cells and local environment influence their fate decisions and function. Reconstructing spatial axes from previously identified spatial markers<sup>124</sup> allowed us to project functional programs of enterocytes to intestinal regions and villi zones<sup>4</sup>. Similarly, spatial context, either measured or reconstructed from spatio-molecular features, could confirm the physical position of endocrine progenitors when the epithelium develops, and the contribution of tip cells to EP formation as indicated by our single-cell transcriptomic data<sup>2</sup> and others<sup>109</sup>. It also may further elucidate how islet architecture contributes to cellular heterogeneity and changes in disease and identify or confirm neogenic and pathological niches. Moving beyond insights in spatial organization and heterogeneity, novel methods leverage spatial information to integrate cell-cell interactions and cell-intrinsic expression to improve cell communication inference and model intercellular communication in cellular niches<sup>185</sup>. This is a step towards unbiased models of tissue organization and how it impacts cellular function. For example, this could elucidate how niche signals contribute to changed ISC identity and fate allocation in the intestinal crypts.



**Figure 6.** Data generation and modeling options to further advance our understanding of endocrine cell populations and their relevance in disease.

#### 4.2.2. From genes to mechanisms

In this work we derived hypotheses on mechanisms that control cellular phenotypes through knowledge enrichment (e.g., pathway enrichment) and literature review of the rich transcriptional profiles. However, while correlated or coordinated gene expression is suggestive, it does not necessarily imply causal links. Combining single-cell transcriptomic data with functional measurements, other omics data or interventions provides opportunities for systematic dissection and better interpretation of regulatory mechanisms and targetable pathways<sup>66,67</sup> (**Figure 6**). With paired transcriptional and functional single-cell data, such as Patch-seq measurements<sup>101</sup>, gene expression changes can be directly linked to different functional activity of a cell. This does not only allow us to establish less ambiguous links of gene programs to function, but also to extrapolate functional activity to samples where functional single-cell measurements are not applicable, for example in frozen tissue or large-scale references as we did in our study<sup>1</sup>. Direct interventions can demonstrate that a gene or process is causative. CRISPR-based technologies for genetic screens in single-cells hold promise to identify or validate novel molecular targets and reconstruct regulatory relationships<sup>186–188</sup>. Epigenomic and proteomic data provide complementary information to further constrain these gene-gene relationships to context specific networks of gene regulation. Finally, perturbation data, even if not directly physiologically relevant, may be leveraged to approximate the boundaries of the transcriptional space of a cell system and predict possible gene responses<sup>78,189</sup>. I envision that together this will improve our ability to infer causal processes from single-cell observations and link differential expression to its effect on a cell's phenotype as computational methods mature.

Taken together, while these novel approaches will not be able to replace experimental validation studies, they will help to generate more powerful hypotheses and predictions on cell and tissue behaviors and nominate causal processes that drive cell state transitions. This will hopefully reveal novel disease biology and clinical targets and reduce the overall cost and time of experimental validation needed.

#### 4.3. Advancing disease relevance with large-scale atlases

The scale of individual single-cell datasets of pancreatic islets and the intestinal epithelium described here and in other studies is still too limited to capture population-wide variability. Moreover, technical confounders and inconsistent cell state identification and annotation make it difficult to relate findings between study or build on knowledge of previous studies beyond

comparing the composition of major cell types or lists of genes and pathways. Cell atlases integrate multiple single-cell maps (datasets) to a harmonized model (**Figure 6**). Such atlases have the potential to cover a broader range of variables, which are expected to impact cell identity and function<sup>190,191</sup>. These include clinical and demographic patient features, but also longitudinal datasets, different animal models, disease conditions or perturbations. Our reference map and analyses of healthy human pancreas are a first step towards generating an islet atlas with a consistent state annotation that captures normal patient variability<sup>1</sup>. We could confirm that the identified  $\beta$ -cell states are robustly detected across study and that age or sex do not strongly contribute to the observed donor variation when we mapped  $\beta$ -cells of 9 publicly available studies (54 human donors) to our reference. Moreover, we adhered to FAIR data guiding principles<sup>166</sup>, which not only facilitates querying of this resource but also integration in future atlases. Similarly, a comprehensive reference atlas of mouse islets provides a resource to model  $\beta$ -cell states across age, models and treatment<sup>114</sup>. I anticipate that soon cell atlases of human islets will more comprehensively chart the cellular landscape of diabetic disease populations. Research consortia enable large-scale sourcing of tissues and data generation<sup>192–194</sup>, single-cell technologies are more robust and computational tools for data integration and reference mapping more advanced. Moreover, bulk RNA-seq datasets from large patient cohorts can be used to complement single-cell atlases and further validate their findings. It will remain a challenge to obtain longitudinal or early disease onset samples, as pancreatic biopsies cannot be safely obtained from living donors and often  $\beta$ -cell dysfunction and loss has already progressed when patients are clinically diagnosed with diabetes. Still, large-scale atlases will better represent patient and disease diversity and reduce technical noise and thus identify cell states and gene programs more relevant to disease.

#### 4.4. Towards unified foundational models of biology

As the complexity and scale of data sets grow and different modalities are combined, computational methods to simplify analyses and integrate data and learnings become ever more important<sup>195,196</sup>. Latent space embeddings, as learned by deep neural networks, have been introduced as a powerful new approach for the increasingly complex tasks (**Figure 6**). Latent representations can be used for most steps in single-cell analysis workflows including data correction, data integration, clustering and differential expression analysis<sup>197,198</sup>. Moreover, latent spaces enable to complement exploratory, unsupervised analyses with supervised predictions and even interpolating or creating data through generative models. Examples of predictive models built on latent spaces include query-to-reference mapping approaches, which transfer learnings from reference to query and thus enable creation and

continuous extension of harmonized large-scale, population-wide and multimodal atlases<sup>81,154,156,199</sup>. Such approaches applied for multi-layer references will not only enable cell label transfer, but also prediction of other layers for single-layer queries including cross-modality or cross-species predictions. Generative models can be used for example for response predictions of unseen perturbations like drugs or genetic perturbations<sup>78,200–202</sup>. Lastly, the breakthrough of large language and foundation models, which connect vast amount of data in an unsupervised fashion, has enabled many novel, more powerful generative applications in other fields. With many efforts to make omics data accessible and reusable<sup>10,65,203,204</sup>, training of such models becomes possible. With the growing data size and modeling advancements such foundation models hold the promise to enable to not only connect molecular but also other data modalities including for example phenotypic, morphological, spatial measurements<sup>195</sup>. This will ultimately provide a unified basis model for many separate applications as well as multi-layered biological answers for simple queries<sup>196</sup>. Such orthogonal and richer data will hopefully also bring us one step closer to modeling gene regulation and mechanisms. Interpretability of deep latent spaces, i.e. linking latent features to genes and gene programs, remains a key challenge. Most current approaches use literature curated databases as priors<sup>205,206</sup>, and therefore do not consider all available information in the data and insights on gene regulation remains restricted and biased to prior knowledge. Thus, as these predictive and generative deep models become more robust and are trained on even larger datasets, they will not replace experiments, but efficiently guide experimental design and with that massively accelerate biological discoveries.

Overall, this thesis provides a deep perspective of endocrine cell populations in the pancreas and intestine during critical windows of life and disease. Through analysis of single-cell data generated to address specific questions in humans and animal models, it sheds light on cells and molecular programs that are involved in development and progression of diabetes and are potential targets for regenerative treatment strategies. In the future, as single-cell omics and computational models continue to advance we will be able to complement such exploratory analyses with more predictive insights: we will integrate these findings into population-wide, multi-modal atlases and latent space embeddings to describe cellular variation across many covariates and create a more complete catalog of endocrine cells. With that we will be able to connect learnings from many datasets and predict mechanisms of disease and drug action to guide future endeavors on the road towards successful regenerative therapies for diabetes.

# Bibliography

1. Tritschler, S. *et al.* A transcriptional cross species map of pancreatic islet cells. *Mol Metab* **66**, 101595 (2022).
2. Bastidas-Ponce, A., Tritschler, S. *et al.* Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. *Development* **146**, (2019).
3. Sachs, S., Bastidas-Ponce, A., Tritschler, S. *et al.* Targeted pharmacological therapy restores  $\beta$ -cell function for diabetes remission. *Nat Metab* **2**, 192–209 (2020).
4. Aliluev, A., Tritschler, S. *et al.* Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice. *Nat Metab* **3**, 1202–1216 (2021).
5. Tritschler, S., Theis, F. J., Lickert, H. & Böttcher, A. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas. *Mol Metab* **6**, 974–990 (2017).
6. Tritschler, S. *et al.* Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, (2019).
7. Böttcher, A. *et al.* Non-canonical Wnt/PCP signalling regulates intestinal stem cell lineage priming towards enteroendocrine and Paneth cell fates. *Nat. Cell Biol.* **23**, 23–31 (2021).
8. Angerer, P. *et al.* Single cells make big data: New challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology* vol. 4 85–91 Preprint at <https://doi.org/10.1016/j.coisb.2017.07.004> (2017).
9. Bakhti, M. *et al.* Establishment of a high-resolution 3D modeling system for studying pancreatic epithelial cell biology in vitro. *Mol Metab* **30**, 16–29 (2019).
10. Fischer, D. S. *et al.* Sfaira accelerates data and model reuse in single cell genomics. *Genome Biol.* **22**, 248 (2021).
11. Salinno, C. *et al.* CD81 marks immature and dedifferentiated pancreatic  $\beta$ -cells. *Mol Metab* **49**, 101188 (2021).
12. Bakhti, M. *et al.* Synaptotagmin-13 orchestrates pancreatic endocrine cell egression and

- islet morphogenesis. *Nat. Commun.* **13**, 4540 (2022).
13. Benakis, C. *et al.* T cells modulate the microglial response to brain ischemia. *Elife* **11**, (2022).
  14. Bommer, C. *et al.* The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study. *The Lancet Diabetes & Endocrinology* vol. 5 423–430 Preprint at [https://doi.org/10.1016/s2213-8587\(17\)30097-9](https://doi.org/10.1016/s2213-8587(17)30097-9) (2017).
  15. Nauck, M. A., Quast, D. R., Wefers, J. & Meier, J. J. GLP-1 receptor agonists in the treatment of type 2 diabetes - state-of-the-art. *Mol Metab* **46**, 101102 (2021).
  16. Müller, T. D. *et al.* Glucagon-like peptide 1 (GLP-1). *Mol Metab* **30**, 72–130 (2019).
  17. Bailey, C. J., Flatt, P. R. & Conlon, J. M. An update on peptide-based therapies for type 2 diabetes and obesity. *Peptides* **161**, 170939 (2023).
  18. Hao, S., Umpierrez, G. E. & Vellanki, P. Intervention with Therapeutic Agents, Understanding the Path to Remission to Type 2 Diabetes: Part 2. *Endocrinol. Metab. Clin. North Am.* **52**, 39–47 (2023).
  19. Gomez-Peralta, F. & Abreu, C. Profile of semaglutide in the management of type 2 diabetes: design, development, and place in therapy. *Drug Des. Devel. Ther.* **13**, 731–738 (2019).
  20. Choi, M. & Burch, M. Bariatric Surgery versus Intensive Medical Therapy in Obese Patients with Diabetes. *Oxford Medicine Online* Preprint at <https://doi.org/10.1093/med/9780199384075.003.0039> (2017).
  21. Docherty, F. M. & Sussel, L. Islet Regeneration: Endogenous and Exogenous Approaches. *Int. J. Mol. Sci.* **22**, (2021).
  22. Basile, G. *et al.* Emerging diabetes therapies: Bringing back the  $\beta$ -cells. *Mol Metab* **60**, 101477 (2022).
  23. Soldovieri, L. *et al.* An update on pancreatic regeneration mechanisms: Searching for paths to a cure for type 2 diabetes. *Mol Metab* **74**, 101754 (2023).
  24. Tschöp, M. & DiMarchi, R. Single-Molecule Combinatorial Therapeutics for Treating Obesity and Diabetes. *Diabetes* **66**, 1766–1769 (2017).

25. Capozzi, M. E., DiMarchi, R. D., Tschöp, M. H., Finan, B. & Campbell, J. E. Targeting the Incretin/Glucagon System With Triagonists to Treat Diabetes. *Endocr. Rev.* **39**, 719–738 (2018).
26. Atanga, R., Singh, V. & In, J. G. Intestinal Enteroendocrine Cells: Present and Future Druggable Targets. *Int. J. Mol. Sci.* **24**, (2023).
27. Nolan, C. J. & Prentki, M. Insulin resistance and insulin hypersecretion in the metabolic syndrome and type 2 diabetes: Time for a conceptual framework shift. *Diab. Vasc. Dis. Res.* **16**, 118–127 (2019).
28. American Diabetes Association Professional Practice Committee *et al.* 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2022. *Diabetes Care* **45**, S125–S143 (2022).
29. Shapiro, A. M. *et al.* Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N. Engl. J. Med.* **343**, 230–238 (2000).
30. Coe, T. M., Markmann, J. F. & Rickert, C. G. Current status of porcine islet xenotransplantation. *Curr. Opin. Organ Transplant.* **25**, 449–456 (2020).
31. Renner, S. *et al.* Comparative aspects of rodent and nonrodent animal models for mechanistic and translational diabetes research. *Theriogenology* **86**, 406–421 (2016).
32. Renner, S. *et al.* Porcine models for studying complications and organ crosstalk in diabetes mellitus. *Cell Tissue Res.* **380**, 341–378 (2020).
33. Ludwig, B. *et al.* Favorable outcome of experimental islet xenotransplantation without immunosuppression in a nonhuman primate model of diabetes. *Proceedings of the National Academy of Sciences* vol. 114 11745–11750 Preprint at <https://doi.org/10.1073/pnas.1708420114> (2017).
34. Pagliuca, F. W. *et al.* Generation of Functional Human Pancreatic  $\beta$  Cells In Vitro. *Cell* vol. 159 428–439 Preprint at <https://doi.org/10.1016/j.cell.2014.09.040> (2014).
35. Millman, J. R. *et al.* Generation of stem cell-derived  $\beta$ -cells from patients with type 1 diabetes. *Nat. Commun.* **7**, 11463 (2016).



36. Balboa, D., Saarimäki-Vire, J. & Otonkoski, T. Concise Review: Human Pluripotent Stem Cells for the Modeling of Pancreatic  $\beta$ -Cell Pathology. *Stem Cells* **37**, 33–41 (2019).
37. Rezaia, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
38. Veres, A. *et al.* Charting cellular identity during human in vitro  $\beta$ -cell differentiation. *Nature* vol. 569 368–373 Preprint at <https://doi.org/10.1038/s41586-019-1168-5> (2019).
39. Balboa, D. *et al.* Functional, metabolic and transcriptional maturation of human pancreatic islets derived from stem cells. *Nat. Biotechnol.* **40**, 1042–1055 (2022).
40. Mahaddalkar, P. U. *et al.* Generation of pancreatic  $\beta$  cells from CD177 anterior definitive endoderm. *Nat. Biotechnol.* **38**, 1061–1072 (2020).
41. Russ, H. A. *et al.* Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *EMBO J.* **34**, 1759–1772 (2015).
42. Augsornworawat, P., Maxwell, K. G., Velazco-Cruz, L. & Millman, J. R. Single-cell transcriptome profiling reveals  $\beta$  cell maturation in stem cell-derived islets after transplantation. *Cell Rep.* **34**, 108850 (2021).
43. Jain, C., Ansarullah, Bilekova, S. & Lickert, H. Targeting pancreatic  $\beta$  cells for diabetes treatment. *Nature Metabolism* vol. 4 1097–1108 Preprint at <https://doi.org/10.1038/s42255-022-00618-5> (2022).
44. Zhou, Q. & Melton, D. A. Pancreas regeneration. *Nature* vol. 557 351–358 Preprint at <https://doi.org/10.1038/s41586-018-0088-0> (2018).
45. Bilekova, S., Sachs, S. & Lickert, H. Pharmacological Targeting of Endoplasmic Reticulum Stress in Pancreatic Beta Cells. *Trends in Pharmacological Sciences* vol. 42 85–95 Preprint at <https://doi.org/10.1016/j.tips.2020.11.011> (2021).
46. Thompson, P. J. *et al.* Targeted Elimination of Senescent Beta Cells Prevents Type 1 Diabetes. *Cell Metab.* **29**, 1045–1060.e10 (2019).
47. Evert, A. B. & Franz, M. J. Why Weight Loss Maintenance Is Difficult. *Diabetes Spectr.* **30**, 153–156 (2017).

48. Gribble, F. M. & Reimann, F. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. *Nat. Rev. Endocrinol.* **15**, 226–237 (2019).
49. Yilmaz, Ö. H. *et al.* mTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* **486**, 490–495 (2012).
50. Koliaki, C., Liatis, S., Dalamaga, M. & Kokkinos, A. The Implication of Gut Hormones in the Regulation of Energy Homeostasis and Their Role in the Pathophysiology of Obesity. *Curr. Obes. Rep.* **9**, 255–271 (2020).
51. Beyaz, S. *et al.* High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* vol. 531 53–58 (2016).
52. Mihaylova, M. M. *et al.* Fasting Activates Fatty Acid Oxidation to Enhance Intestinal Stem Cell Function during Homeostasis and Aging. *Cell Stem Cell* **22**, 769–778.e4 (2018).
53. Le Gall, M. *et al.* Intestinal plasticity in response to nutrition and gastrointestinal surgery. *Nutr. Rev.* **77**, 129–143 (2019).
54. Dailey, M. J. Nutrient-induced intestinal adaption and its effect in obesity. *Physiol. Behav.* **136**, 74–78 (2014).
55. Lean, M. E. J. & Malkova, D. Altered gut and adipose tissue hormones in overweight and obese individuals: cause or consequence? *Int. J. Obes.* **40**, 622–632 (2015).
56. Evers, S. S., Sandoval, D. A. & Seeley, R. J. The Physiology and Molecular Underpinnings of the Effects of Bariatric Surgery on Obesity and Diabetes. *Annu. Rev. Physiol.* **79**, 313–334 (2017).
57. Wilkinson, L. *et al.* Effect of semaglutide 2.4 mg once weekly on 10-year type 2 diabetes risk in adults with overweight or obesity. *Obesity* **31**, 2249–2259 (2023).
58. Roscioni, S. S., Migliorini, A., Gegg, M. & Lickert, H. Impact of islet architecture on  $\beta$ -cell heterogeneity, plasticity and function. *Nat. Rev. Endocrinol.* **12**, 695–709 (2016).
59. Pipeleers, D. G. Heterogeneity in pancreatic beta-cell population. *Diabetes* vol. 41 777–781 Preprint at <https://doi.org/10.2337/diabetes.41.7.777> (1992).
60. Gutierrez, G. D., Gromada, J. & Sussel, L. Heterogeneity of the Pancreatic Beta Cell.

*Frontiers in Genetics* vol. 8 Preprint at <https://doi.org/10.3389/fgene.2017.00022> (2017).

61. Beumer, J. *et al.* Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. *Nat. Cell Biol.* **20**, 909–916 (2018).
62. Gehart, H. *et al.* Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell* **176**, 1158–1173.e16 (2019).
63. Beumer, J., Gehart, H. & Clevers, H. Enteroendocrine Dynamics - New Tools Reveal Hormonal Plasticity in the Gut. *Endocr. Rev.* **41**, (2020).
64. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
65. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, (2017).
66. Rood, J. E., Maartens, A., Hupalowska, A., Teichmann, S. A. & Regev, A. Impact of the Human Cell Atlas on medicine. *Nat. Med.* **28**, 2486–2496 (2022).
67. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
68. Rostom, R., Svensson, V., Teichmann, S. A. & Kar, G. Computational approaches for interpreting scRNA-seq data. *FEBS Lett.* **591**, 2213–2225 (2017).
69. Haynes, W. A., Tomczak, A. & Khatri, P. Gene annotation bias impedes biomedical research. *Sci. Rep.* **8**, 1362 (2018).
70. Timmons, J. A., Szkop, K. J. & Gallagher, I. J. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol.* **16**, 186 (2015).
71. Website. <https://data.humancellatlas.org/>,
72. Zappia, L. & Theis, F. J. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol.* **22**, 301 (2021).
73. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* **14**, e1006245 (2018).
74. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

75. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
76. Badia-I-Mompel, P. *et al.* Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet.* **24**, 739–754 (2023).
77. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).
78. Lotfollahi, M., Wolf, F. A. & Theis, F. J. scGen predicts single-cell perturbation responses. *Nat. Methods* **16**, 715–721 (2019).
79. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
80. Kharchenko, P. V. The triumphs and limitations of computational methods for scRNA-seq. *Nat. Methods* **18**, 723–732 (2021).
81. De Donno, C. *et al.* Population-level integration of single-cell datasets enables multi-scale analysis across samples. *Nat. Methods* **20**, 1683–1692 (2023).
82. Gayoso, A. *et al.* Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nat. Methods* **18**, 272–282 (2021).
83. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
84. Sonesson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).
85. Baik, B., Yoon, S. & Nam, D. Benchmarking RNA-seq differential expression analysis methods using spike-in and simulation data. *PLoS One* **15**, e0232271 (2020).
86. Tran, H. T. N. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 12 (2020).
87. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
88. Duò, A., Robinson, M. D. & Sonesson, C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res.* **7**, 1141 (2018).
89. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a

- tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
90. Heumos, L. *et al.* Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023).
  91. Lafzi, A., Moutinho, C., Picelli, S. & Heyn, H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat. Protoc.* **13**, 2742–2757 (2018).
  92. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
  93. Virshup, I. *et al.* The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat. Biotechnol.* **41**, 604–606 (2023).
  94. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
  95. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).
  96. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst* **3**, 385–394.e3 (2016).
  97. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346–360.e4 (2016).
  98. Xin, Y. *et al.* RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab.* **24**, 608–615 (2016).
  99. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* **171**, 321–330.e14 (2017).
  100. Xin, Y. *et al.* Pseudotime Ordering of Single Human  $\beta$ -Cells Reveals States of Insulin Production and Unfolded Protein Response. *Diabetes* **67**, 1783–1794 (2018).
  101. Camunas-Soler, J. *et al.* Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. *Cell Metab.* **31**, 1017–1031.e4 (2020).
  102. Lawlor, N. *et al.* Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Research* vol. 27 208–222 Preprint at <https://doi.org/10.1101/gr.212720.116> (2017).

103. Shrestha, S. *et al.* Combinatorial transcription factor profiles predict mature and functional human islet  $\alpha$  and  $\beta$  cells. *JCI Insight* **6**, (2021).
104. Fasolino, M. *et al.* Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. *Nat Metab* **4**, 284–299 (2022).
105. Fang, Z. *et al.* Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key  $\beta$ -Cell-Specific Disease Genes. *Cell Rep.* **26**, 3132–3144.e7 (2019).
106. Tatsuoka, H. *et al.* Single-Cell Transcriptome Analysis Dissects the Replicating Process of Pancreatic Beta Cells in Partial Pancreatectomy Model. *iScience* **23**, 101774 (2020).
107. Piñeros, A. R. *et al.* Single-Cell Transcriptional Profiling of Mouse Islets Following Short-Term Obesogenic Dietary Intervention. *Metabolites* **10**, (2020).
108. Sznurkowska, M. K. *et al.* Defining Lineage Potential and Fate Behavior of Precursors during Pancreas Development. *Dev. Cell* **46**, 360–375.e5 (2018).
109. Scavuzzo, M. A. *et al.* Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. *Nat. Commun.* **9**, 3356 (2018).
110. Qiu, W.-L. *et al.* Deciphering Pancreatic Islet  $\beta$  Cell and  $\alpha$  Cell Maturation Pathways and Characteristic Features at the Single-Cell Level. *Cell Metab.* **27**, 702 (2018).
111. Zeng, C. *et al.* Pseudotemporal Ordering of Single Cells Reveals Metabolic Control of Postnatal  $\beta$  Cell Proliferation. *Cell Metab.* **25**, 1160–1175.e11 (2017).
112. Dorrell, C. *et al.* Human islets contain four distinct subtypes of  $\beta$  cells. *Nat. Commun.* **7**, 11756 (2016).
113. Wang, Y. J. *et al.* Single-Cell Mass Cytometry Analysis of the Human Endocrine Pancreas. *Cell Metab.* **24**, 616–626 (2016).
114. Hrovatin, K. *et al.* Delineating mouse  $\beta$ -cell identity during lifetime and in diabetes with a single cell atlas. *Nat Metab* **5**, 1615–1637 (2023).
115. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714–730.e22 (2019).
116. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).

117. Beumer, J. *et al.* High-Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells. *Cell* **182**, 1062–1064 (2020).
118. Parikh, K. *et al.* Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* **567**, 49–55 (2019).
119. Huang, B. *et al.* Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. *Cell* **179**, 1160–1176.e24 (2019).
120. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021).
121. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
122. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
123. Basak, O. *et al.* Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. *Cell Stem Cell* **20**, 177–190.e4 (2017).
124. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156–1167.e15 (2018).
125. Kleinert, M. *et al.* Animal models of obesity and diabetes mellitus. *Nat. Rev. Endocrinol.* **14**, 140–162 (2018).
126. Bakhti, M., Böttcher, A. & Lickert, H. Modelling the endocrine pancreas in health and disease. *Nat. Rev. Endocrinol.* **15**, 155–171 (2019).
127. Zettler, S. *et al.* A decade of experience with genetically tailored pig models for diabetes and metabolic research. *Anim. Reprod.* **17**, (2020).
128. Renner, S. *et al.* Porcine models for studying complications and organ crosstalk in diabetes mellitus. *Cell Tissue Res.* **380**, 341–378 (2020).
129. Renner, S. *et al.* Comparative aspects of rodent and nonrodent animal models for mechanistic and translational diabetes research. *Theriogenology* **86**, 406–421 (2016).
130. Ludwig, B. *et al.* Favorable outcome of experimental islet xenotransplantation without

- immunosuppression in a nonhuman primate model of diabetes. *Proceedings of the National Academy of Sciences* vol. 114 11745–11750 Preprint at <https://doi.org/10.1073/pnas.1708420114> (2017).
131. Coe, T. M., Markmann, J. F. & Rickert, C. G. Current status of porcine islet xenotransplantation. *Curr. Opin. Organ Transplant.* **25**, 449–456 (2020).
132. Kim, S. *et al.* Molecular and genetic regulation of pig pancreatic islet cell development. *Development* **147**, (2020).
133. GitHub - scverse/scanpy: Single-cell analysis in Python. Scales to >1M cells. *GitHub* <https://github.com/scverse/scanpy>.
134. Mereu, E. *et al.* Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* **38**, 747–755 (2020).
135. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
136. Vieth, B., Parekh, S., Ziegenhain, C., Enard, W. & Hellmann, I. A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat. Commun.* **10**, 4667 (2019).
137. Cole, M. B. *et al.* Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq. *Cell Syst* **8**, 315–328.e8 (2019).
138. Weinreb, C., Wolock, S. & Klein, A. M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018).
139. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).
140. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
141. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
142. Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7426–7431 (2005).
143. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set



- overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016).
144. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML] Preprint* at <http://arxiv.org/abs/1802.03426> (2018).
145. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
146. GitHub - vtraag/louvain-igraph: Implementation of the Louvain algorithm for community detection with various methods for use with igraph in python. *GitHub* <https://github.com/vtraag/louvain-igraph>.
147. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
148. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
149. Pasquini, G., Rojo Arias, J. E., Schäfer, P. & Busskamp, V. Automated methods for cell type annotation on scRNA-seq data. *Comput. Struct. Biotechnol. J.* **19**, 961–969 (2021).
150. Xi, N. M. & Li, J. J. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. *Cell Syst* **12**, 176–194.e6 (2021).
151. Neavin, D. *et al.* Demuxafy: Improvement in droplet assignment by integrating multiple single-cell demultiplexing and doublet detection methods. *bioRxiv* 2022.03.07.483367 (2022) doi:10.1101/2022.03.07.483367.
152. Xi, N. M. & Li, J. J. Protocol for executing and benchmarking eight computational doublet-detection methods in single-cell RNA sequencing data analysis. *STAR Protoc* **2**, 100699 (2021).
153. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).
154. Lotfollahi, M. *et al.* Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* **40**, 121–130 (2022).
155. Kang, J. B. *et al.* Efficient and precise single-cell reference atlas mapping with

- Symphony. *Nat. Commun.* **12**, 5890 (2021).
156. Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
157. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
158. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
159. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
160. Forrow, A. & Schiebinger, G. LineageOT is a unified framework for lineage tracing and trajectory inference. *Nat. Commun.* **12**, 4940 (2021).
161. Klein, D. *et al.* Mapping cells through time and space with moscot. *bioRxiv* (2023) doi:10.1101/2023.05.11.540374.
162. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
163. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
164. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).
165. Büttner, M., Ostner, J., Müller, C. L., Theis, F. J. & Schubert, B. scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021).
166. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
167. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160 (2016).
168. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
169. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single

- cells. *Nat. Methods* **14**, 865–868 (2017).
170. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).
171. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
172. Larsson, L., Frisén, J. & Lundeberg, J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods* **18**, 15–18 (2021).
173. Ogbeide, S., Giannese, F., Mincarelli, L. & Macaulay, I. C. Into the multiverse: advances in single-cell multiomic profiling. *Trends Genet.* **38**, 831–843 (2022).
174. Duvall, E. *et al.* Single-cell transcriptome and accessible chromatin dynamics during endocrine pancreas development. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2201267119 (2022).
175. de la O, S., Yao, X., Chang, S., Liu, Z. & Sneddon, J. B. Single-cell chromatin accessibility of developing murine pancreas identifies cell state-specific gene regulatory programs. *Mol Metab* **73**, 101735 (2023).
176. Chakravarthy, H. *et al.* Converting Adult Pancreatic Islet  $\alpha$  Cells into  $\beta$  Cells by Targeting Both Dnmt1 and Arx. *Cell Metab.* **25**, 622–634 (2017).
177. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic  $\alpha$  to  $\beta$  cell reprogramming. *J. Clin. Invest.* **123**, 1275–1284 (2013).
178. Ackermann, A. M., Wang, Z., Schug, J., Najj, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab* **5**, 233–244 (2016).
179. Wang, G. *et al.* Integration of single-cell multiomic measurements across disease states with genetics identifies mechanisms of beta cell dysfunction in type 2 diabetes. Preprint at <https://doi.org/10.1101/2022.12.31.522386>.
180. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
181. Kester, L. & van Oudenaarden, A. Single-Cell Transcriptomics Meets Lineage Tracing.

- Cell Stem Cell* **23**, 166–179 (2018).
182. Rui, J. *et al.*  $\beta$  Cells that Resist Immunological Attack Develop during Progression of Autoimmune Diabetes in NOD Mice. *Cell Metab.* **25**, 727–738 (2017).
183. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nature Biotechnology* vol. 40 308–318 Preprint at <https://doi.org/10.1038/s41587-021-01182-1> (2022).
184. Tian, L., Chen, F. & Macosko, E. Z. The expanding vistas of spatial transcriptomics. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01448-2.
185. Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* **41**, 332–336 (2023).
186. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e17 (2016).
187. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882.e21 (2016).
188. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022).
189. Ji, Y., Lotfollahi, M., Wolf, F. A. & Theis, F. J. Machine learning for perturbational single-cell omics. *Cell Syst* **12**, 522–537 (2021).
190. Sikkema, L. *et al.* An integrated cell atlas of the lung in health and disease. *Nat. Med.* **29**, 1563–1577 (2023).
191. Osumi-Sutherland, D. *et al.* Cell type ontologies of the Human Cell Atlas. *Nat. Cell Biol.* **23**, 1129–1135 (2021).
192. Campbell-Thompson, M. *et al.* Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. *Diabetes. Metab. Res. Rev.* **28**, 608–617 (2012).
193. Kaestner, K. H., Powers, A. C., Naji, A., HPAP Consortium & Atkinson, M. A. NIH Initiative to Improve Understanding of the Pancreas, Islet, and Autoimmunity in Type 1

- Diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes* **68**, 1394–1402 (2019).
194. Shapira, S. N., Naji, A., Atkinson, M. A., Powers, A. C. & Kaestner, K. H. Understanding islet dysfunction in type 2 diabetes through multidimensional pancreatic phenotyping: The Human Pancreas Analysis Program. *Cell Metab.* **34**, 1906–1913 (2022).
195. Polychronidou, M. *et al.* Single-cell biology: what does the future hold? *Mol. Syst. Biol.* **19**, e11799 (2023).
196. Cui, H. *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* (2024) doi:10.1038/s41592-024-02201-0.
197. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
198. Erfanian, N. *et al.* Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed. Pharmacother.* **165**, 115077 (2023).
199. Michielsen, L. *et al.* Single-cell reference mapping to construct and extend cell-type hierarchies. *NAR Genom Bioinform* **5**, lqad070 (2023).
200. Lotfollahi, M., Naghipourfar, M., Theis, F. J. & Wolf, F. A. Conditional out-of-distribution generation for unpaired data using transfer VAE. *Bioinformatics* **36**, i610–i617 (2020).
201. Lotfollahi, M. *et al.* Predicting cellular responses to complex perturbations in high-throughput screens. *Mol. Syst. Biol.* **19**, e11517 (2023).
202. Bunne, C. *et al.* Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* **20**, 1759–1768 (2023).
203. CZI Single-Cell Biology Program *et al.* CZ CELL×GENE Discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *bioRxiv* (2023) doi:10.1101/2023.10.30.563174.
204. Single Cell Portal. [https://singlecell.broadinstitute.org/single\\_cell](https://singlecell.broadinstitute.org/single_cell).
205. Lotfollahi, M. *et al.* Biologically informed deep learning to query gene programs in single-cell atlases. *Nat. Cell Biol.* **25**, 337–350 (2023).
206. Seninge, L., Anastopoulos, I., Ding, H. & Stuart, J. VEGA is an interpretable generative

model for inferring biological network activity in single-cell transcriptomics. *Nat. Commun.* **12**, 5684 (2021).

# Appendix

## Appendix A

This is the published version of the article in *Molecular Metabolism* following peer review. The article is open access thus the published version is inserted here. This article is published in *Molecular Metabolism* under a Creative Commons Attribution license (CC BY 4.0).

**Sophie Tritschler**, Moritz Thomas, Anika Böttcher, Barbara Ludwig, Janine Schmid, Undine Schubert, Elisabeth Kemter, Eckhard Wolf, Heiko Lickert, Fabian J. Theis, **A transcriptional cross species map of pancreatic islet cells**, *Molecular Metabolism*, Volume 66, 101595 (2022)



# A transcriptional cross species map of pancreatic islet cells



Sophie Tritschler<sup>1,2,3</sup>, Moritz Thomas<sup>3,4</sup>, Anika Böttcher<sup>2,5</sup>, Barbara Ludwig<sup>6,7,8</sup>, Janine Schmid<sup>6</sup>, Undine Schubert<sup>6</sup>, Elisabeth Kemter<sup>8,9,10</sup>, Eckhard Wolf<sup>8,9,10</sup>, Heiko Lickert<sup>2,5,8,11,\*\*</sup>, Fabian J. Theis<sup>1,12,\*</sup>

## ABSTRACT

**Objective:** Pancreatic islets of Langerhans secrete hormones to regulate systemic glucose levels. Emerging evidence suggests that islet cells are functionally heterogeneous to allow a fine-tuned and efficient endocrine response to physiological changes. A precise description of the molecular basis of this heterogeneity, in particular linking animal models to human islets, is an important step towards identifying the factors critical for endocrine cell function in physiological and pathophysiological conditions.

**Methods:** In this study, we used single-cell RNA sequencing to profile more than 50'000 endocrine cells isolated from healthy human, pig and mouse pancreatic islets and characterize transcriptional heterogeneity and evolutionary conservation of those cells across the three species. We systematically delineated endocrine cell types and  $\alpha$ - and  $\beta$ -cell heterogeneity through prior knowledge- and data-driven gene sets shared across species, which altogether capture common and differential cellular properties, transcriptional dynamics and putative driving factors of state transitions.

**Results:** We showed that global endocrine expression profiles correlate, and that critical identity and functional markers are shared between species, while only approximately 20% of cell type enriched expression is conserved. We resolved distinct human  $\alpha$ - and  $\beta$ -cell states that form continuous transcriptional landscapes. These states differentially activate maturation and hormone secretion programs, which are related to regulatory hormone receptor expression, signaling pathways and different types of cellular stress responses. Finally, we mapped mouse and pig cells to the human reference and observed that the spectrum of human  $\alpha$ - and  $\beta$ -cell heterogeneity and aspects of such functional gene expression are better recapitulated in the pig than mouse data.

**Conclusions:** Here, we provide a high-resolution transcriptional map of healthy human islet cells and their murine and porcine counterparts, which is easily queryable via an online interface. This comprehensive resource informs future efforts that focus on pancreatic endocrine function, failure and regeneration, and enables to assess molecular conservation in islet biology across species for translational purposes.

© 2022 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords** Pancreatic islets;  $\beta$ -Cell;  $\alpha$ -Cell; Single-cell RNAseq; Cross species conservation; Translation

## 1. INTRODUCTION

Pancreatic  $\beta$ -cells are essential endocrine cells, which regulate systemic glucose homeostasis together with the other endocrine islet cells - glucagon-producing  $\alpha$ -cells, somatostatin-producing  $\delta$ -cells, pancreatic polypeptide-producing PP-cells and ghrelin-producing  $\epsilon$ -cells. In diabetic patients,  $\beta$ -cells are lost or become

dysfunctional, which leads to chronically elevated blood glucose levels. Even in healthy individuals,  $\beta$ -cells are heterogeneous and differ in their responsiveness to glucose, insulin secretion capacity, maturation state, stress response and other functional phenotypes [1–3]. Similarly, varying phenotypes and cell states of  $\alpha$ -cells have been described [4–6]. It has been proposed that these molecular and functional cell states complement each other to fine tune and

<sup>1</sup>Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany <sup>2</sup>Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, 85764 Neuherberg, Germany <sup>3</sup>Technical University of Munich, School of Life Sciences Weihenstephan, 85354 Freising, Germany <sup>4</sup>Institute of AI for Health, Helmholtz Zentrum München, 85764 Neuherberg, Germany <sup>5</sup>Institute of Stem Cell Research, Helmholtz Zentrum München, 85764 Neuherberg, Germany <sup>6</sup>Department of Medicine III, University Hospital Carl Gustav Carus, Technical University of Dresden, 01307 Dresden, Germany <sup>7</sup>Paul Langerhans Institute Dresden of Helmholtz Zentrum München, University Hospital Carl Gustav Carus, Technical University of Dresden, 01307 Dresden, Germany <sup>8</sup>German Center for Diabetes Research (DZD), 85764 Neuherberg, Germany <sup>9</sup>Chair for Molecular Animal Breeding and Biotechnology, Gene Center, LMU Munich, 81377 Munich, Germany <sup>10</sup>Center for Innovative Medical Models (CiMM), Department of Veterinary Sciences, LMU Munich, 85764 Oberschleißheim, Germany <sup>11</sup>Technical University of Munich, Medical Faculty, 81675 Munich, Germany <sup>12</sup>Technical University of Munich, Department of Mathematics, 85748 Garching b. Munich, Germany

\*Corresponding author. Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. E-mail: [fabian.theis@helmholtz-muenchen.de](mailto:fabian.theis@helmholtz-muenchen.de) (F.J. Theis).

\*\*Corresponding author. Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, 85764 Neuherberg, Germany. E-mail: [heiko.lickert@helmholtz-muenchen.de](mailto:heiko.lickert@helmholtz-muenchen.de) (H. Lickert).

Received March 5, 2022 • Revision received August 20, 2022 • Accepted September 3, 2022 • Available online 13 September 2022

<https://doi.org/10.1016/j.molmet.2022.101595>

efficiently adapt the endocrine response to physiological changes in their environment [3,7,8]. Heterogeneity can also arise from individual cells that cycle asynchronously between phases of active insulin biosynthesis, recovery and rest [9], different tissue locations or phenotypic variation between cells of different ages [10]. Although it is unclear to which extent the endocrine heterogeneity is important for normal pancreatic endocrine function, a precise description of the functional and molecular differences between distinct cell states informs drug discovery and development of anti-diabetic drugs [4,11–14]. Most importantly, it will help to establish a reference for a mature, functional  $\beta$ -cell as a clinical endpoint. Moreover, aspects of the molecular programs that characterize less-functional or stressed states, may overlap with programs that contribute to pathological  $\beta$ -cell dysfunction in diabetes and thus reveal novel molecular targets. Lastly, it can indicate which subset of cells has the potential to respond to a treatment, which affects the efficacy of a therapeutic approach. Today, most of the pre-clinical research of the endocrine system relies on animal models as access to pancreatic tissue from patients is limited. Endocrine cells are mostly studied in rodents. However, differences in endocrine development and whole-body anatomy and physiology between human and rodents lowers the predictive value of rodent models for human physiology and therapeutic success [15]. As an alternative to rodents, pigs are a large-animal model with a higher translational promise: The anatomy and physiology of pigs is more similar to humans, porcine islets are a potential source for islet xenotransplantation, and, ethical concerns about animal studies are smaller for pigs than for non-human primates [16–19]. Still, it is unclear whether human functional states and molecular profiles of endocrine cells are better conserved in pigs than rodents [20]. Only recently, endocrine heterogeneity can be systematically characterized at the molecular level by profiling individual cells with high-throughput single-cell RNA sequencing [12]. Most phenotypic states are reflected in the gene or protein expression profile of a cell and can thus be captured and resolved by single-cell approaches. Single-cell studies have provided cell-by-cell descriptions of healthy and diabetic pancreatic islets from mice [11,21,22] and human donors [4,9,13,14,23–25], however in these early studies resolution was limited by low cell numbers - which makes it difficult to identify rare cell states and to infer cell state transitions - and there is so far no systematic cross-species comparison. Here, we leveraged single-cell transcriptomics to finely resolve human endocrine heterogeneity and its conservation in pig and mouse islets. We describe endocrine cell type signatures and gradients as well as distinct  $\alpha$ - and  $\beta$ -cell states that can be related to distinct biological properties like function, maturation and cellular stress. Our data represents a queryable resource to provide insight into shared endocrine cell states and expression profiles in humans, pigs and mice, which can be easily accessed and explored online and adheres to the FAIR data guiding principles [26].

## 2. RESULTS

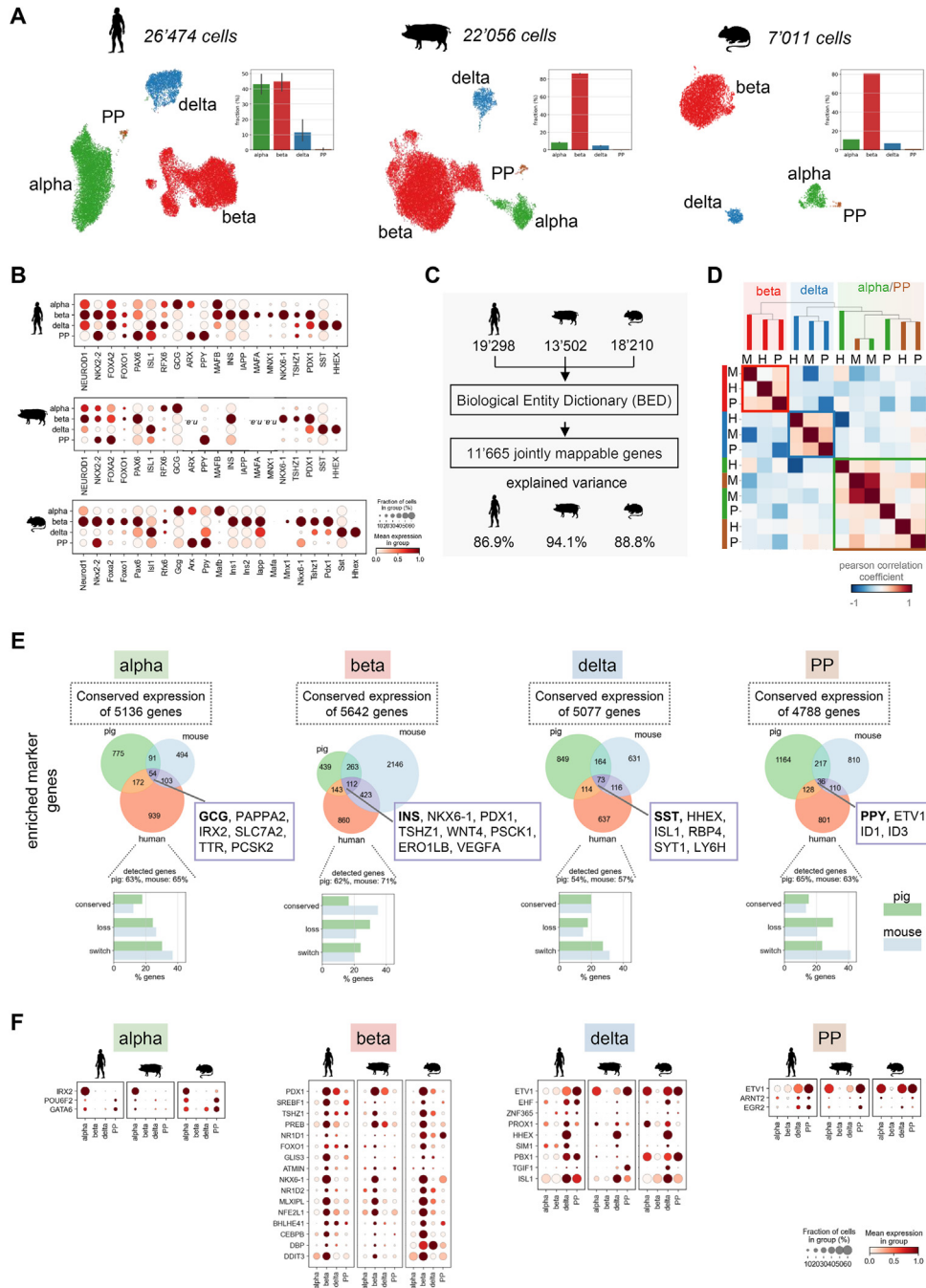
### 2.1. Conservation of endocrine signatures across human, pig, and mouse

We sequenced >50'000 single cells from pancreatic islets isolated from 5 healthy human donors (age 22–74 years, male and female), a Göttingen minipig (2 replicates, age 3 years 8 months, female) and 3 mice (pooled, C57BLJ/6, age 23.5 weeks, male) to describe transcriptome-wide expression signatures of human endocrine cell populations and their conservation in animal models (Figure 1A, Figure S1A, B). To facilitate exploration and reuse of our data set we published

it in the cellxgene portal (<https://cellxgene.cziscience.com/collections/0a77d4c0-d5d0-40f0-aa1a-5e1429bcdb7e>) and added it to the sfaira data zoo [27], which both follow the concept of FAIR data [26]. In all three species we identified the four main endocrine cell types:  $\alpha$ -,  $\beta$ -,  $\delta$ -, PP-cells. We captured a few rare *GHRH* positive  $\epsilon$ -cells in the human, but not in pig and mouse samples, and therefore did not consider them for downstream analyses. Likewise, we excluded poly-hormonal cells as it is difficult to distinguish the profile of true poly-hormonal cells from doublet cells (Supplementary Table 1). In human islets the ratio of  $\alpha$ - and  $\beta$ -cells was relatively balanced, while in pig and mouse islets  $\beta$ -cells were most abundant (~80%). These cell type frequencies are consistent with reported quantification in histological sections [28,29], which indicates our data is less confounded by technical artifacts than previous single-cell studies with low  $\beta$ -cell frequencies [14,23,25]. Human cells expressed established islet hormones and transcription factors defining endocrine cell identities. These expression patterns were conserved in pig and mouse clusters with a few known exceptions (Figure 1B). For example, the transcription factor *MAFB* was expressed in human  $\alpha$ -,  $\beta$ - and  $\delta$ -cells, but only in mouse  $\alpha$ -cells. In pig, we detected low levels of *MAFB* in  $\alpha$ -,  $\beta$ - and  $\delta$ -cells similar to human islets as it was recently described in bulk expression profiles of sorted islet cells [20]. Such low detection levels are a general issue in RNA-seq studies of pig cells. The functional annotation of the pig genome is still less complete than for mouse and human genes, although continuity and quality of the reference sequence has been greatly improved [30–32]. Due to incomplete annotation of protein-coding genes, a subset of reads cannot be confidently mapped and are thus discarded. In our data this included the transcription factors *MAFA* or *ARX*, which were not detected in pig cells (Figure 1B). The lower mapping rate for pig sequencing data can limit the interpretability of genes that are not expressed.

To directly compare gene expression across species, we identified mappable gene orthologs using the Biological Entity Dictionary (BED) [33] tool (Figure 1C). Out of approximately 19'300 human, 13'500 pig and 18'200 mouse genes (annotated and detected), 11'665 genes were mappable across all three species. The 11'665 genes explained on average 90% of the total variance in each species (human = 87%, pig = 94%, mouse = 89%, Figure 1C). We computed pairwise correlation and clustering of cell type profiles in the principal component analysis (PCA) representation of the scaled and concatenated cross-species data to assess global transcriptional similarity of human, pig and mouse endocrine cell types (Figure 1D). We did not consider the two top-variance components, because they were almost entirely driven by cross-species variation (Figure S1C). Cell types correlated stronger among each other than among species, which indicates that globally cell type expression profiles were conserved (mean Pearson's rho for  $\alpha$ -cells = 0.15, for  $\beta$ -cells = 0.12, for  $\delta$ -cells = 0.23, for PP-cells = 0.2, for human-cells = -0.15, for pig-cells = -0.26 and for mouse-cells = 0.02). Moreover,  $\alpha$ - and PP-cells were closely related to each other and more distant to  $\beta$ - and  $\delta$ -cells in all three species. During development mutual inhibition of lineage determinants promotes endocrine progenitors towards a  $\alpha$ -/PP- or  $\beta$ -/ $\delta$ -cell fate [34,35], thus, this developmental proximity of  $\alpha$ -/PP and  $\beta$ -/ $\delta$ -cells is reproduced in adult islets. Further, this suggests that developmental programs of endocrine subtype specification are conserved across species.

Next, we evaluated the overlap of gene expression between species in each cell type (Figure 1E, Supplementary Table 2). We found that on average 5'160 out of 11'665 mappable genes showed conserved expression in >5% of the cells in each endocrine cell type across species. This indicates that only 50–60% of genes expressed in



**Figure 1: Conservation of endocrine signatures in human, pig, and mouse islets.** A) UMAP plots of scRNA-seq data of human, pig and mouse pancreatic islets capturing all 4 major endocrine populations. Barplots show cell type compositions, which reflect islet composition *in vivo*. B) Expression of islet hormones and known endocrine and lineage transcription factors in human, pig and mouse endocrine cell types. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene. *N. a.* means genes were not detected. C) Overview of gene orthologue mapping between species to assess conservation of the human transcriptional signature. Explained variance is the fraction of the total variance captured by the subset of mappable genes. D) Correlation matrix of gene expression indicates global conservation of transcriptional profiles of endocrine cell types across species. Cell types are grouped by hierarchical clustering. Pairwise correlation is computed in the principal component analysis space after excluding the top two variance components, which are entirely driven by cross-species variation (see also Figure S1C).  $\alpha$ -,  $\beta$ - and  $\delta$ -cells were subsampled to 2000 cells to balance cell type representation. E) Conservation of endocrine gene and marker expression. *Top*: Venn diagram showing overlap between species of enriched marker genes for each endocrine cell type. Only marker genes that are mappable across species are shown. Selected known overlapping cell type markers and number of genes with conserved expression are indicated. Enriched marker genes are defined as genes expressed in >5% of the cells of the corresponding cell type and showing increased expression versus all other cell types ( $\log_2$ -fold change > 0.5). *Bottom*: Conservation of human enriched marker genes in pig and mouse cell types. % of human enriched marker genes expressed/detected is indicated. Conserved: enriched marker in same cell type as human; loss: detected but not an enriched marker; switch: enriched marker in different cell type than human. F) Expression of enriched and conserved transcription factors for each endocrine cell type in human, pig and mouse. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

human cell types are shared with their mouse and pig counterparts (Figure S1D). The majority of the other 40–50% were either only expressed in another cell type (“loss of expression”) or not expressed or detected. The remaining 5% were not expressed in human but detected in pig or mouse cells (“gain of expression”). For example, we detected high mRNA levels of *free fatty acid receptor 4* *FFAR4*, as well as calcium-sensing receptor *CASR* in human  $\beta$ - and  $\delta$ -cells (Figure S1E). The expression of both genes was low or lost in mouse and pig  $\beta$ -cells but conserved in  $\delta$ -cells (4.7% of cells in pig). In addition, mouse  $\alpha$ -cells “gained” expression of *FFAR4* while pig  $\alpha$ -cells “gained” *CASR*. Similarly, the synaptic protein *neuronal pentraxin-2* (*NPTX2*) was strongly expressed in human  $\beta$ - and  $\delta$ -cells, all pig endocrine subtypes, but mostly lost or not detected in mouse cells. Instead, mouse cells expressed *neuronal pentraxin-1* (*NPTX1*). The subtype expression pattern of the transcription factors *DNA-binding protein inhibitor 1D1-4* was highly conserved in humans and pigs, but not in mice. These examples highlight cell type specific species differences in receptors and regulatory or signaling proteins relevant for islet function. As noted previously, not detected expression of a gene can be due to either biological species differences or technical factors such as genome annotation or sequencing depth. For validation, we compared our results to reported core genes derived from human and mouse bulk  $\beta$ -cell transcriptomes [36] (Figure S1F). From the 85.5% of core genes (8105/9474 core genes) captured within the 11’665 mappable genes, we found that 77% overlapped with those we identified as conserved between human and mouse  $\beta$ -cells. This indicates that our approach approximates conservation consistent with previous reports. Differences may be due to the distinct data types, how conservation is defined and or detection limits in scRNA-seq data. Beyond global gene expression profiles, we focused on cell type enriched marker genes to approximate conservation of cell type-specific functions (Figure 1E, Supplementary Table 3). As a positive control, we verified that we identify established marker genes in all species, which included *GCG*, *IRX2* and *TTR* for  $\alpha$ -, *INS*, *PDX1* and *NKX6-1* for  $\beta$ -, *SST* and *HHEX* for  $\delta$ - and *PPY* for PP-cells. Surprisingly, of the remaining identified human marker genes only 5–10% were shared with both mouse and pig. The small overlap was not biased by one species, because the overlap with human markers was similar for mouse and pig markers. Overall, we observed that in all cell types less than 20% of the human markers were conserved, approximately 20% were expressed but did not appear as marker genes (‘loss’), and 30% marked other populations (‘switch’). The rest was not detected or expressed. We thus conclude that while critical identity and functional marker genes are conserved, cell type specific expression is evolutionarily more labile. We noted that, especially in mice, fewer enriched marker genes were detected and conserved in  $\alpha$ - and PP-cells than in  $\beta$ - or  $\delta$ -cells, which may be explained by the high similarity of mouse  $\alpha$ - and PP-cell profiles.

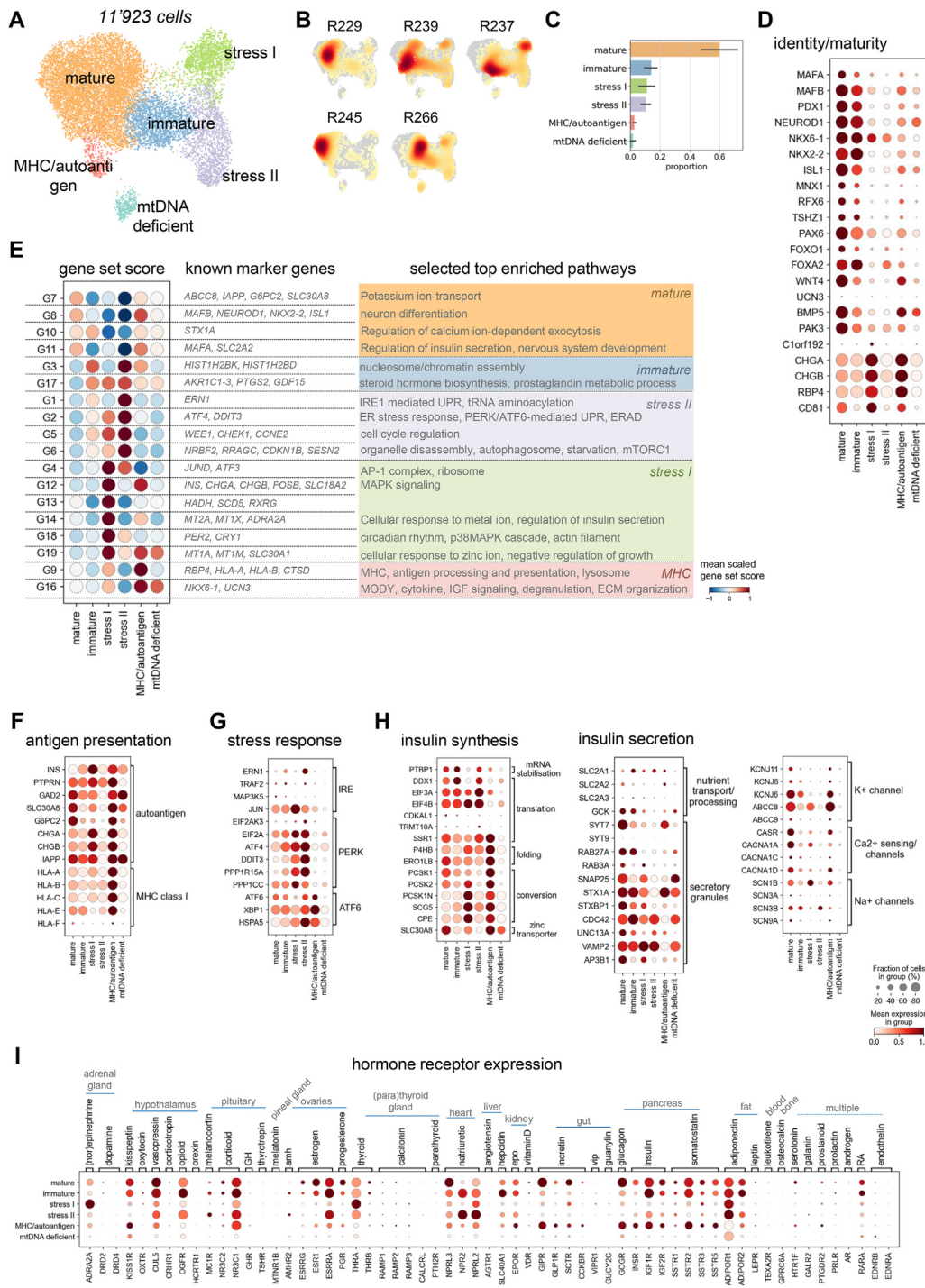
Finally, we assessed the similarity of transcription factor (TF) expression patterns. TFs are key components of the gene regulatory networks that determine endocrine cell identity during development and maintain identity and function in adult islets. Thus, we considered TF patterns as another measure for proximity of animal models to humans (Figure 1F). We assumed, TF networks are most likely best evolutionary conserved within the shared marker genes and subset to shared TFs. Moreover, to quantify similarity we considered TF expression across cell types, because for modeling transcriptional regulation in a cell type, not only TF expression but also cell type-specificity should be conserved. Lastly, we computed a correlation measure that includes the mean expression as well as the fraction of cells expressing a TF in a cluster to leverage all information contained

in single-cell data (Methods). With this approach, we observed that  $\alpha$ - and  $\beta$ -cell TF patterns were better conserved between human and pig (pearson’s rho = 0.97, p-value =  $10^{-7}$  for  $\alpha$ -cells, pearson’s rho = 0.73, p =  $10^{-12}$  for  $\beta$ -cells) than between human and mouse (pearson’s rho = 0.73, p = 0.006 for  $\alpha$ -cells, pearson’s rho = 0.57, p =  $10^{-7}$  for  $\beta$ -cells) (Figure S1G).  $\alpha$ - and  $\beta$ -cell TF patterns also correlated stronger between human and pig than human and mouse when considering all TFs we identified as cell-type enriched markers in humans (Figure S1H), or, all TFs with conserved expression (not necessarily cell-type enriched) (Figure S1I). Conversely, for  $\delta$ - and PP-cells, there were no pronounced differences between species when subset to conserved marker TFs (Figure S1G). Human and mouse  $\delta$ - and PP- TF patterns correlated stronger within all enriched marker TFs (Figure S1H), while human and pig  $\delta$ - and PP-TF patterns correlated stronger within all TFs with conserved expression (Figure S1I). Thus, this analysis suggests that  $\alpha$ - and  $\beta$ -cell TF expression and likely target gene regulation is closer to human in pig than in mouse models.

## 2.2. $\beta$ -Cell heterogeneity and phenotypic states in human islets

To understand  $\beta$ -cell heterogeneity in human islets, we clustered the human  $\beta$ -cells at higher resolution and identified six  $\beta$ -cell clusters (Figure 2A). These clusters did not form separated clusters, but rather connected states in the continuous  $\beta$ -cell manifold. All six clusters were represented in all five donors, but subtype composition varied across donors (Figure 2B,C). Approximately 60% of the cells formed a large cluster we annotated as *mature*  $\beta$ -cells, because they highly expressed canonical  $\beta$ -cell identity and maturity genes [37] (Figure 2C,D), and scored high for the  $\beta$ -cell hallmark pathway (Figure S2A). The other clusters made up less than 20% of all  $\beta$ -cells. Two clusters activated hallmark pathways associated with unfolded protein response, stress and apoptosis, which we therefore referred to as *stress I* and *stress II* cells (Figure S2A). Identity and maturity markers as well as  $\beta$ -cell hallmark scores decreased from the *mature* to the stress-clusters, which suggests a gradual loss of  $\beta$ -cell identity and maturity (Figure 2D). The state between the *mature* and the stress-clusters most resembled *immature* cells. In this intermediate state, pathways associated with the cell cycle and the PI3K-Akt-mTOR signaling axis were increased, which was previously reported to characterize less mature  $\beta$ -cells in mice [37,38] (Figure S2A). However, other reported markers of murine immature  $\beta$ -cells like *CHGB*, *RBP4* and *CD81* showed variable expression in the  $\beta$ -cell states that did not fully correlate with loss of maturity and identity markers (Figure 2D). We could not annotate the two remaining clusters based on this analysis, because the top scoring hallmark pathways were not related to an interpretable  $\beta$ -cell state, but described processes of other systems or tissues. Finally, we saw no strong upregulation of  $\beta$ -cell disallowed genes in any non-mature cluster compared to the *mature* cluster (Figure S2B). Thus, we identified clusters with established  $\beta$ -cell profiles, alongside novel transcriptional  $\beta$ -cell states.

$\beta$ -cell-specific processes can be better captured when gene sets are identified with an unbiased, data-driven approach. We therefore clustered the 3000 most variable genes into groups of highly correlated and or anti-correlated genes (hereafter referred to as *gene sets*) and then related these gene sets to cellular processes based on known marker genes and pathway enrichment for interpretation (Figure 2E, Supplementary Table 4, Methods). This approach was previously described to identify *de novo* gene sets in single-cell data [39] and is commonly used in correlation network analysis [40]. In contrast to describing cell states with marker genes, it gathers genes into context-specific groups independent of the predefined cell states, i.e. the same set of genes can be activated in multiple cell states. We identified four



**Figure 2: Transcriptional  $\beta$ -cell heterogeneity and states in human islets.** A) UMAP plot of 11'923 human  $\beta$ -cells. Colors highlight clustering into six different  $\beta$ -cell states. B) Cell densities in UMAP space for five human donors shows that all  $\beta$ -cell clusters are represented by all donors. ID indicates donor ID for ADI IsletCore (see also Figure S1A). C) Fraction of cells per  $\beta$ -cell cluster. Error bar indicates donor variation. D) Expression of selected known  $\beta$ -cell identity and maturity markers. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene. E) Gene sets capturing variation in human  $\beta$ -cells that describe biological processes. Gene sets are groups of highly correlated and/or anti-correlated genes identified using hierarchical clustering on the correlation matrix of the top 3000 variable genes. *Left:* Scaled mean score for each gene set per  $\beta$ -cell cluster. For each gene set selected known  $\beta$ -cell identity or functional marker genes are indicated. *Right:* Summary of selected enriched pathways for each gene set indicating biological processes associated to gene sets. Coloring indicates the highest scoring  $\beta$ -cell cluster. F–H) Expression of genes encoding MHC class I components and  $\beta$ -cell autoantigens (F), members of the three canonical ER stress response arms (G), and insulin synthesis and secretion pathways (H). Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene. I) Expression of receptors for the majority of circulating hormones in human  $\beta$ -cell clusters. The tissue or organ origin and the type of hormone are indicated. Only receptors detected in  $>5\%$  of cells of any cluster are shown. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

gene sets (G7-8, 10–11) scoring high in *mature*  $\beta$ -cells that contain key markers and enriched pathways of  $\beta$ -cell identity, glucose sensing and insulin secretion (Figure 2E). These gene sets were decreased in the *immature*, *stress I* and *stress II* clusters.

Beyond canonical  $\beta$ -cell function, one gene set (G9) was enriched for factors involved in antigen processing and presentation including major histocompatibility complex (MHC) class I and lysosome. Cells scoring high for the MHC/antigen processing-associated gene set formed a small cluster we could not previously annotate and also highly expressed  $\beta$ -cell identity and function genes as well as reported  $\beta$ -cell autoantigens (Figure 2F). We therefore referred to the cluster as *MHC/autoantigen*. While healthy  $\beta$ - and other endocrine cells steadily present self peptides via MHC class I complex, hyperexpression of MHC class I genes has been observed in islets of T1D patients. Increased levels of MHC class I were suggested to contribute to aberrant antigen presentation and autoimmune-mediated  $\beta$ -cell destruction [41]. To confirm that this gene set captures biologically relevant  $\beta$ -cell features, we compared the *MHC/autoantigen* state to  $\beta$ -cells from T1D patients [42], and observed a high T1D  $\beta$ -cell-derived score in *MHC/autoantigen* cells (Figure S2D). *Vice versa*, T1D  $\beta$ -cells highly expressed MHC class I genes and activated the MHC/antigen processing gene set (G9) when compared to healthy  $\beta$ -cells (Figure 2E, F). Also in  $\alpha$ - and  $\delta$ -cells a small subset of cells scored high for this gene set, which suggests a similar MHC-high state exists in other endocrine cell types (Figure S2G). Besides an increased MHC and lysosomal gene expression, *MHC/autoantigen* scored low for a gene set enriched for ribosomal genes (G4) (Figure 2E). This may indicate reduced ribosomal biogenesis and translation. Consistently, the expression of multiple regulatory factors of translation (e.g. translation initiation factors) was decreased in the *MHC/autoantigen* state (Figure S2H). Moreover, *MHC/autoantigen* cells lowly expressed genes governing gene transcription including transcription initiation factors and subunits of RNA polymerase, which likely was linked to a reduced number of total genes expressed per cell (Figure S2H, I). G16, which contained  $\beta$ -cell markers *UCN3* and *NKX6-1*, also scored highest in the *MHC/autoantigen* cluster. However, low overall variance of the activation score for G16 indicated that the magnitude of the activation level differences was small and thus the gene set was similarly activated in all  $\beta$ -cells (Figure S2C). Together, this suggests the presence of rare  $\beta$ - as well as  $\alpha$ - and  $\delta$ -cells in healthy islets, which downregulate global transcription and translation, but maintain  $\beta$ -cell identity and enhance MHC class I-mediated antigen processing and presentation. When insulin demand is high, the ER protein folding capacity of  $\beta$ -cells can be overwhelmed and misfolded proinsulin accumulates. To counteract the overload and its resulting stress,  $\beta$ -cells activate a UPR-mediated stress response [43,44]. For this adaptive UPR, also constitutive, low autophagy is considered important to remove the misfolded proteins and damaged organelles. We identified three gene sets, which captured these cellular stress response pathways and autophagosome and organelle disassembly (Figure 2E). All three gene sets were highly activated in the *stress II* cluster and a subset in the *stress I* cluster. Consistent with the gene set analysis, the three main global stress response arms - IRE, PERK and ATF6-mediated stress response-were differentially activated in the  $\beta$ -cell states (Figure 2G, Figure S2J). The PERK-arm was induced in the *stress I* and *stress II* cluster, ATF6 in the *stress II* and *MHC/autoantigen* cluster, while IRE was only active in the *stress II* cluster. *Stress I* cells scored high for further gene sets enriched for the stress-induced transcription factor *ATF3*, AP-1 complex, metallothionein, circadian rhythm (Figure 2E). Metallothionein and circadian rhythm genes are a part of the transcriptional program recently reported to be regulated by glucocorticoid

signaling in human islets [45]. Glucocorticoid signaling has been associated with  $\beta$ -cell dysfunction and we therefore further compared the *stress I* profile to the transcriptional response glucocorticoid signaling induced. Like in glucocorticoid-treated islets, in *stress I* cells components of STAT and TGF $\beta$ -signaling as well as other islet growth factors including *Vascular endothelial growth factor A* (VEGFA) and *Platelet-derived growth factor subunit A* (PDGFA) were decreased (Figure S2K). Lastly, we annotated the remaining small cluster of cells as *mtDNA deficient* because mitochondria-encoded gene expression was low (Figure S2I). In this cluster most gene sets scored low, identity and maturity genes were decreased and also other data quality metrics were low (Figure 2D, Figure S2I). We therefore could not exclude that this cluster contained dying cells. In summary, our single-cell sequencing data captured distinct  $\beta$ -cell states that may reflect the transcriptional response to different stress factors. While maturity and identity markers and gene sets were not expressed in a large fraction of cells of non-mature  $\beta$ -cell states, stress-linked gene sets showed baseline activation in all  $\beta$ -cell states.

Finally, we sought to associate the distinct  $\beta$ -cell states with two key properties of  $\beta$ -cell function: insulin synthesis and secretion. We observed that all  $\beta$ -cell subpopulations expressed key genes of insulin synthesis (Figure 2H). Surprisingly, *stress I* and *MHC/autoantigen* cells expressed a higher level of *prohormone convertase 2* (*PCK2*) than *prohormone convertase 1* (*PCK1*) unlike the other  $\beta$ -cells. *PCK*-genes encode enzymes that cleave pro-hormones including insulin and glucagon. Consistent with the increased *PCK2* expression, also expression of *prohormone convertase subtilisin/kexin type 1 inhibitor* (*PCK1M*) - a *PCK1* inhibitor - and the *Neuroendocrine protein 7B2* (*SCG5*) - a chaperone of *PCK2*, which facilitates transport and function of *PCK2* - was increased in the *stress I* and *MHC/autoantigen* clusters. In healthy human donors, *PCK1* levels are reportedly higher in  $\beta$ -cells, while *PCK2* levels are higher in  $\alpha$ -cells [46]. A defective maturation of proinsulin is implicated in both T1D and T2D and plasma proinsulin to insulin ratio serves as a clinical index for  $\beta$ -cell dysfunction [47–50]. Our analysis suggests that variable *PCK* expression is part of the transcriptional programs turned on in  $\beta$ -cell substates, which eventually result in functional  $\beta$ -cell heterogeneity. The activation of key insulin secretion processes was more heterogeneous (Figure 2H, Figure S2L). Relative to *mature*  $\beta$ -cells, multiple genes linked to glucose sensing, and secretory granules as well as ion channels were decreased in *immature*, *stress I* and *stress II* cells, but not in the *MHC/autoantigen* cluster. Beyond glucose and other nutrients, various circulating body hormones regulate insulin secretion. To identify the target  $\beta$ -cell states of these hormones we explored the expression of their cognate receptors (Figure 2I). Reduced receptor expression of known insulin secretion stimuli including other islet hormones, gut incretins, adipose tissue hormones or estrogen correlated with reduced expression of insulin secretion genes in *immature*, *stress I*, *stress II* clusters. In *stress I* cells decreased insulin secretion might be associated with increased  $\alpha$ -2-adrenergic receptor (*ADRA2A*) expression and stimulation of inhibitory adrenergic signaling leading to reduced cAMP levels [51,52]. Consistently, the expression of several components of cAMP signaling was decreased in *stress I* cells (Figure S2M). In *stress II* and *immature* cells we observed a strong increase of *atrial natriuretic receptor 2* (*NPR2*) and the *Anti-Müllerian hormone receptor* (*AMHR2*). The effects of natriuretic peptides are still unclear, but insulinotropic and mitogenic effects on  $\beta$ -cells have been suggested [53,54]. To further corroborate that the described transcriptional heterogeneity is associated with functional heterogeneity we linked the  $\beta$ -cell states to electrophysiological measurements of exocytosis and ion channel activity in published single-cell “Patch-

seq” data of human islet cells (“Patch-seq”: whole-cell patch-clamp measurements combined with RNA sequencing) [4]. To map the Patch-seq cells to our reference  $\beta$ -cell states, we represented them as activation scores of the  $\beta$ -cell gene sets (Methods). The 230  $\beta$ -cells from healthy donors were similarly distributed across  $\beta$ -cell states and had similar marker expression and gene set activation profiles compared to our dataset (Figure S2N, O). As suggested from our transcriptional characterization, for *immature* and *stress II*  $\beta$ -cell decreased exocytotic function was measured compared to *mature*  $\beta$ -cells (Figure S2O). *Immature* cells also showed decreased Na<sup>+</sup> channel activity. No MHC-like and too few *stress I* cells were detected in the Patch-Seq data.

To confirm that the identified transcriptional  $\beta$ -cell states are robustly detected across study, age and sex we mapped  $\beta$ -cells of 9 single-cell studies (n = 54 donors) [9,13,14,23,25,55–58] to our reference  $\beta$ -cell map in the representative gene set space (Figure S3A, Methods). Approximately 60% of cells mapped to the *mature*  $\beta$ -cell state, and 10–25% to *immature*  $\beta$ -cell state in all studies. Also *stress I*, *stress II* and MHC/autoantigen-like cells were consistently captured in multiple studies with a sufficiently large  $\beta$ -cell number (median >700 cells per donor).  $\beta$ -Cell state fractions were not significantly increased in female or male donors or correlated with age (Figure S3B, C), which indicates that the observed donor variation is not strongly linked to these variables in the integrated datasets.

Collectively, these results established that changes in  $\beta$ -cell function and maturation are reflected in the transcriptional profile of a cell and include activation of stress pathways and differential hormone receptor expression.

### 2.3. $\beta$ -Cell maturation factors in human islets

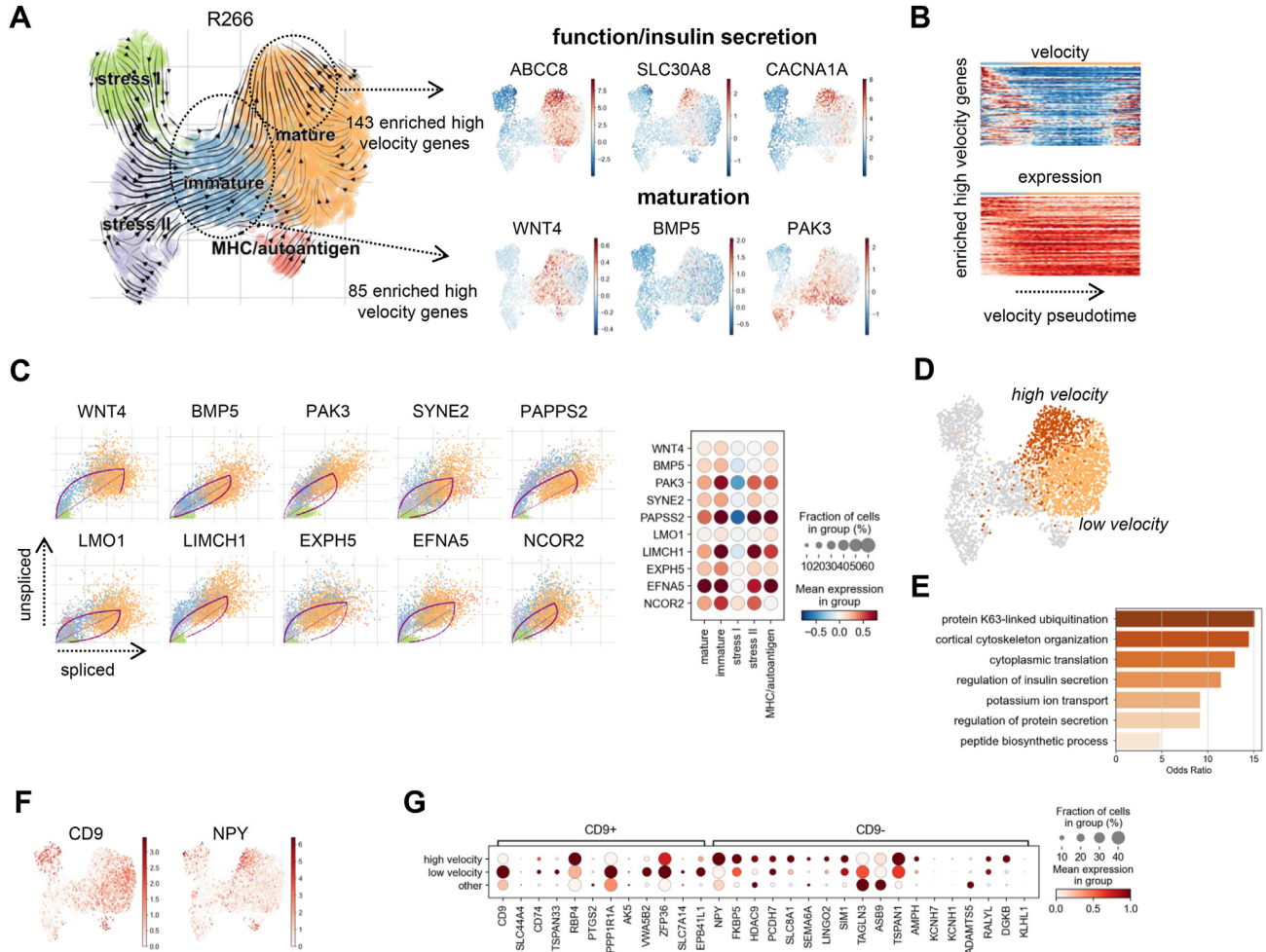
For clinical research it is crucial to identify the transcriptional programs critical to induce or maintain a functional  $\beta$ -cell with high insulin biosynthesis and secretion capacity. Single-cell sequencing can reconstruct gene expression dynamics by RNA velocity inference [59,60] and thereby reveal factors underlying a transcriptional state change. We applied RNA velocity analysis to  $\beta$ -cells of each donor separately, since current velocity inference methods cannot account for batch- and/or donor-variation. We then focused our analyses on one donor (ID R266), in which all  $\beta$ -states were well represented (Figure 3A), and confirmed the outcomes in the other four donors (Figure S4A). We identified two regions with high dynamics that captured *in silico* transcriptional state changes associated with  $\beta$ -cell maturation and insulin secretion, respectively. For the flow from *immature* to *mature* cells, we predicted high velocity for the signaling proteins *WNT4*, *BMP5* and *PAK3*, which are known markers of *mature*  $\beta$ -cells [38,61,62] (Figure 3A). This showed that maturity factors were actively transcribed in our immature cells, which suggests that the inferred dynamics may recapitulate aspects of  $\beta$ -cell maturation. Other genes with a similar dynamic behavior - i. d. high velocity in the *immature* cluster and high expression in the *mature* cluster - are additional putative maturation factors (Figure 3B,C, Supplementary Table 5). For example, we identified the co-regulatory *nuclear receptor co-repressor 2* (*NCOR2*) as well as *LIM* and *calponin-homology domains 1* (*LIMCH1*) - a positive regulator of non-muscle myosin II promoting focal adhesion assembly - which, to our knowledge, have not been previously associated with  $\beta$ -cell maturation (Figure 3C). Further also *ephrinA5* (*EFNA5*), a well known factor of neurogenesis and potential regulator of insulin secretion in  $\beta$ -cells [63], showed high velocity in *immature* cells (Figure 3C). The inferred dynamic behavior of these genes was confirmed in the other donors (Figure 4A, B). We verified the

transcriptional activity of the identified maturation-associated genes during  $\beta$ -cell maturation in single-cell data of human  $\beta$ -cell development from two studies [64,65] (Figure S5A-B, D-E). The expression of *WNT4*, *BMP5* and *PAK3* as well as *PAPSS2*, *LMO1*, *NCOR2*, *LIMCH1* and *EFNA5* and other identified factors was increased in immature  $\beta$ -cells compared to  $\beta$ -cell progenitors and precursors in fetal islets, which corroborates their role in  $\beta$ -cell maturation (Figure S5C, F).

Within the *mature*  $\beta$ -cell cluster, our analysis indicated a static and dynamic region of cells (Figure 3A,D). High velocity genes in the *mature* cluster were enriched for insulin secretion pathways and genes, which suggests that these dynamics describe transcriptional state changes from lower to higher insulin biosynthesis and/or secretion activity (Figure 3D, E, Supplementary Table 5). The high and low velocity clusters were also separated by *CD9* and *NPY* expression (Figure 3F). *CD9* has been proposed as a marker of functional  $\beta$ -cell heterogeneity, which together with *ST8SIA1* separates  $\beta$ -cells into four subpopulations [66]. Additional markers of *CD9*<sup>+</sup> and *CD9*<sup>-</sup> cells were differentially expressed between high and low velocity *mature* cells (Figure 3G). Within this classification scheme, *NPY* is a marker for *CD9*<sup>-</sup> *ST8SIA1*<sup>+</sup> cells, which showed higher glucose-stimulated insulin secretion capacity consistent with the transcriptional activity in insulin biosynthesis and/or secretion observed here. We found high and low velocity clusters with a similar marker expression profile also in the *mature* cluster of three out of the four other human donors (Figure S4C). In summary, our RNA velocity analysis predicts factors that promote possible state transitions in the continuous transcriptional  $\beta$ -cell landscape to and within mature  $\beta$ -cells. The predicted cellular flows from stressed/immature-like to mature and within mature cells indicate that these are likely interchangeable transcriptional states located along gene expression gradients and not stable  $\beta$ -cell subpopulations.

### 2.4. Human $\alpha$ -cell states

To describe molecular  $\alpha$ -cell heterogeneity in human islets, we refined the  $\alpha$ -cell clustering and identified four  $\alpha$ -cell states, which were represented in all 5 donors (Figure 4A–C). As for  $\beta$ -cells we used known marker genes and pathways as well as data-driven gene sets to annotate and characterize the  $\alpha$ -cell states (Figure 4D–F). We annotated a cluster of approximately 50% of the  $\alpha$ -cells as *mature* (Figure 4A–F). The *mature* cells highly expressed known  $\alpha$ -cell or endocrine identity and maturation factors as well as glucose transporters, hormone receptors, secretory-granule linked genes and ion channels important for  $\alpha$ -cell function (Figure 4D,E). These key markers as well as pathways linked to  $\alpha$ -cell function including glucagon secretion, insulin regulation and the mitochondrial respiratory chain were also captured by four  $\alpha$ -cell gene sets (G7–8, 12–13), which were activated in the *mature*  $\alpha$ -cells (Figure 4F, Supplementary Table 4). More than 30% of  $\alpha$ -cells showed an increase of PERK-mediated stress response genes and gene set scores and a decrease of  $\alpha$ -cell identity and function factors similar to *stress II*  $\beta$ -cells and were therefore annotated as *stress II*  $\alpha$ -cells (Figure 4F,G). 1% of cells were MHC-like  $\alpha$ -cells with increased MHC gene expression and gene set activation (Figure S2G). The remaining  $\alpha$ -cell cluster had an *immature* or precursor-like profile (Figure 4F,H). Multiple developmental markers including *SOX4*, *SOX11*, *NRG1*, *ID1-4*, *EPHB2* and *EPHB6* were increased, while  $\alpha$ -cell function genes were decreased. *Immature*  $\alpha$ -cells also activated gene sets enriched for TGF $\beta$  signaling, cell adhesion, ECM components, cytokines and interferon response (G2–5) as well as several direct transcriptional targets of the TGF $\beta$  signaling or interferon response pathway

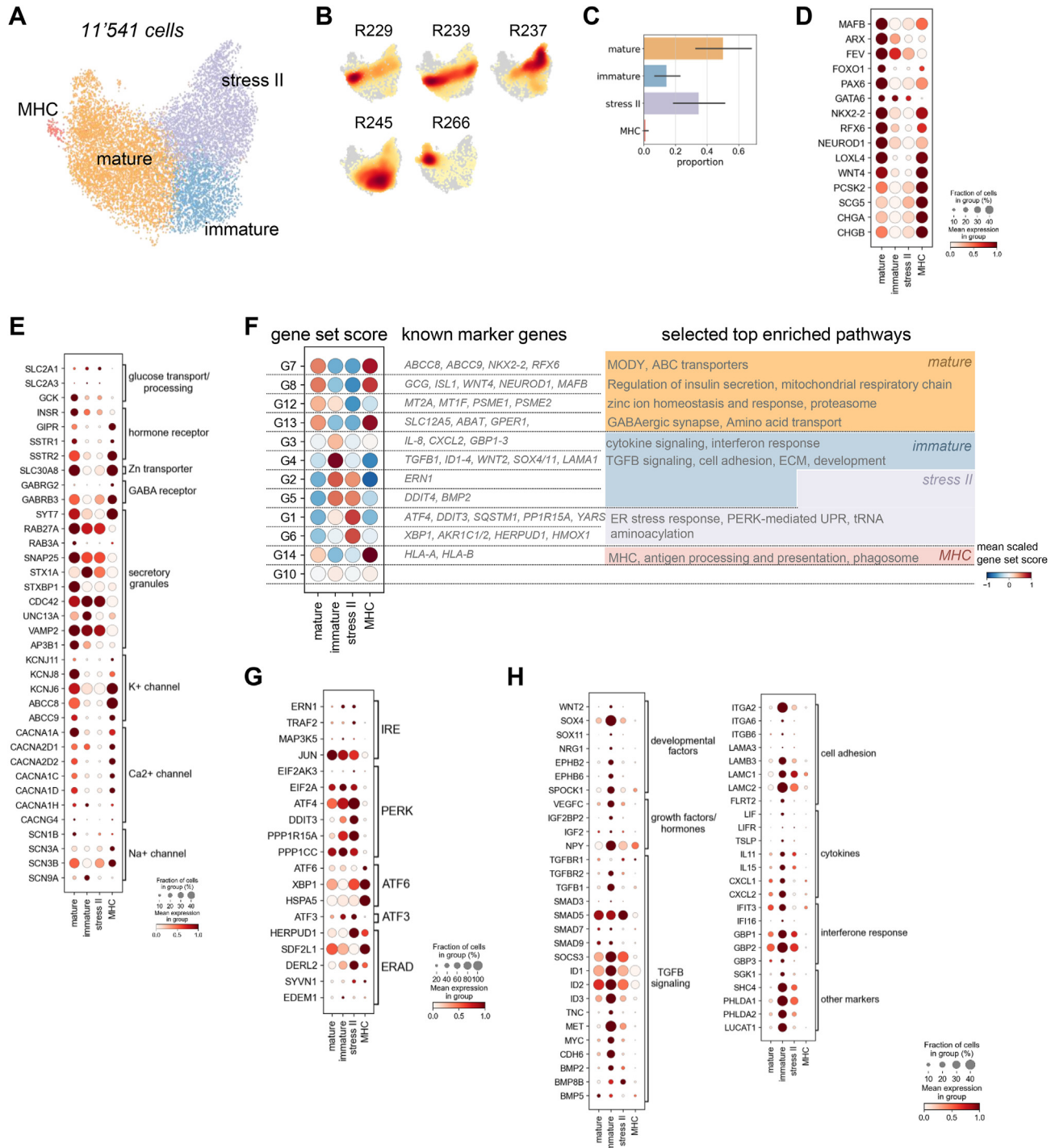


**Figure 3: Predicted transcriptional dynamics in human  $\beta$ -cell maturation and insulin secretion.** A) Cellular dynamics revealing areas of high induction and or repression of gene expression in  $\beta$ -cells of one human donor (R-ID 266). *Left:* Cell transitions are inferred from estimated RNA velocities and the direction of inferred movement plotted as streamlines on the UMAP. Colors indicate  $\beta$ -cell clusters. Circles highlight two areas of high velocity. Disconnected mtDNA deficient cluster was excluded. *Right:* UMAP showing the velocity of selected genes with increased velocity in the corresponding circled area. Top genes indicate induction of transcription of genes involved in  $\beta$ -cell function and insulin secretion. Bottom genes are associated with  $\beta$ -cell maturation. B) Velocity (top) and expression (bottom) of genes showing high velocity in immature  $\beta$ -cells along the cellular transition from immature to mature  $\beta$ -cells inferred from velocities. Cells were ordered by velocity pseudotime. Velocities and expression were scaled per gene. C) *Left:* Gene-resolved velocities of factors driving the transition from immature to the mature  $\beta$ -cell cluster. Purple lines indicate dynamics fitted with a full dynamical model. *Right:* Dotplot showing mean velocity per  $\beta$ -cell cluster. Selected known genes involved in  $\beta$ -cell maturation and potential novel genes important for maturation are shown. D) UMAP indicating two clusters of mature  $\beta$ -cells with high or low velocity. E) Selected top enriched Gene Ontology (GO) terms in high velocity genes of mature  $\beta$ -cells indicate induction of genes involved in insulin secretion. Gene enrichment was performed with EnrichR using a modification of the Fisher's exact test. F) Expression of two known markers of  $\beta$ -cell heterogeneity, CD9 and NPY, separates the two mature clusters in D). G) Expression of genes previously described to separate CD9+ and CD9-  $\beta$ -cells in high and low velocity mature  $\beta$ -cells. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

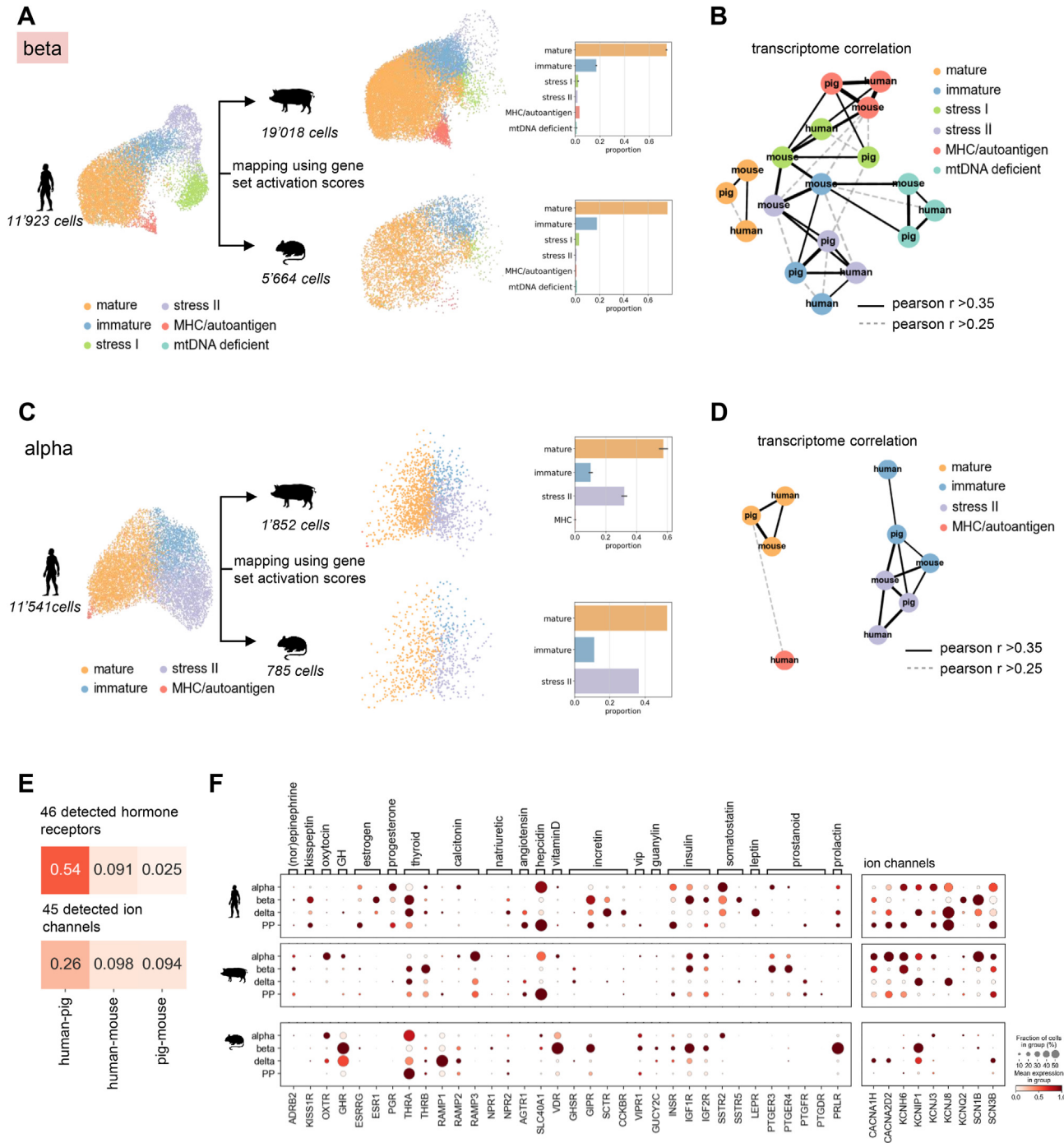
(Figure 4F,H). We verified activation of these gene sets in endocrine precursors and *immature*  $\alpha$ -cells in single-cell data of human pancreatic development [64,65] (Figure S5A, S6A,B). Fetal FEV<sup>+</sup> endocrine and  $\alpha$ -cell precursors scored higher than  $\alpha$ -cells for the *immature* and TGF $\beta$ -linked gene sets, but not for the inflammatory responses (Figure S6A, B). In addition, a subset of the identified markers of *immature*  $\alpha$ -cells were expressed in fetal precursors and  $\alpha$ -cells, which together confirms that parts of the profile of the *immature* adult  $\alpha$ -cell state resembles that of developing  $\alpha$ -cells (Figure S6B). Stress-linked  $\alpha$ -cells formed less distinct clusters than stress-linked  $\beta$ -cells (Figure S6C), which indicates that  $\alpha$ -cells were transcriptionally more homogenous and elicited a smaller stress response.

Finally, we leveraged published Patch-seq data to link the observed transcriptional states to  $\alpha$ -cell electrophysiology [4] (Figure S6D-G). Cells from healthy donors mapped to the *mature*, *immature* and *stress II* reference  $\alpha$ -cell states, hence these transcriptional states are robustly detected in different human data sets and donors (Figure S6D, E). Like in our reference map, *immature* cells had increased expression of developmental markers, TGF $\beta$  signaling and interferon response genes (Figure S6F). *Stress II* cells upregulated a canonical stress response (Figure S6F). In both *immature* and *stress II* cells Na<sup>+</sup> and Early Ca<sup>2+</sup> currents were decreased, while the other electrophysiological parameters were unchanged (Figure S6G). Molecular heterogeneity described by a set of marker genes was already associated with differences in Na<sup>+</sup> and Early Ca<sup>2+</sup> currents by [4]. Here, we





**Figure 4: Transcriptional  $\alpha$ -cell heterogeneity and states in human islets.** A) UMAP plot of 11'541 human  $\alpha$ -cells. Colors highlight clustering into four different  $\alpha$ -cell states. B) Cell densities in UMAP space for five human donors shows that all  $\alpha$ -cell clusters are represented by all donors. ID indicates donor ID for ADI IsletCore (see also Figure S1A). C) Fraction of cells per  $\alpha$ -cell clusters. Error bar indicates donor variation. D-G) Characterization of  $\alpha$ -cell clusters. D-E, G-H) Expression of selected known  $\alpha$ -cell identity and maturity markers (D), functional markers (E), adaptive stress response genes (G) and genes involved in pathways describing immature  $\alpha$ -cells (H). Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene. F) Gene sets capturing variation in human  $\alpha$ -cells that describe biological processes. *Left*: Scaled mean score for each gene set per  $\alpha$ -cell cluster. For each gene set selected  $\alpha$ -cell identity or functional marker genes are indicated. *Right*: Summary of selected enriched pathways for each gene set indicating biological processes associated to gene sets. Coloring indicates the highest scoring  $\alpha$ -cell cluster.



**Figure 5: Cross-species mapping of  $\alpha$ - and  $\beta$ -cell states.** A-D) Cross-species mapping of  $\alpha$ - and  $\beta$ -cells (A, C) by gene set activation scores. UMAP plot (left) shows human cells, where each cell is represented by an activation score of the corresponding cell gene sets. Pig and mouse cells were mapped to the human reference data through projecting on the human gene set representation. Embedding and labels are mapped using the Scanpy ingest functionality (see Methods). The barplot indicates the frequencies of mapped clusters for pig and mouse. B, D) Graph showing global transcriptome correlation of  $\beta$ - (B) and  $\alpha$ - (D) cell clusters across species. Edge weights indicate Pearson correlation coefficient (see also Figure S4B). Nodes are colored by  $\beta$ -cell clusters. E) Pairwise correlation of the expression pattern across endocrine cell states computed using detected hormone or hormone-like receptors (top) or ion channels (bottom).  $\alpha$ - and  $\beta$ -cells were subset to mature state. List of hormone receptors was manually curated. List of ion channels contains calcium, sodium, potassium and transient receptor potential ion channels. Pearson correlation is computed using the harmonic average of mean expression and fraction of cells expressing a gene in a group across all cell types (see Methods). Pearson correlation coefficient is indicated. F) Expression of selected hormone receptors (left) and ion channels (right) showing differential expression patterns in endocrine cell states across species.  $\alpha$ - and  $\beta$ -cells were subset to mature state. Hormone and peptide ligands for receptors are indicated. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

established that two transcriptionally distinct states may underlie this functional  $\alpha$ -cell heterogeneity highlighting two potential routes that lead to decreased function.

### 2.5. Cross-species mapping of human $\alpha$ - and $\beta$ -cell heterogeneity

Gene sets are a data representation, which captures the human  $\alpha$ - and  $\beta$ -cell biology but removes species- or batch-specific details and overcomes technical artifacts like the limited annotation and capture rate in pig. If one assumes that the subset of mappable genes is sufficient to indicate activation of the full gene set, the gene set space corresponds to normalizing the data per functional gene set unit. To assess conservation of the human  $\alpha$ - and  $\beta$ -cell states, we represented each cell as an activation score of the human  $\alpha$ - or  $\beta$ -cell gene sets, respectively, and projected mouse and pig cells to the human reference map (Figure 5A,C).

The majority of pig and mouse  $\beta$ -cells mapped to the *mature* human reference cluster and scored high for the identified maturity gene sets (Figure 5A, Figure S7A). The mapped *mature* cells highly expressed  $\beta$ -cell identity and maturity markers and their gene expression profiles strongly correlated with the human *mature* profile (Figure 5B, Figure S7B, C), which validates our gene set-based mapping strategy. A smaller cluster of pig and mouse cells resembled *immature* cells and showed decreased levels of maturity gene set scores and markers (Figure 5A, Figure S7A, B). Moreover, a small fraction of cells mapped to the *stress I* and *stress II* references (Figure 5A). In mice, the expression profiles of *immature*, *stress I* and *stress II* correlated stronger with each other, cells clustered more tightly, and activation level differences of markers and gene sets were smaller than for human and pig  $\beta$ -cells (Figure 5B, Figure S7A-D). For example, multiple stress response genes including *ATF3*, *DDIT3*, *PPP1R15A*, *HSPA1B*, *DNAJB1*, *SYNV1*, *DERL3*, *FKBP11*, *SXRN1* were expressed in most *mature* and non-mature mouse  $\beta$ -cells, while they were more specifically increased in *stress I* or *stress II* clusters of pig and human  $\beta$ -cells (Figure S7D). Hence, mouse  $\beta$ -cells were more homogeneous than human and pig  $\beta$ -cells and adopted a *mature* or immature-like state with high basal expression of stress-response factors but not a distinct stress-associated state. Lastly, we identified in both pig and mouse  $\beta$ -clusters that mapped to the *MHC/autoantigen* human  $\beta$ -cells, which activated the MHC/autoantigen-associated gene set (G9) and decreased the ribosome/translation-associated gene set (G4), and whose profiles strongly correlated with their human counterparts. This indicates that the *MHC/autoantigen*  $\beta$ -cell state is evolutionarily conserved.

Pig and mouse  $\alpha$ -cells mapped to *mature*, *immature* and *stress II* reference states and were similarly distributed as human  $\alpha$ -cells (Figure 5C). In *mature* cells identity and maturation markers as well as maturity gene set activation were conserved and their transcriptomes correlated across species (Figure 5D, Figure S7E-G). The transcriptomes of *immature* and *stress II* cells correlated strongly across and within species (Figure 5D, Figure S7F). Like in human  $\alpha$ -cells, *immature* cells had increased activation of TGF $\beta$ -associated genes and a subset of other developmental factors (Figure S7H). However, we did not detect increased cell adhesion/ECM factors or an inflammatory response in pig and mouse cells. Similar to  $\beta$ -cells, in mouse expression level differences of stress-associated genes were smaller and *stress II* cells less distinct from mature/immature cells than in human and pig (Figure S7G, I). To confirm that the cross-species comparison and observed states are robust across datasets we mapped  $\alpha$ - and  $\beta$ -cells of three additional healthy mice [67] to our human references (Figure S8, Methods). For both  $\alpha$ - and  $\beta$ -cells, detected states and state fractions (Figure S8A,C) and gene set

activation (Figure S8B,D) were consistent with results observed for the mouse data used in this study. Together, our analyses suggest that the spectrum of human transcriptional  $\alpha$ - and  $\beta$ -cell heterogeneity including stress-associated states were better captured in our pig than mouse data.

Finally, we investigated conservation of the transcriptional profile of human *mature* states. We first focused on mappable genes within the  $\alpha$ - and  $\beta$ -maturity gene sets, respectively. Of these genes more than 60% were conserved in *mature*  $\beta$ -cells and more than 70% were conserved in *mature*  $\alpha$ -cells of pigs and mice (Figure S7J). Moreover, putative human  $\beta$ -cell maturation factors identified by RNA velocity analysis were expressed in mouse and pig *mature*  $\beta$ -cells (Figure S7K). Finally, to approximate conservation of hormone/peptide signaling and excitability in mature cells we explored hormone or hormone-like receptors and ion channels in mature  $\alpha$ - and  $\beta$ -cells and the other endocrine cell types  $\delta$ - and PP-cells. Overall, the expression patterns across endocrine cell types of both detected hormone receptors and ion channels (calcium, potassium, sodium and transient receptor potential ion channels) correlated stronger between human and pig than human and mouse (Figure 5E). Differentially expressed receptors in mouse when compared to human islets included for example the *prolactin receptor (PRLR)*, *leptin receptor (LEPR)*, *Vitamin D receptor (VDR)*, *growth hormone receptor (GHR)*, *Natriuretic peptide receptor A (NPR1)*, *Estrogen receptor 1 (ESR1)*, *Progesterone receptor (PGR)*, *Vasoactive intestinal polypeptide receptor (VIPR)*, *guanylate cyclase-C receptor (GUCY2C)*, *secretin receptor (SCTR)*, *prostanoid receptors (PTGER3, PTGER4, PTGFR)* as well as *ferroportin (SLC40A1)* (Figure 5F). *PRLR*, *VDR*, *VIPR*, *NPR1*, *GUCY2C* and *GHR* were highly expressed in mouse but low or absent in pig and human mature  $\beta$ -cells and instead detected in other endocrine cell types. Similarly, *ADRB2*, *PGR* and *ESR1* were expressed in human but not in mouse  $\beta$ -cells, and, *ADRB2* but not *PGR* and *ESR1* was also detected in pig  $\beta$ -cells. We confirmed that all of these receptors were unique or enriched in mouse or human  $\beta$ -cells, respectively, in bulk  $\beta$ -cell transcriptomes of human and mouse islets [36]. Surprisingly, pig  $\beta$ - and  $\alpha$ -cells expressed *PTGER3* and *PTGER4*, which in mice have been reported as  $\beta$ -cell dedifferentiation markers. Especially, *PTGER3* was strongly upregulated in STZ-treated diabetic  $\beta$ -cells (Figure S7L). In humans, *PTGER3* and *PTGER4* were expressed in  $\alpha$ -cells. Ion channels with differential expression in mouse and human  $\beta$ -cells included potassium channel *KCNJ8*, sodium channel *SCN3B* and calcium channels *CACNA1H* and *CACNA2D2* (Figure 5F). *KCNJ8* was expressed in all human endocrine cell types and in pig  $\delta$ -cells but not detected in mice. *SCN3B*, *CACNA1H* and *CACNA2D2* were expressed in all human and pig cell types, but only in mouse  $\delta$ -cells. Like prostanoid receptors, these channels were increased in diabetic  $\beta$ -cells of STZ-treated mice (Figure S7L).

In summary, the identified species-specific expression patterns of hormone receptors and ion channels suggest that these functional genes are better conserved in pig than mouse endocrine cells. Moreover, they exemplify the value of this data resource to explore differences between human and two commonly used animal models.

### 3. DISCUSSION

Our single-cell data of human, pig and mouse endocrine islet cells is a foundational resource for advancing our understanding of human endocrine heterogeneity and its conservation in clinically relevant animal models. We characterized a compendium of human transcriptional  $\alpha$ - and  $\beta$ -cell states, which represent a reference to investigate endocrine cell function, maturation and disease-associated

phenotypes. The distinct non-mature  $\alpha$ - and  $\beta$ -cell states (immature/stress/MHC) do not necessarily represent cells found as such *in vivo* in healthy patients, but likely have been induced during tissue isolation, processing, storage and transport. Moreover, the *in silico* predicted transcriptional dynamics indicate that these states are likely physiological and interchangeable states different from stable sub-populations, which transition only upon specific signaling cues and can be followed by lineage tracing [38]. Nevertheless, the captured cell states model mature, functional  $\alpha$ - and  $\beta$ -cells as well as different types of endocrine cell stress. For example, our analyses revealed novel putative  $\beta$ -cell maturation markers (e.g. *NCOR2*, *LIMCH1*, *EFNA5*) and a distinct, conserved immature  $\alpha$ -cell state with increased expression of developmental markers (e.g. *WNT2*, *SOX4*, *SOX11*), members of the TGF- $\beta$  signaling pathway (e.g. *TGFB1*, *ID1-3*, *SOCS3*, *TNC*), integrins (e.g. *ITGA2*, *ITGA6*) and a cytokine response. Endocrine precursor cells of fetal human islets share parts of the transcriptional profile of immature-like  $\alpha$ -cells [64]. Stressed  $\alpha$ - and  $\beta$ -cells differentially express markers of hormone biosynthesis and secretion and regulatory hormone receptors and match cells with divergent electrophysiological properties, which may mirror aspects of the pathological phenotype reported for type 1 and type 2 diabetic islet cells [68]. We found that  $\beta$ -cells responded diversely to the multiple exogenous stressors they were exposed to during processing and described three distinct states linked to stress. These included a rare, but conserved  $\beta$ -cell state with a reduced expression of factors governing general transcription and translation, but increased MHC-class I and antigen expression. This suggests that in a state of high stress, in which global transcription is diminished,  $\beta$ -cells can maintain expression of identity genes and enhance antigen presentation activity, of which the latter is a gene program also observed in  $\beta$ -cells of T1D patients. Overall, we hope that this comprehensive human islet cell map will guide future hypotheses on the control and molecular basis of functioning islet cells and their response to stress, while also informing the path to successful therapeutic reestablishment of islet cell function in diabetic patients.

Despite correlation of whole transcriptional profiles and TF expression patterns of cell states, the conservation of human gene expression is surprisingly low (50–60%). We may have underestimated conservation due to detection limits inherent to single-cell RNAseq data and, for pig, due to the sparser coverage and annotation of the genome. Nonetheless, our findings suggest that large parts of gene expression patterns are evolutionarily labile, while important identity and functional marker genes and TF expression patterns are conserved. This is consistent with previous reports that showed similarly low conservation of cell type enriched genes between human, mouse and zebrafish [69]. These species-differences likely do not result in altered functional or phenotypic cell states, but they can become relevant in animal studies designed to identify pathological programs and clinical targets. Our analyses provide evidence that pigs can be a surrogate model of gene expression relevant for human endocrine cell function. We showed that, overall, expression and cell type-specificity of regulatory units like TFs, hormone/peptide signaling and cell excitability are better mirrored in pig than mouse islet cells. For example, mature human and pig  $\alpha$ - or  $\beta$ -cells shared functional regulators not observed in mouse, which included the TFs *ID1-4*, the surface hormone receptors *ADRB2* and *PTGER3* and the ion channels *SCN3B*, *CACNA1H* and *CACNA2D2*. These examples correspond well with reported differences between human and mouse  $\beta$ -cells [36], and illustrate the value of this data resource to reveal species-specific expression of targets governing glucose sensing and hormone secretion and to complement existing data sources of humans and mice. Finally, we observed that in our data

the extent of human transcriptional  $\alpha$ - and  $\beta$ -cell heterogeneity - especially expression gradients of stress-associated genes - is better conserved in pigs than in mice. While  $\alpha$ - and  $\beta$ -cells of all three species adopted mature and more immature-like states, only human and pig cells formed distinct stressed states. In mice, stress-response factors (e.g. *DDIT3*, *PPP1R15A*, *DERL3*, *ATF3*, *DNAJB3*, *HSPA1B*) were expressed more homogeneously with a high basal level even in the mature state.

Altogether, our cross-species islet map provides a framework for investigating the transcriptional programs of human endocrine cells and represents a FAIR data resource [26] that can inform future studies where mouse and pig will fail to model human islet biology.

## 4. METHODS

### 4.1. Cell sources

Primary human islets were obtained from the IsletCore facility (Edmonton, AB, Canada) with informed consent. Detailed donor information can be accessed via <https://www.epicore.ualberta.ca/isletcore/> using the R-IDs indicated for each donor in Figure S1.

A female retired breeder Göttingen minipig (age: 3 years, 8 months) was purchased from Ellegaard (Denmark) and housed under standard conditions (19–23 °C; 40–70% relative humidity; 12:12 h day/night cycle). Pancreas retrieval and islet isolation was conducted as previously described [70]. Briefly, pancreas was preserved in Custodiol®-HTK solution for 2.5 h (cold Ischemia time). For islet isolation cold perfusion solution (Corning®, NY, USA) with Collagenase NB8 (4 U/g tissue), neutral protease (0.4 U/g tissue; both Serva, Heidelberg, Germany) and 100 mg DNase (Roche Diagnostics, Mannheim, Germany) were infused into the pancreatic duct. The digestion was performed by a modified Ricordi method at low temperature (34 °C) and with minimal mechanical force. Islets were separated from exocrine tissue by centrifugation on a discontinuous Ficoll (Sigma–Aldrich, Taufkirchen, Germany) density gradient in a COBE 2991 cell processor (Terumo BCT). After purification, islets were cultured in CMRL 1066 medium supplemented with 10% heat inactivated FBS, 100U/mL penicillin, 0.1 mg/ml streptomycin (all Gibco®, Darmstadt, Germany) and 32.5 mM L-glutathione (Sigma–Aldrich, Taufkirchen, Germany) at 37 °C in a 5% CO<sub>2</sub> incubator.

### 4.2. Single-cell suspension

To obtain a single-cell suspension of human and pig islets, 60 islets were hand-picked into a 1.5 ml Eppendorf tube, pelleted (280 g, 1 min), washed with PBS (minus Mg or Ca, Gibco) and digested with Tryp-LE (Gibco) at 37 °C for 12 min. During the incubation step with Tryp-LE, islets were mechanically disaggregated with a 1 ml pipet tip every 2–3 min. The digestive reaction was then stopped by adding FACS-buffer (PBS, 2% FCS, 2 mM EDTA) and cells were pelleted (280 g, 3 min). Cells were stained with trypan blue to visualize dead cells and counted with a hemocytometer.

### 4.3. Single-cell sequencing

Single-cell libraries were generated using the Chromium Single Cell 3' library and gel bead kit v2 (PN #120237) from 10x Genomics. Briefly, we targeted 10'000 cells per sample by loading 16,000 cells per sample onto a channel of the 10x chip to produce Gel Bead-in-Emulsions (GEMs). This underwent reverse transcription to barcode RNA before cleanup and cDNA amplification followed by enzymatic fragmentation and 5' adaptor and sample index attachment. Libraries were sequenced on the HiSeq4000 (Illumina) with 150 bp paired-end sequencing of read2.

#### 4.4. Preprocessing and quality control of scRNA-seq data

For human and pig single-cell samples, the Cell Ranger analysis pipeline (v2.0.0) provided by 10x Genomics was used to demultiplex binary base call (BCL) files, to align and filter reads and to count barcodes and unique molecular identifiers (UMI). Barcodes with high quality were selected based on the distribution of total UMI counts per cell using the standard Cell Ranger algorithm for cell detection. All downstream analyses were run with python3 (v $\geq$ 3.5) using the Scanpy package [71] (v $\geq$ 1.4, <https://github.com/theislab/scanpy>) except stated differently. Python package versions that may affect numerical results are indicated in the available jupyter notebooks (See Data and Code availability). Genes with expression in less than 20 cells were excluded. Low quality or outlier cells were removed if the fraction of mitochondria-encoded counts was above 20%; (2) and based on total UMI counts and total genes. In human samples, thresholds were defined per sample after visual inspection of the total UMI count and total gene distributions as recommended [72] (for threshold values, see Data and Code availability and provided analysis notebooks). Cell-by-gene count matrices of all samples of one species were then concatenated to a single matrix. To account for differences in sequencing depth, UMI counts of each cell were normalized using the SCRAN algorithm [73] as implemented in the scan R package [74] and values were log-transformed (log (count+1)). Sample differences in human and pig samples were corrected as recommended [75] using the python implementation of ComBat [76] (<https://github.com/brentp/combat.py>) adopted by Scanpy (pp.combat) with default parameters and specifying each sample as one batch. Zero values were kept as zero even after correction to avoid spurious sample-to-sample differences around zero.

For mouse single-cell data [11] the filtered and annotated raw count matrix was downloaded from the Gene Expression Omnibus (GEO) (GEO accession number: GSE128565). The raw count matrix was filtered by subsetting to cells present in the filtered count matrix. Counts of each cell were normalized by total counts of that cell (pp.normalize\_total with *exclude\_highly\_expressed* = True). Highly expressed genes (genes with more than 5% of total counts in a cell) were excluded from total counts for each cell before normalization. Counts were then log-transformed (log (count+1)).

These count matrices were used as input for further analyses unless indicated. Data from each species was analyzed separately until cross species mapping described below. Custom scripts with source code for all analyses of scRNA-seq data are available as jupyter notebooks in a github repository ([https://github.com/theislab/2022\\_Tritschler\\_pancreas\\_cross\\_species](https://github.com/theislab/2022_Tritschler_pancreas_cross_species)) and the scRNA-seq data can be explored in the cellxgene data portal (<https://cellxgene.cziscience.com/collections/0a77d4c0-d5d0-40f0-aa1a-5e1429bcbd7e>).

#### 4.5. Single cell manifolds, clustering and annotation

The manifolds and clusterings for the human, pig and murine endocrine cells and the human  $\alpha$ - and  $\beta$ -cells were computed separately by performing the following steps. A single-cell neighborhood graph (kNN-graph was computed on the top principal components: 50 first for endocrine cells and  $\alpha$ -cells, 25 first for  $\beta$ -cells) using 15 neighbors. Genes with expression in less than 10 cells were excluded. To calculate the principal components top highly variable genes were used as identified by the highly\_variable identification function in Scanpy (pp.highly\_variable, top 4000 for mouse endocrine cells, top 2000 for others). Clustering was performed using louvain-based clustering [77] as implemented in louvain-igraph (v0.6.1 <https://github.com/vtraag/louvain-igraph>) and adopted by Scanpy (tl.louvain). The resolution parameter was varied in different parts of the data

manifold to account for strong changes in resolution (for details, see Data availability and provided analysis notebooks). For single-cell manifolds and visualization UMAP was run as recommended [78] and adopted by Scanpy. From the initial data mono-hormonal endocrine cells were annotated based on expression of genes encoding the four main islet hormones: insulin for  $\beta$ -cells, glucagon for  $\alpha$ -cells, somatostatin for  $\delta$ -cells, pancreatic poly-peptide for PP cells and ghrelin for epsilon cells. Clusters expressing known markers of non-endocrine cells (for example *SPP1* for ductal cells, *PRSS2* for acinar cells, *PLVAP* for endothelial cells, *PTPRC* for immune cells or *COL1A1* for fibroblasts and stellate cells), cells identified as doublets based on scores computed with the Scrublet algorithm [79] (v0.2.1, <https://github.com/AllonKleinLab/scrublet>) and co-expression of marker genes, and polyhormonal cells expressing multiple pancreatic hormones were excluded.  $\alpha$ - And  $\beta$ -cell states were annotated as described in the main text. Clusters expressing the same hormones, markers or gene sets ( $\alpha$ - and  $\beta$ -cell states) were merged (see also Data availability and provided analysis notebooks).

#### 4.6. Gene orthologue mapping

To identify the genes mappable between species we used the R-based biological entity dictionary (BED). Briefly, first, ensembl gene names of pig samples were converted to human and mouse ensembl gene names, and then subset to the genes shared across species, detected in the data and with an ID set as preferred by the BED tool. For genes that did not map 1:1 between pig and human or pig and mouse (approximately 5% of all genes) the gene with the maximal expression in the corresponding species-data was kept. The list of mappable and detected genes is provided in the github repository ([https://github.com/theislab/2022\\_Tritschler\\_pancreas\\_cross\\_species/BED\\_mapping\\_genes.csv](https://github.com/theislab/2022_Tritschler_pancreas_cross_species/BED_mapping_genes.csv)).

#### 4.7. Marker gene detection and comparison

Enriched marker genes of endocrine cell types were identified by comparing the mean expression of cells of one cell type to the mean expression of cells in all other cell types within each species. Genes that were expressed in at least 5% of the cells of the cell type and were increased by at least 1.4 fold (log<sub>2</sub> (fold change) > 0.5) were defined as enriched marker genes.

#### 4.8. Correlation based-gene sets of human $\alpha$ - and $\beta$ -cells

Gene sets of human  $\alpha$ - and  $\beta$ -cells were identified by clustering the top 3000 variable genes based on their pairwise—pearson correlation values across human  $\alpha$ - or  $\beta$ -cells, respectively, as previously described in [39] to identify *de novo* gene sets. Genes detected in less than 20  $\alpha$ -/ $\beta$ -cells were excluded. Clustering was performed using Ward's method and euclidean distance as implemented in the scipy python package [80] (v.1.5.4). Functional enrichment of gene sets was performed as described below. Gene sets with very low average correlation (<0.005) were excluded from downstream analyses.

#### 4.9. Similarity of gene expression patterns

Similarity of gene expression patterns was estimated by pearson correlation coefficients of gene expression across cell types or states to account for cell type or state-specificity. To leverage all information gained from single cell resolution, Pearson correlation coefficients were computed using the harmonic average of mean expression and fraction of cells expressing a gene in a group across all cell types. To account for differences in detection limits and sequencing depth the fraction of cells expressing a gene in a group was normalized to the mean fraction per group and species.

#### 4.10. Pathway and transcription factor sources and pathway enrichment

Pathway enrichment of gene lists and sets was performed using EnrichR [81] as adopted by the enrichr functionality in the gseapy package (<https://github.com/zqfang/GSEAPy>). To evaluate hallmarks and stress pathway activations, hallmark and ontology gene sets were downloaded from the Molecular Signatures Database v7.2 of the Broad Institute. To identify transcription factors within gene lists a list of human transcription factors was downloaded from the Human Transcription Factor Database [82] (<http://bioinfo.life.hust.edu.cn/HumanTFDB>, v1.01).

#### 4.11. Gene set activation and cell scores

Gene set or pathway activation in a cell was computed using the cell scoring function described by [83] and implemented in Scanpy (`tl.score_genes`). Briefly, the activation score of a cell is the average expression of genes of the gene set in a cell subtracted with the average expression of genes of a randomly sampled background set with expression values within the same range.

#### 4.12. Characterization of T1D $\beta$ -cells

Raw count matrices of cells from healthy and T1D patients generated by [42] were downloaded from GEO (Accession number GSE121863). Genes expressed in less than 10 cells were excluded. Raw counts of each cell were normalized by total counts of that cell not considering highly expressed genes for the total count normalization factor of a cell (`pp.normalize_total` with `exclude_highly_expressed = True`) and log-transformed (`log(count+1)`). Mono-hormonal  $\beta$ -cells were identified by iterative clustering and annotation as described above. The T1D  $\beta$ -cell score was computed based on the top 50 differentially expressed genes between  $\beta$ -cells from healthy and T1D donors (Welch's t-test, `tl.rank_genes_groups`).

#### 4.13. Characterization of fetal human precursor $\alpha$ - and $\beta$ -cells

Raw count matrices generated by [64] were downloaded from the data visualization center descartes (<https://descartes.brotmanbaty.org/bbi/human-gene-expression-during-development/>). The `rsd`-file was loaded into R and an AnnData object was generated for downstream analysis with the `ry2` (v3.3.5, <https://github.com/ry2/ry2>) and `anndata2ri` (v1.0.4, <https://github.com/theislabs/anndata2ri>) python packages. Raw count matrices generated by [65] using the 10X Genomics technology were downloaded from OMix (<https://bigd.big.ac.cn/omix/>) using the identifier OMIX236. An AnnData object was generated for downstream analysis.

Both datasets were processed and analyzed following the same steps: Genes expressed in less than 10 cells were excluded. Raw counts of each cell were normalized by total counts of that cell. Highly expressed genes in a cell were not considered for the total count normalization factor of that cell (`pp.normalize_total` with `exclude_highly_expressed = True`). Counts were then log-transformed (`log(count+1)`). Pancreatic cell types and endocrine clusters were identified by clustering and annotation using markers described above. To distinguish epithelial from mesenchymal cell clusters the markers `EPCAM` and `VIM` were used. In [65], to detect neuronal or neuroendocrine cell clusters `ASCL1` was used, for trunk and ductal clusters `HES1`, `SAT1` were used, and for tip and acinar clusters `CTRB1`, `GP2`, `RBPJL` were used. Endocrine progenitors were identified based on the expression of progenitor marker genes `SOX4` and `NEUROG3`, precursors using marker gene `FEV` and `PAX4` ( $\beta$ -cell lineage) and `ARX` ( $\alpha$ -cell lineage) amongst others.

#### 4.14. Inference of $\beta$ -cell dynamics using RNA velocity

To infer cellular dynamics in  $\beta$ -cells, RNA velocities were estimated for each human donor with a steady-state model as initially proposed by [60] and adopted and extended by [59] and in the `scVelo` python package (v0.2.2, <https://github.com/theislabs/scvelo>). Splicing information of reads (spliced/unspliced) was extracted from the `bam`-files using the `velocyto` pipeline (<http://velocyto.org>). The resulting loompy file was then read into an AnnData object for downstream analysis with `scVelo` and `Scanpy`. To estimate velocities and infer cellular transitions the following steps were performed as recommended. First, genes were filtered with shared spliced and unspliced expression in less than 10 cells, the spliced and unspliced count layers were normalized to the initial total count per cell and log transformed (`log(count+1)`), and top 4000 variable genes were selected. Next, first- and second-order moments were calculated for each cell across its nearest neighbors of a kNN in principal components space (number of neighbors = 30, number of PCs = 30). Then velocities were estimated by fitting a steady-state model of transcription for each gene. Finally, a velocity graph was computed from the cosine similarities between the cell state change predicted by the velocity vector and possible cell transitions in the kNN. To compute the graph only genes with a likelihood >0.1 were considered. Using this graph the estimated velocities were then projected to the original UMAP space. To identify enriched velocity genes in mature and immature cells a differential expression test on velocities was applied comparing the velocity of one to all other clusters (Welch t-test with overestimated variance, `tl.rank_velocity_genes`). The velocity pseudotime was computed based on the directed velocity graph as implemented in `scVelo` (`tl.velocity_pseudotime`). The velocity pseudotime is a directed random-walk based distance measure between cells.

#### 4.15. Cross-species mapping of $\alpha$ - and $\beta$ -cell states

Mouse and pig  $\alpha$ - and  $\beta$ -cells were mapped separately onto the human  $\alpha$ - and  $\beta$ -cell reference states using the `Scanpy` `ingest` functionality (`tl.ingest`). Briefly, genes were subset to mappable genes and cells were scored for activation of the identified human gene sets. The gene set score matrix was scaled to standard variation (`pp.scale`). A single-cell manifold was then computed for human cells in gene set space applying the UMAP algorithm on the calculated kNN in PC space. Mouse and pig cells were mapped to the human reference through projecting to the PC space of the human cells. To map the single-cell embedding the UMAP package is used. Cell type labels are mapped using a kNN classifier.

Additional publicly available mouse data to confirm the cross species mapping were downloaded from GEO with accession number GSE162512 [67] and an AnnData object was generated. Cells with less than 200 total counts or 200 total genes expressed were filtered. Genes expressed in less than 10 cells were excluded. Raw counts of each cell were normalized by total counts of that cell not considering highly expressed genes for the total count normalization factor of a cell (`pp.normalize_total` with `exclude_highly_expressed = True`) and log-transformed (`log(count+1)`). Single-cell manifold generation, clustering and cluster annotation were performed as described above for the data of this study using top 2000 highly variable genes, 50 top principal components, a neighborhood size of 15 and known marker genes.

#### 4.16. Mapping of Patch-Seq data to $\alpha$ - and $\beta$ -cell states

Raw count matrices and metadata files including cell type annotations of Patch-Seq data from  $\alpha$ - and  $\beta$ -cells generated by [4] were

downloaded from <https://github.com/jcamunas/patchseq/tree/master/data>. An AnnData object was generated from the text-files for downstream analysis. Genes expressed in less than 5 cells or with less than 10 total counts were excluded. Raw counts of each cell were normalized by total counts of that cell. Counts were then log-transformed ( $\log(\text{count}+1)$ ). Data was subset to  $\alpha$ - and  $\beta$ -cells using the provided cell type labels and mapped to our human reference states as described above for the cross-species mapping. Genes in gene sets were subset to 15'864 overlapping genes between the two studies before scoring. The data was then subset to patch-clamped cells from healthy donors. Cell states with  $<3$  cells were excluded.

#### 4.17. Mapping of 9 publicly available datasets to $\beta$ -cell states

Raw count matrices and metadata of publicly available single-cell RNAseq datasets of pancreatic islets of healthy human donors were downloaded from GEO from accession numbers GSE114297 [9], GSE84133 [13], GSE86469 [55], GSE85241 [14], GSE81547 [25], GSE183568 [56], GSE101207 [58], and the cellxgene data portal (<https://cellxgene.cziscience.com/collections/51544e44-293b-4c2b-8c26-560678423380>) [57]. An AnnData object was generated for downstream analysis. Cells with less than 200 total counts or genes expressed were filtered. Genes expressed in less than 10 cells were excluded. Raw counts of each cell were normalized by total counts of that cell not considering highly expressed genes for the total count normalization factor of a cell (`pp.normalize_total` with `exclude_highly_expressed = True`) and log-transformed ( $\log(\text{count}+1)$ ). Additionally, the processed count matrix was downloaded from ArrayExpress (EBI) with accession number E-MTAB-5061 [23], an AnnData object was generated and counts were log-transformed ( $\log(\text{count}+1)$ ).

Single-cell manifold generation, clustering and cluster annotation were performed as described above for the data of this study using top 2000 highly variable genes, 50 top principal components, a neighborhood size of 15 and known marker genes. For GSE81547 [25] and GSE101207 [58] data of individual donors was integrated before computing the UMAP and clusters using the BBKNN alignment method [84]. For datasets from E-MTAB-5061 [23], GSE84133 [13,57] original cell type labels were kept.

The datasets of each study were then subset to  $\alpha$ - and  $\beta$ -cells using the cell type labels and mapped to our human reference states as described above for the cross-species mapping. Genes in gene sets were subset to genes overlapping with this study before scoring.

#### 4.18. Data and code availability

Annotated single-cell data can be explored and queried in the cellxgene data portal (<https://cellxgene.cziscience.com/collections/0a77d4c0-d5d0-40f0-aa1a-5e1429bcbd7e>) and were added to the sfaira data zoo [27]. Pig data was mapped and subset to human genes in the cellxgene portal. Raw data and count matrices of scRNA-seq data are available on GEO (accession number: GSE198623). Custom python scripts written for performing scRNA-seq analysis are available as jupyter notebooks in a github repository ([https://github.com/theislab/2022\\_Tritschler\\_pancreas\\_cross\\_species](https://github.com/theislab/2022_Tritschler_pancreas_cross_species)). Python package versions that may affect numerical results as well as specific parameters and threshold values for all analyses are indicated in the scripts.

#### CREDIT AUTHOR CONTRIBUTIONS

S.T.: Conceptualization, Methodology, Software, Data curation and analysis, Visualization, Writing- Original draft; M.T.: Software, Data curation and analysis, Writing- Reviewing and Editing, A.B.:

Investigation, B.L.: Resources; J.S.: Investigation, Resources; U.S.: Investigation, Resources; E.K.: Investigation; E.W.: Supervision; H.L.: Conceptualization, Supervision, Resources, Funding acquisition, Writing- Reviewing and Editing; F.J.T.: Conceptualization, Supervision, Resources, Funding acquisition, Writing- Reviewing and Editing

#### DATA AVAILABILITY

Data and source code were made publicly available on GEO (accession number: GSE198623) and in a github repository ([https://github.com/theislab/2022\\_Tritschler\\_pancreas\\_cross\\_species](https://github.com/theislab/2022_Tritschler_pancreas_cross_species)). Data can be explored and queried in the cellxgene data portal (<https://cellxgene.cziscience.com/collections/0a77d4c0-d5d0-40f0-aa1a-5e1429bcbd7e>).

#### ACKNOWLEDGEMENT

We thank the Alberta Diabetes Institute IsletCore for the human islets, F. A. Wolf for fruitful discussions and constructive feedback on the computational analysis, and K. Hrovatin and S. Sachs for reviewing the manuscript and figures. Moreover, we thank D.S. Fischer and the cellxgene team, specifically J. Cool, J. Hilton, J. Yu-Sheng Chien and B. Aevermann, for their support with publishing the data in sfaira and cellxgene data portal. This project was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, Project number 458958943 and TRR127). Further, this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 874839. We also acknowledge support by the Federal Ministry of Education and Research (BMBF) due to an enactment of the German Bundestag under Grant No. 031L0251 e-Islet-Organersatz. S.T. is supported by a DFG Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). M.T. acknowledges financial support by the Volkswagen Foundation (project OntoTime). F.J.T. acknowledges support by the BMBF (grant #L031L0214 A, grant# 01IS18036A and grant# 01IS18053A), by the Helmholtz Association (Incubator grant sparse2big, grant # ZT-I-0007), by Helmholtz Association's Initiative and Networking Fund through Helmholtz AI [grant # ZT-I-PF-5-01] and by the Chan Zuckerberg Initiative DAF (advised fund of Silicon Valley Community Foundation, 2018–182835 and 2019–207271). This research was funded in part, by the Wellcome Trust Grant 108413/A/15/D.

#### CONFLICT OF INTEREST

F.J.T. consults for Immunai Inc., Singularity Bio B.V., CytoReason Ltd, and Omniscope Ltd, and has ownership interest in Dermagnostix GmbH and Cellarity, Inc. S.T. is an employee of Cellarity, Inc. and has stake-holder interests; the present work was carried out as an employee of Helmholtz Munich.

#### APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molmet.2022.101595>.

#### REFERENCES

- [1] Roscioni, S.S., Migliorini, A., Gegg, M., Lickert, H., 2016. Impact of islet architecture on  $\beta$ -cell heterogeneity, plasticity and function. *Nature Reviews Endocrinology* 12:695–709.
- [2] Pipeleers, D.G., 1992. Heterogeneity in pancreatic beta-cell population. *Diabetes* 41:777–781.
- [3] Gutierrez, G.D., Gromada, J., Sussel, L., 2017. Heterogeneity of the pancreatic beta cell. *Frontiers in Genetics* 8:22.

- [4] Camunas-Soler, J., Dai, X.-Q., Hang, Y., Bautista, A., Lyon, J., Suzuki, K., et al., 2020. Patch-seq links single-cell transcriptomes to human islet dysfunction in diabetes. *Cell Metabolism* 31:1017–1031.e4.
- [5] Ghazvini Zadeh, E.H., Huang, Z., Xia, J., Li, D., Davidson, H.W., Li, W.-H., 2020. ZIGIR, a granule-specific Zn indicator, reveals human islet  $\alpha$  cell heterogeneity. *Cell Reports* 32:107904.
- [6] Dai X-Q, Camunas-Soler J, Briant LJB, dos Santos T, Spigelman AF, Walker EM, et al. Heterogenous impairment of  $\alpha$ -cell function in type 2 diabetes is linked to cell maturation state. *Cell Metabolism*. doi:10.1101/2021.04.08.435504
- [7] Benninger, R.K.P., Hodson, D.J., 2018. New understanding of  $\beta$ -cell heterogeneity and in situ islet function. *Diabetes* 67:537–547.
- [8] Benninger, R.K.P., Kravets, V., 2021. The physiological role of  $\beta$ -cell heterogeneity in pancreatic islet function. *Nature Reviews Endocrinology*. <https://doi.org/10.1038/s41574-021-00568-0>.
- [9] Xin, Y., Dominguez Gutierrez, G., Okamoto, H., Kim, J., Lee, A.-H., Adler, C., et al., 2018. Pseudotime ordering of single human  $\beta$ -cells reveals states of insulin production and unfolded protein response. *Diabetes* 67:1783–1794.
- [10] Aguayo-Mazzucato, C., 2020. Functional changes in beta cells during ageing and senescence. *Diabetologia* 63:2022–2029.
- [11] Sachs, S., Bastidas-Ponce, A., Tritschler, S., Bakhti, M., Böttcher, A., Sánchez-Garrido, M.A., et al., 2020. Targeted pharmacological therapy restores  $\beta$ -cell function for diabetes remission. *Nat Metab* 2:192–209.
- [12] Tritschler, S., Theis, F.J., Lickert, H., Böttcher, A., 2017. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas. *Mol Metab* 6:974–990.
- [13] Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., et al., 2016. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst* 3:346–360 e4.
- [14] Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., et al., 2016. A single-cell transcriptome atlas of the human pancreas. *Cell Syst* 3:385–394 e3.
- [15] Bakhti, M., Böttcher, A., Lickert, H., 2019. Modelling the endocrine pancreas in health and disease. *Nature Reviews Endocrinology* 15:155–171.
- [16] Ludwig, B., Ludwig, S., Steffen, A., Knauf, Y., Zimmerman, B., Heinke, S., et al., 2017. Favorable outcome of experimental islet xenotransplantation without immunosuppression in a nonhuman primate model of diabetes. *Proc Natl Acad Sci*, 11745–11750. <https://doi.org/10.1073/pnas.1708420114>.
- [17] Renner, S., Blutke, A., Claus, S., Deeg, C.A., Kemter, E., Merkus, D., et al., 2020. Porcine models for studying complications and organ crosstalk in diabetes mellitus. *Cell Tissue Res* 380:341–378.
- [18] Renner, S., Dobenecker, B., Blutke, A., Zöls, S., Wanke, R., Ritzmann, M., et al., 2016. Comparative aspects of rodent and nonrodent animal models for mechanistic and translational diabetes research. *Theriogenology* 86: 406–421.
- [19] Coe, T.M., Markmann, J.F., Rickert, C.G., 2020. Current status of porcine islet xenotransplantation. *Current Opinion Organ Transpl* 25:449–456.
- [20] Kim, S., Whitener, R.L., Peiris, H., Gu, X., Chang, C.A., Lam, J.Y., et al., 2020. Molecular and genetic regulation of pig pancreatic islet cell development. *Development* 147. <https://doi.org/10.1242/dev.186213>.
- [21] Thompson, P.J., Shah, A., Ntranos, V., Van Gool, F., Atkinson, M., Bhushan, A., 2019. Targeted elimination of senescent beta cells prevents type 1 diabetes. *Cell Metabolism* 29:1045–1060 e10.
- [22] Tatsuoka, H., Sakamoto, S., Yabe, D., Kabai, R., Kato, U., Okumura, T., et al., 2020. Single-cell transcriptome analysis dissects the replicating process of pancreatic beta cells in partial pancreatectomy model. *iScience* 23:101774.
- [23] Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.-M., Andréasson, A.-C., Sun, X., et al., 2016. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metabolism* 24:593–607.
- [24] Xin, Y., Kim, J., Okamoto, H., Ni, M., Wei, Y., Adler, C., et al., 2016. RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metabolism* 24:608–615.
- [25] Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., et al., 2017. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* 171:321–330 e14.
- [26] Wilkinson, M.D., Dumontier, M., Ijz, Aalbersberg, Appleton, G., Axton, M., Baak, A., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3:160018.
- [27] Fischer, D.S., Dony, L., König, M., Moed, A., Zappia, L., Heumos, L., et al., 2021. Sfaira accelerates data and model reuse in single cell genomics. *Genome Biology* 22:248.
- [28] Steiner, D.J., Kim, A., Miller, K., Hara, M., 2010. Pancreatic islet plasticity: interspecies comparison of islet architecture and composition. *Islets* 2:135–145.
- [29] Kim, A., Miller, K., Jo, J., Kiimnik, G., Wojcik, P., Hara, M., 2009. Islet architecture: a comparative study. *Islets* 1:129–136.
- [30] Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D.M., Billis, K., et al., 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* 9. <https://doi.org/10.1093/gigascience/giaa051>.
- [31] Summers, K.M., Bush, S.J., Wu, C., Su, A.I., Muriuki, C., Clark, E.L., et al., 2019. Functional annotation of the transcriptome of the pig, based upon network analysis of an RNAseq transcriptional atlas. *Front Genet* 10:1355.
- [32] Li, M., Chen, L., Tian, S., Lin, Y., Tang, Q., Zhou, X., et al., 2017. Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res* 27:865–874.
- [33] Godard, P., van Eyll, J., 2018. BED: a Biological Entity Dictionary based on a graph data model. *F1000Res* 7:195.
- [34] Bastidas-Ponce, A., Scheibner, K., Lickert, H., Bakhti, M., 2017. Cellular and molecular mechanisms coordinating pancreas development. *Development* 144:2873–2888.
- [35] Napolitano, T., Avolio, F., Courtney, M., Vieira, A., Druelle, N., Ben-Othman, N., et al., 2015. Pax4 acts as a key player in pancreas development and plasticity. *Semin Cell Dev Biol* 44:107–114.
- [36] Benner, C., van der Meulen, T., Cacères, E., Tigyi, K., Donaldson, C.J., Huisling, M.O., 2014. The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. *BMC Genomics* 15:620.
- [37] Salinno, C., Cota, P., Bastidas-Ponce, A., Tarquis-Medina, M., Lickert, H., Bakhti, M., 2019.  $\beta$ -Cell maturation and identity in health and disease. *Int J Mol Sci* 20. <https://doi.org/10.3390/ijms20215417>.
- [38] Bader, E., Migliorini, A., Gegg, M., Moruzzi, N., Gerdes, J., Roscioni, S.S., et al., 2016. Identification of proliferative and mature  $\beta$ -cells in the islets of Langerhans. *Nature* 535:430–434.
- [39] Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L., et al., 2016. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat Methods* 13:241–244.
- [40] Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-9-559>.
- [41] Richardson, S.J., Rodriguez-Calvo, T., Gerling, I.C., Mathews, C.E., Kaddis, J.S., Russell, M.A., et al., 2016. Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes. *Diabetologia* 59: 2448–2458.
- [42] Russell, M.A., Redick, S.D., Blodgett, D.M., Richardson, S.J., Leete, P., Krogvold, L., et al., 2019. HLA class II antigen processing and presentation pathway components demonstrated by transcriptome and protein analyses of islet  $\beta$ -cells from donors with type 1 diabetes. *Diabetes* 68:988–1001.
- [43] Fonseca, S.G., Burcin, M., Gromada, J., Urano, F., 2009. Endoplasmic reticulum stress in beta-cells and development of diabetes. *Current Opinion in Pharmacology* 9:763–770.



- [44] Rabhi, N., Salas, E., Froguel, P., Annicotte, J.-S., 2014. Role of the unfolded protein response in  $\beta$  cell compensation and failure during diabetes. *Journal of Diabetes Research* 2014:795171.
- [45] Aylward A, Okino M-L, Benaglio P, Chiou J, Beebe E, Padilla JA, et al. Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *PLoS Genetics*. doi:10.1101/2020.05.15.038679
- [46] Ramzy, A., Asadi, A., Kieffer, T.J., 2020. Revisiting proinsulin processing: evidence that human  $\beta$ -cells process proinsulin with prohormone convertase (PC) 1/3 but not PC2. *Diabetes* 69:1451–1462.
- [47] Pfützner, A., Kunt, T., Hohberg, C., Mondok, A., Pahler, S., Konrad, T., et al., 2004. Fasting intact proinsulin is a highly specific predictor of insulin resistance in type 2 diabetes. *Diabetes Care* 27:682–687.
- [48] El Shabrawy, A.M., Elbana, K.A., Abdelsalam, N.M., 2019. Proinsulin/insulin ratio as a predictor of insulin resistance and B-cell dysfunction in obese Egyptians ((insulin resistance & B-cell dysfunction in obese Egyptians)). *Diabetes & Metabolic Syndrome* 13:2094–2096.
- [49] Sims, E.K., Bahnson, H.T., Nyalwidhe, J., Haataja, L., Davis, A.K., Speake, C., et al., 2019. Proinsulin secretion is a persistent feature of type 1 diabetes. *Diabetes Care* 42:258–264.
- [50] Then, C., Gar, C., Thorand, B., Huth, C., Then, H., Meisinger, C., et al., 2020. Proinsulin to insulin ratio is associated with incident type 2 diabetes but not with vascular complications in the KORA F4/FF4 study. *BMJ Open Diabetes Res Care* 8. <https://doi.org/10.1136/bmjdr-2020-001425>.
- [51] Singh, A., Gibert, Y., Dwyer, K.M., 2018. The adenosine, adrenergic and opioid pathways in the regulation of insulin secretion, beta cell proliferation and regeneration. *Pancreatology* 18:615–623.
- [52] Schuit, F., Pipeleers, D., 1986. Differences in adrenergic recognition by pancreatic A and B cells. *Science*, 875–877. <https://doi.org/10.1126/science.2871625>.
- [53] You, H., Laychock, S.G., 2009. Atrial natriuretic peptide promotes pancreatic islet beta-cell growth and Akt/Foxo1a/cyclin D2 signaling. *Endocrinology* 150: 5455–5465.
- [54] Undank, S., Kaiser, J., Sikimic, J., Düfer, M., Krippeit-Drews, P., Drews, G., 2017. Atrial natriuretic peptide affects stimulus-secretion coupling of pancreatic  $\beta$ -cells. *Diabetes* 66:2840–2848.
- [55] Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., et al., 2017. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Research*, 208–222. <https://doi.org/10.1101/gr.212720.116>.
- [56] Shrestha, S., Saunders, D.C., Walker, J.T., Camunas-Soler, J., Dai, X.-Q., Haliyur, R., et al., 2021. Combinatorial transcription factor profiles predict mature and functional human islet  $\alpha$  and  $\beta$  cells. *JCI Insight* 6. <https://doi.org/10.1172/jci.insight.151621>.
- [57] Fasolino M, Schwartz GW, Golson ML, Wang YJ, Morgan A, Liu C, et al Multiomics single-cell analysis of human pancreatic islets reveals novel cellular states in health and type 1 diabetes, bioRxiv. 2021. doi:10.1101/2021.01.28.428598
- [58] Fang, Z., Weng, C., Li, H., Tao, R., Mai, W., Liu, X., et al., 2019. Single-cell heterogeneity analysis and CRISPR screen identify key  $\beta$ -cell-specific disease genes. *Cell Reports*. p. 3132–44. <https://doi.org/10.1016/j.celrep.2019.02.043> e7.
- [59] Bergen, V., Lange, M., Peidli, S., Wolf, F.A., Theis, F.J., 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology* 38:1408–1414.
- [60] La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., et al., 2018. RNA velocity of single cells. *Nature* 560:494–498.
- [61] Szabat, M., Pourghaderi, P., Soukhatcheva, G., Verchere, C.B., Warnock, G.L., Piret, J.M., et al., 2011. Kinetics and genomic profiling of adult human and mouse  $\beta$ -cell maturation. *Islets* 3:175–187.
- [62] Piccand, J., Meunier, A., Merle, C., Jia, Z., Barnier, J.-V., Gradwohl, G., 2014. Pak3 promotes cell cycle exit and differentiation of  $\beta$ -cells in the embryonic pancreas and is necessary to maintain glucose homeostasis in adult mice. *Diabetes* 63:203–215.
- [63] Konstantinova, I., Nikolova, G., Ohara-Imaizumi, M., Meda, P., Kucera, T., Zarbalis, K., et al., 2007. EphA-Ephrin-A-mediated beta cell communication regulates insulin secretion from pancreatic islets. *Cell* 129:359–370.
- [64] Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., et al., 2020. A human cell atlas of fetal gene expression. *Science*, 370. <https://doi.org/10.1126/science.aba7721>.
- [65] Yu, X.-X., Qiu, W.-L., Yang, L., Wang, Y.-C., He, M.-Y., Wang, D., et al., 2021. Sequential progenitor states mark the generation of pancreatic endocrine lineages in mice and humans. *Cell Research* 31:886–903.
- [66] Dorrell, C., Schug, J., Canaday, P.S., Russ, H.A., Tarlow, B.D., Grompe, M.T., et al., 2016. Human islets contain four distinct subtypes of  $\beta$  cells. *Nature Communications* 7:11756.
- [67] Piñeros, A.R., Gao, H., Wu, W., Liu, Y., Tersey, S.A., Mirmira, R.G., 2020. Single-cell transcriptional profiling of mouse islets following short-term obesogenic dietary intervention. *Metabolites* 10. <https://doi.org/10.3390/metabo10120513>.
- [68] Bilekova, S., Sachs, S., Lickert, H., 2021. Pharmacological targeting of endoplasmic reticulum stress in pancreatic beta cells. *Trends in Pharmacological Sciences* 42:85–95.
- [69] Tarifeño-Saldivia, E., Lavergne, A., Bernard, A., Padamata, K., Bergemann, D., Voz, M.L., et al., 2017. Transcriptome analysis of pancreatic cells across distant species highlights novel important regulator genes. *BMC Biology* 15:21.
- [70] Steffen, A., Kiss, T., Schmid, J., Schubert, U., Heinke, S., Lehmann, S., et al., 2017. Production of high-quality islets from goettingen minipigs: choice of organ preservation solution, donor pool, and optimal cold ischemia time. *Xenotransplantation* 24. <https://doi.org/10.1111/xen.12284>.
- [71] Wolf, F.A., Angerer, P., Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* 19:15.
- [72] Luecken, M.D., Theis, F.J., 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology* 15:e8746.
- [73] Lun, A.T.L., Bach, K., Marioni, J.C., 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology* 17:75.
- [74] Lun, A.T.L., McCarthy, D.J., Marioni, J.C., 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 5: 2122.
- [75] Büttner, M., Miao, Z., Wolf, F.A., Teichmann, S.A., Theis, F.J., 2019. A test metric for assessing single-cell RNA-seq batch correction. *Nature Methods* 16: 43–49.
- [76] Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8:118–127.
- [77] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* P10008. <https://doi.org/10.1088/1742-5468/2008/10/p10008>.
- [78] Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., et al., 2018. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.4314>.
- [79] Wolock, S.L., Lopez, R., Klein, A.M., 2019. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst* 8:281–291 e9.
- [80] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17:261–272.
- [81] Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., et al., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* 44:W90–W97.

- [82] Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q., Guo, A.-Y., 2019. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Research* 47:D33–D38.
- [83] Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* 33: 495–502.
- [84] Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., Park, J.-E., 2020. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36:964–965.
- [85] Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).

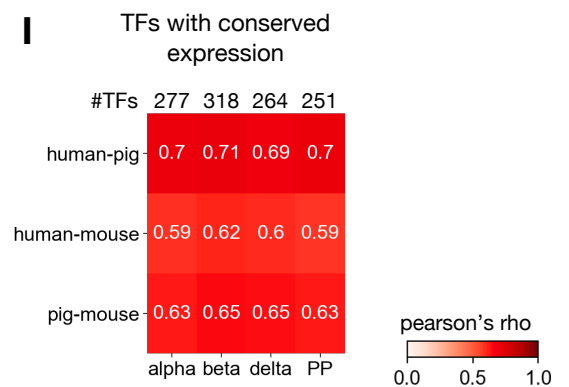
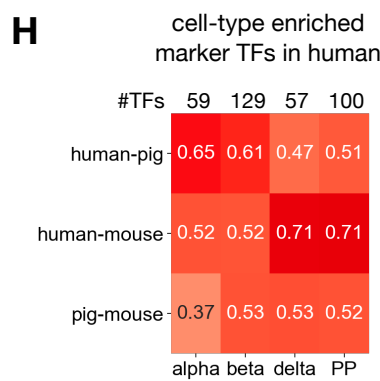
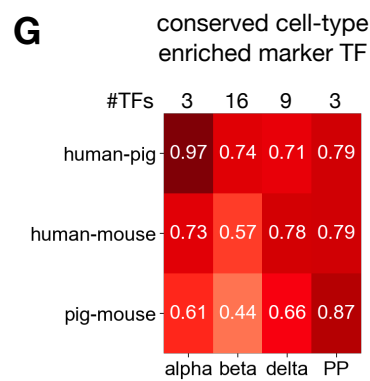
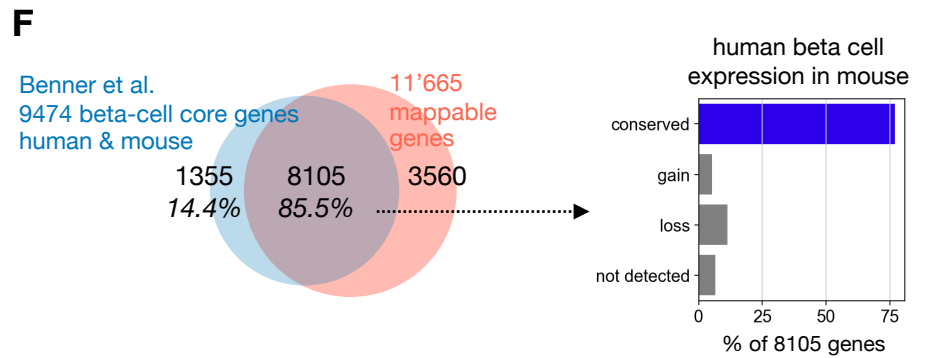
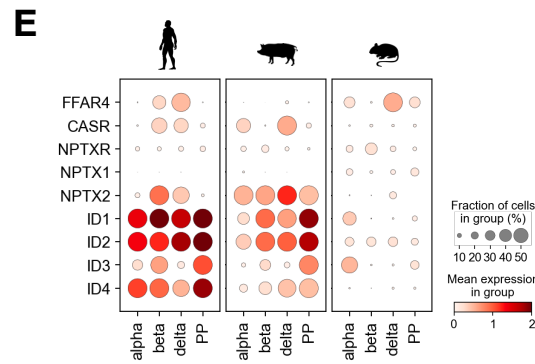
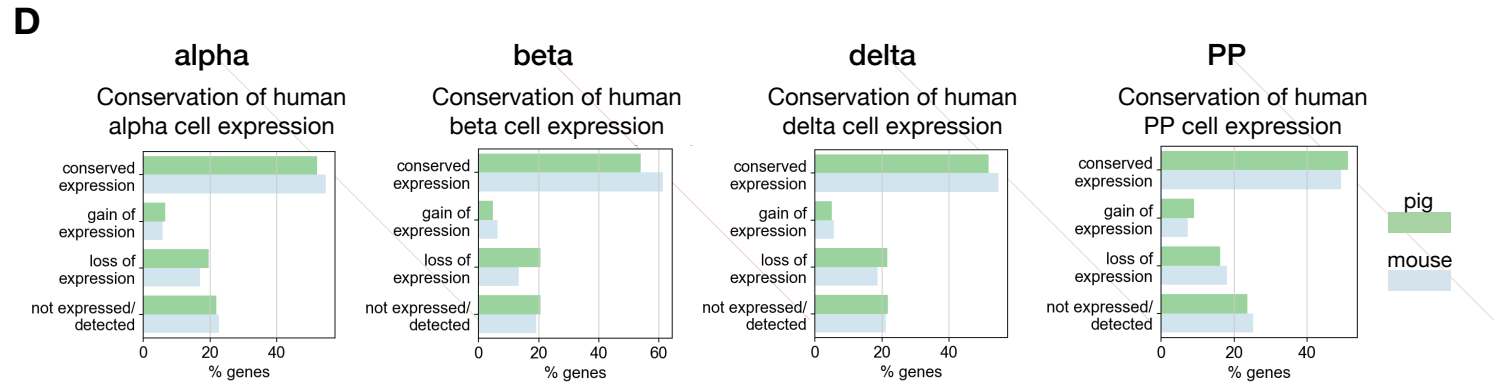
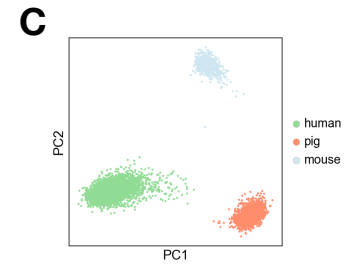
# Supplementary Figure 1 Conservation of gene expression in scRNA-seq data of human, mouse and pig islet cells

**A**

ID	sex	age	BMI	HbA1c %
R229	female	22	23,0	5,3
R239	female	24	22,0	5,5
R237	male	61	19,6	5,9
R245	male	63	22,3	5,6
R266	female	74	29,2	6,0

**B**

sample	# cells	median #counts/cell	median #genes/cell
human	26474	15277	3159
pig	22056	8831	1836
mouse	7011	16927	2423



## Supplementary Figure 1 Conservation of gene expression in scRNA-seq data of human, mouse and pig islet cells

A) Metadata of the 5 human donors. ID indicates donor ID for ADI IsletCore (see Material & Methods).

B) Quality control metrics of scRNA-seq data.

C) Scatter plot of the top two principal components. Cells are colored by species.

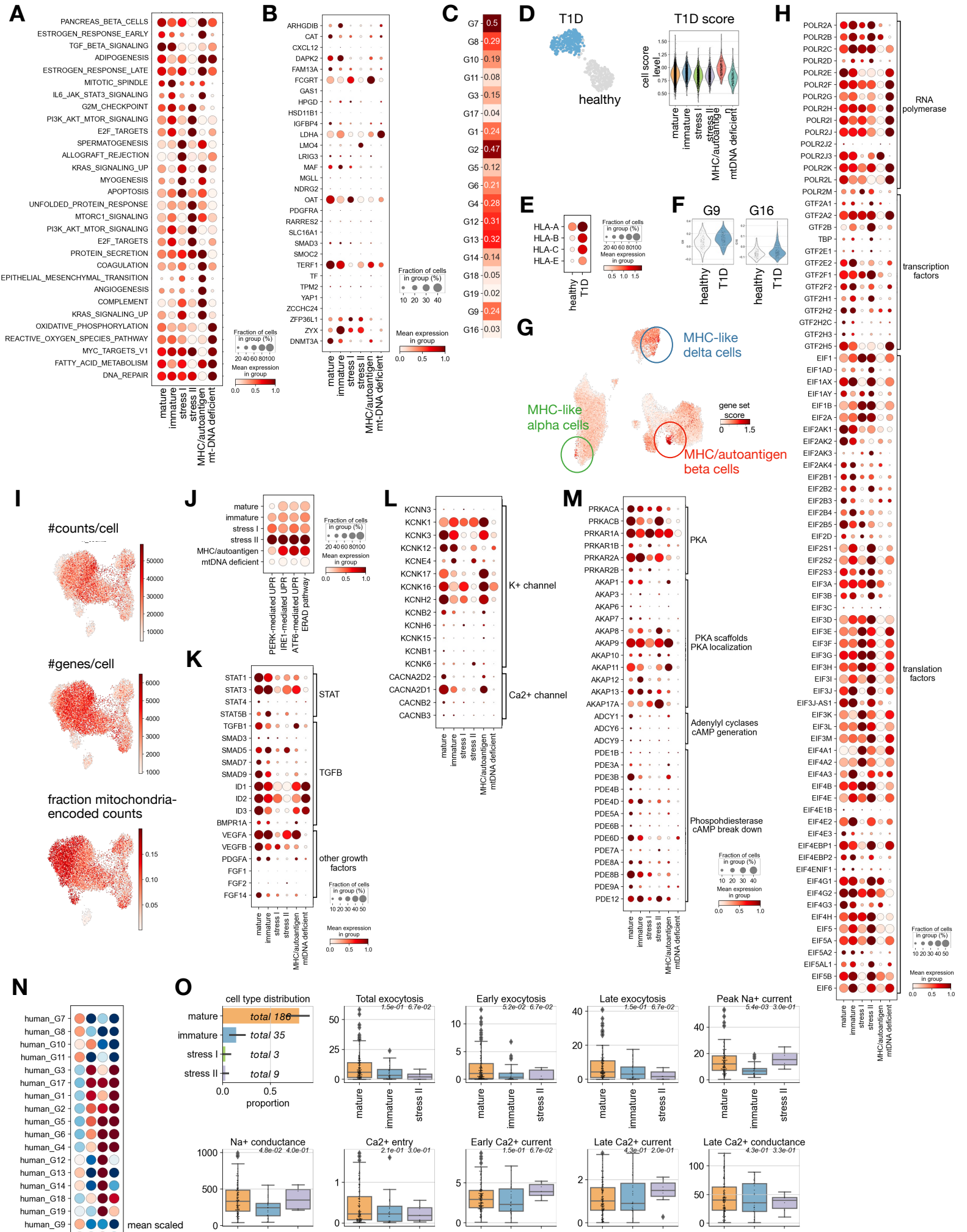
D) Summary of conservation of human gene expression in pig and mouse in endocrine cell types. A gene is considered expressed if detected in >5% of the cells of the cell type.

E) Expression of selected genes in human, pig and mouse endocrine cell types exemplifying conservation, “gain” and “loss” of expression shown in C. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene.

F) Comparison of conserved  $\beta$ -cell genes to  $\beta$ -cell core genes derived from human and mouse bulk  $\beta$ -cell transcriptomes [36]. Left: Venn Diagram showing the overlap of reported  $\beta$ -cell core genes (9'474) and our list of mappable genes (11'665). Right: Barplot indicating conservation of 8105 overlapping  $\beta$ -cell core genes between human and mouse  $\beta$ -cells. A gene is considered expressed if detected in >5% of the cells of the cell type.

G-I) Pairwise correlation of TF expression patterns between species for each cell type. Pearson correlation is computed on a subset of TF as indicated using the harmonic average of mean expression and fraction of cells expressing a gene in a group across all cell types (Material & Methods). Pearson correlation coefficient is indicated. G) Cell-type enriched marker TFs conserved across species as shown in Figure 1G. H) All TFs enriched in human cell-types. I) All TFs with conserved expression across species.

# Supplementary Figure 2 Transcriptional profiling of human $\beta$ -cell states



## Supplementary Figure 2 Transcriptional profiling of human $\beta$ -cell states

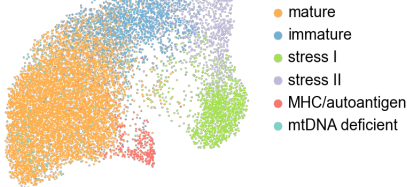
- A) Cell scores indicating hallmark pathway activation in  $\beta$ -cell clusters. Top 5 enriched hallmarks are shown per cluster. Scaled scores per pathway are shown.
- B) Expression of  $\beta$ -cell disallowed genes in  $\beta$ -cell clusters. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.
- C) Variance of non-scaled gene set scores across all  $\beta$ -cells indicating magnitude of activation level differences across clusters.
- D-F) Comparison of the transcriptional profile of the identified MHC/autoantigen  $\beta$ -cell cluster to  $\beta$ -cells from T1D patients [42]. D) UMAP plot of  $\beta$ -cells from healthy and T1D patients. T1D score indicates increased expression of T1D-associated genes in the MHC/autoantigen cluster. The T1D score is computed from the top differentially expressed genes between  $\beta$ -cells of T1D patients and healthy individuals. E) Expression of MHC genes in healthy and T1D  $\beta$ -cells. F) Gene sets increased in MHC/autoantigen cluster are also increased in T1D  $\beta$ -cells.
- G) UMAP plot of endocrine cells colored by MHC/autoantigen gene set (G9) scores. Circles highlight clusters with high activation scores in  $\alpha$ -,  $\beta$ - and  $\delta$ -cells.
- H) Expression of RNA polymerase II and general transcription and translation factors expressed in >200  $\beta$ -cells. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.
- I) UMAP plot of human  $\beta$ -cells colored by data quality metrics. Top: Percentage of counts from mitochondria-encoded RNA, middle: total number of counts per cell, bottom: total number of genes per cell.
- J) Cells scores indicating stress pathway activation in  $\beta$ -cell clusters. Scores were computed based on the expression of genes in the corresponding GO pathways (see Methods).
- K) Expression of transcription, signaling and growth factors in  $\beta$ -cell clusters. Genes were described to be significantly downregulated by glucocorticoid signaling in human islets [45]. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.
- L,M) Expression of ion channels (L) and selected components of cAMP signaling pathway (M) expressed in >200  $\beta$ -cells. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.
- N,O) Excitability of  $\beta$ -cell states measured in single-cell Patch-Seq data [4]. State labels were mapped in the  $\beta$ -cell gene set representation using the Scanpy ingest functionality. N)  $\beta$ -cell gene set activation in Patch-Seq cells. Scaled mean scores for each gene set per  $\beta$ -cell state are shown. O) Boxplots showing the distribution of different electrophysiological measurements per  $\beta$ -cell state (top left). Line indicates the median, values are FDR of differential test against mature state. Extreme values above 97% or below 3% - quantiles were excluded. Data were analyzed by a Mann-Whitney-U test and Benjamini-Hochberg correction for multiple testing per state comparison. Top left: Barplot showing  $\beta$ -cell state composition and total number of cells per state. Error bar indicating donor variation.

# Supplementary Figure 3 Cross-study mapping of $\beta$ -cell states

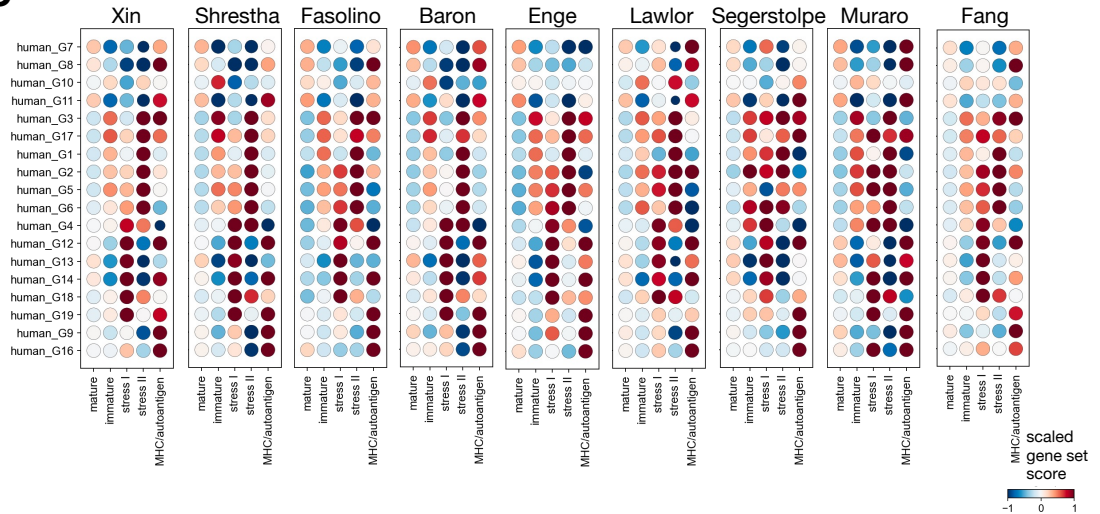
**A**

Tritschler et al, 2022  
n=5, median cells=2196, median genes=3390

11'923 cells



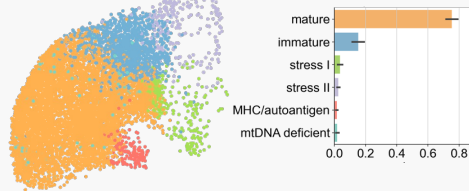
**C**



**B**

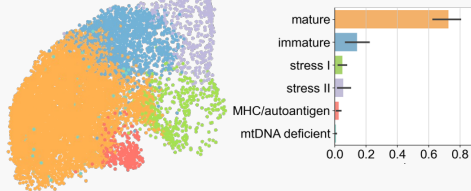
Xin et al, 2018, GSE114297  
n=12, median cells=718, median genes=2409

8'317 cells



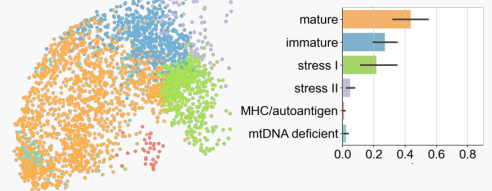
Shrestha et al, 2021, GSE183568  
n=5, median cells=2209, median genes=2505

11'033 cells



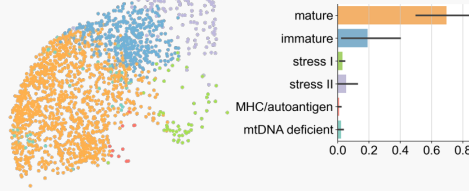
Fasolino et al, 2022  
n=11, median cells=141, median genes=2950

3158 cells



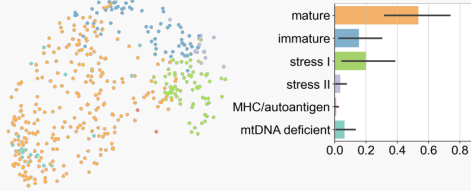
Baron et al, 2016, GSE84133  
n=3, median cells=787, median genes=1946

2030 cells



Enge et al, 2017, GSE81547  
n=8, median cells=45, median genes=3111

472 cells



Lawlor et al, 2017, GSE86469  
n=5, median cells=25, median genes=7984

162 cells



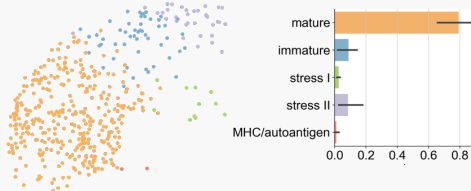
Segerstolpe et al, 2016, E-METAB-5061  
n=6, median cells=33, median genes=4725

171 cells



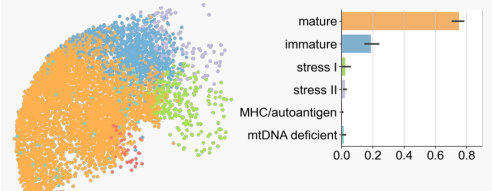
Muraro et al, 2016, GSE85241  
n=4, median cells=146, median genes=5297

480 cells



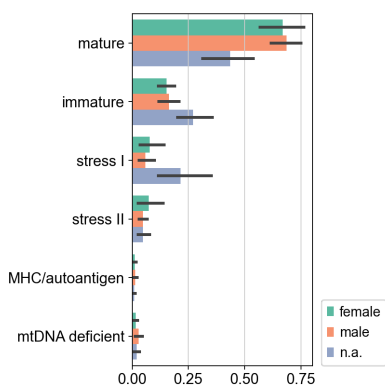
Fang et al, 2019, GSE101207  
n=6, median cells=1151, median genes=691

6731 cells



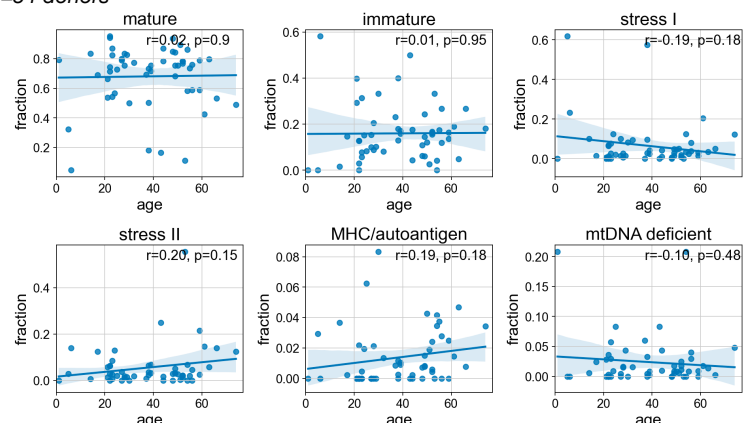
**D**

36 male, 18 female, 11 n.a.



**E**

n=54 donors



### Supplementary Figure 3 Cross-study mapping of $\beta$ -cell states

A-C)  $\beta$ -cell states across 9 studies and 54 donors. A) Reference UMAP showing  $\beta$ -cells in gene set representation, where each cell is represented by an activation score of the corresponding cell gene sets. B) Mapping of  $\beta$ -cells from publicly available studies to the reference UMAP in A. Cells were mapped through projecting on the reference gene set representation. Embedding and labels are mapped using the Scanpy ingest functionality (see Methods). The barplot indicates the frequencies of mapped clusters. Number of donors and median numbers of cells and genes per donor are indicated.

C)  $\beta$ -cell gene set activation in mapped  $\beta$ -cell states per study.

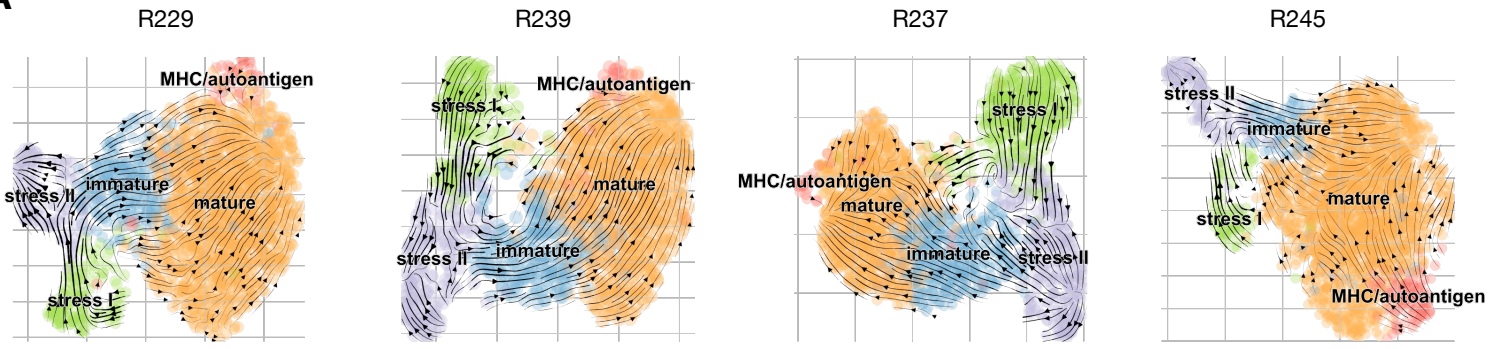
D) Barplot showing fraction of  $\beta$ -cell states in male and female donors of all studies. N.a. indicates donor for which sex information was not available.

E) Scatterplots showing linear relationship between fraction of cells per  $\beta$ -cell cluster and age. Line shows linear regression fit, shaded area shows the 95% confidence interval for the regression. Pearson correlation coefficient ( $r$ ) and  $p$ -value ( $p$ ) testing for non-correlation are indicated.

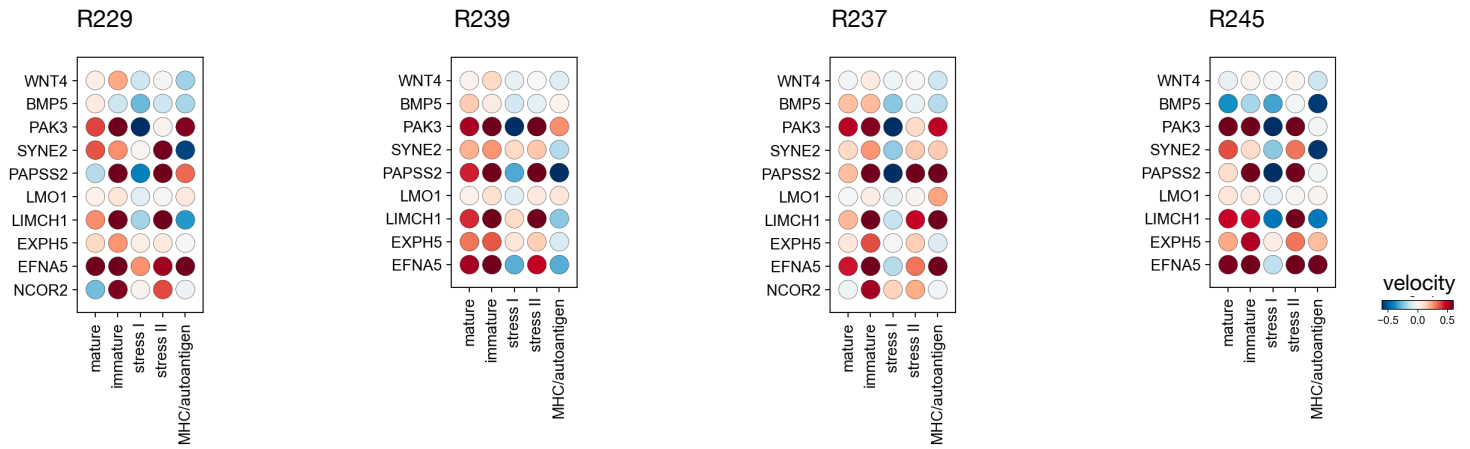


# Supplementary Figure 4 RNA velocity analysis in $\beta$ -cell across human donors

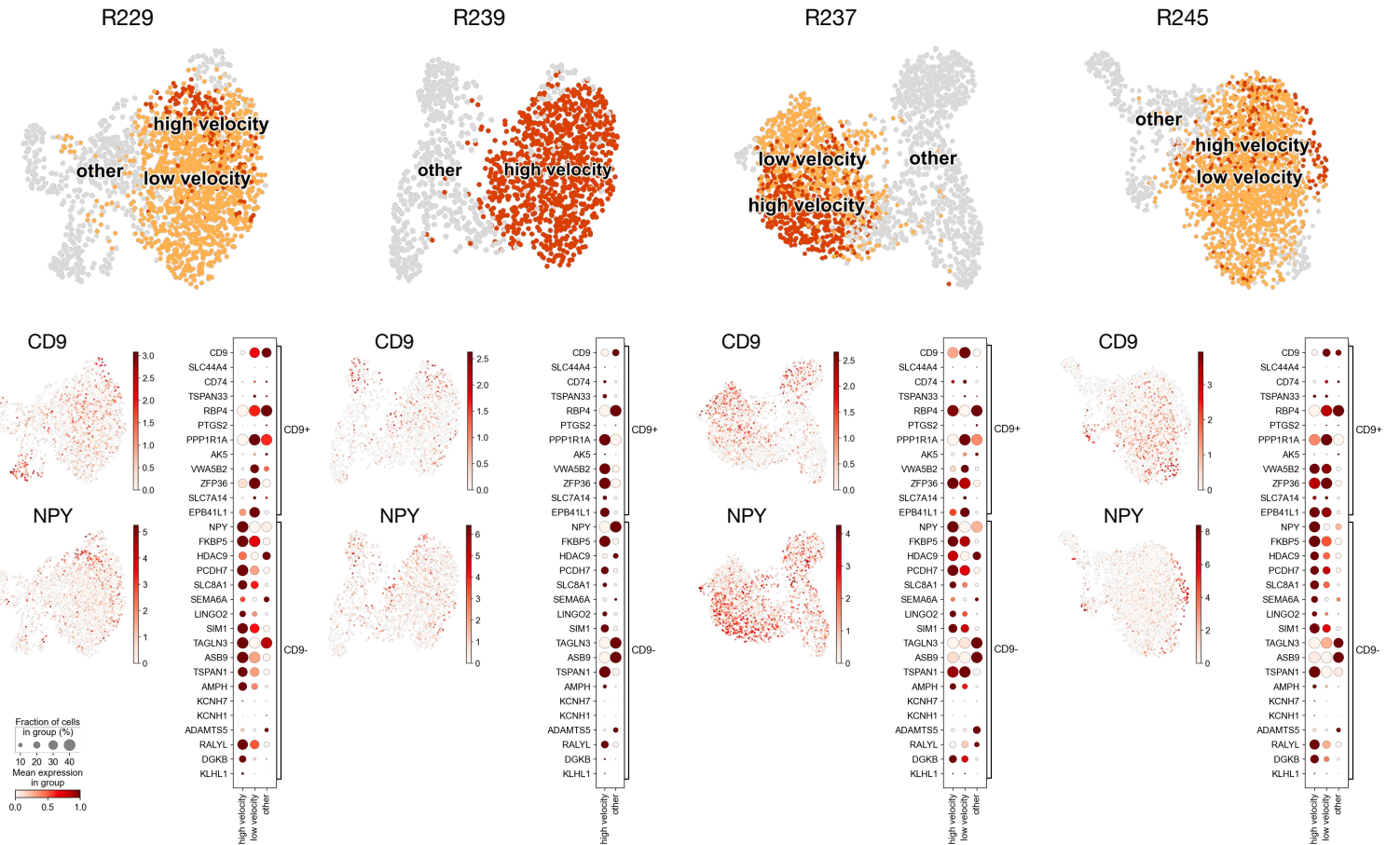
**A**



**B**



**C**



## Supplementary Figure 4 RNA velocity analysis in $\beta$ -cell across human donors

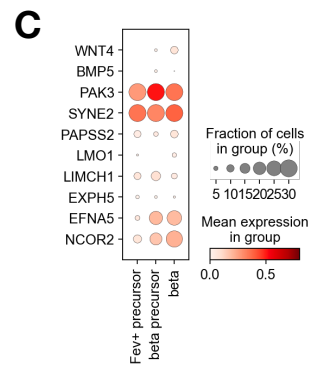
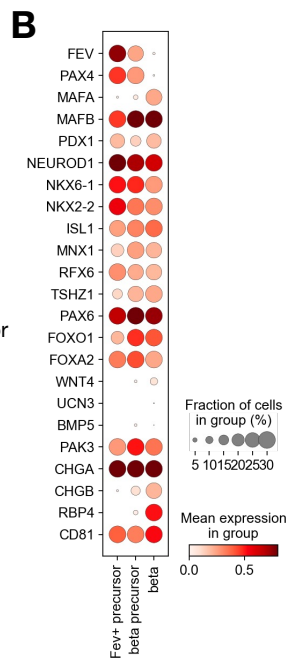
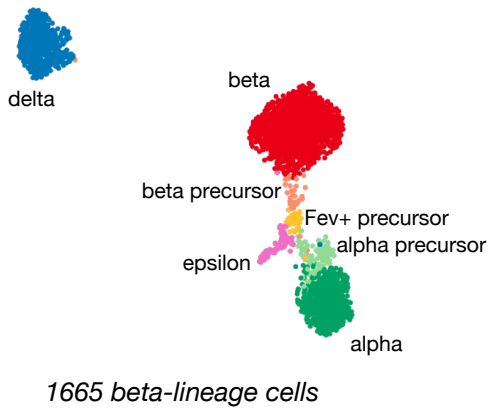
A) Cellular dynamics in  $\beta$ -cells resolved by donor. Cell transitions are inferred from estimated RNA velocities and the direction of inferred movement plotted as streamlines on the UMAP. Colors indicate  $\beta$ -cell clusters.

B) Dotplots showing mean velocities per  $\beta$ -cell cluster resolved by donor. Selected known genes involved in  $\beta$ -cell maturation and potential novel genes important for maturation are shown.

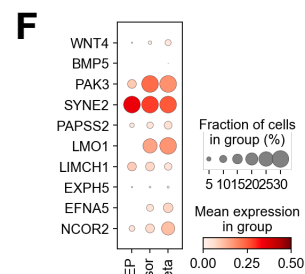
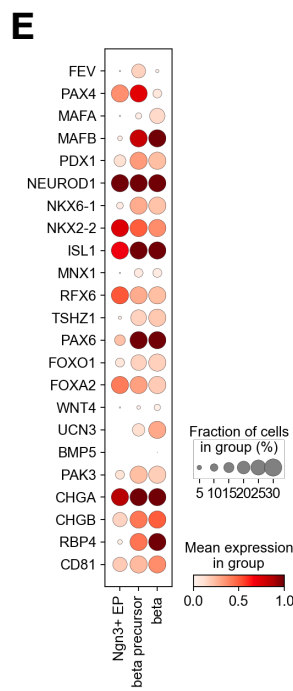
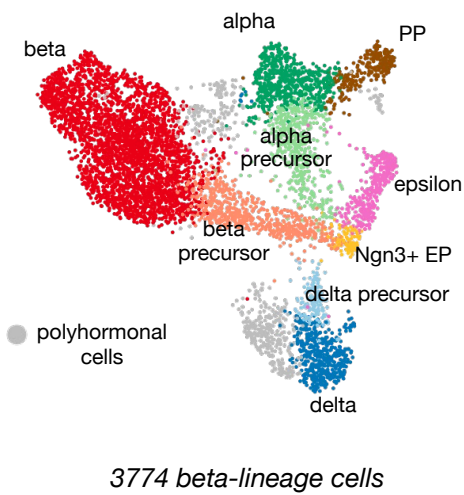
C) Inferred high or low velocity clusters of mature  $\beta$ -cells. *Top*: UMAP indicating clustering into high or low velocity cells. *Bottom*: Expression of genes previously described to separate CD9+ and CD9-  $\beta$ -cells in high and low velocity mature  $\beta$ -cells. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

# Supplementary Figure 5 Maturation factor expression in human fetal $\beta$ -cell development in publicly available datasets

**A** Cao et al., 2020



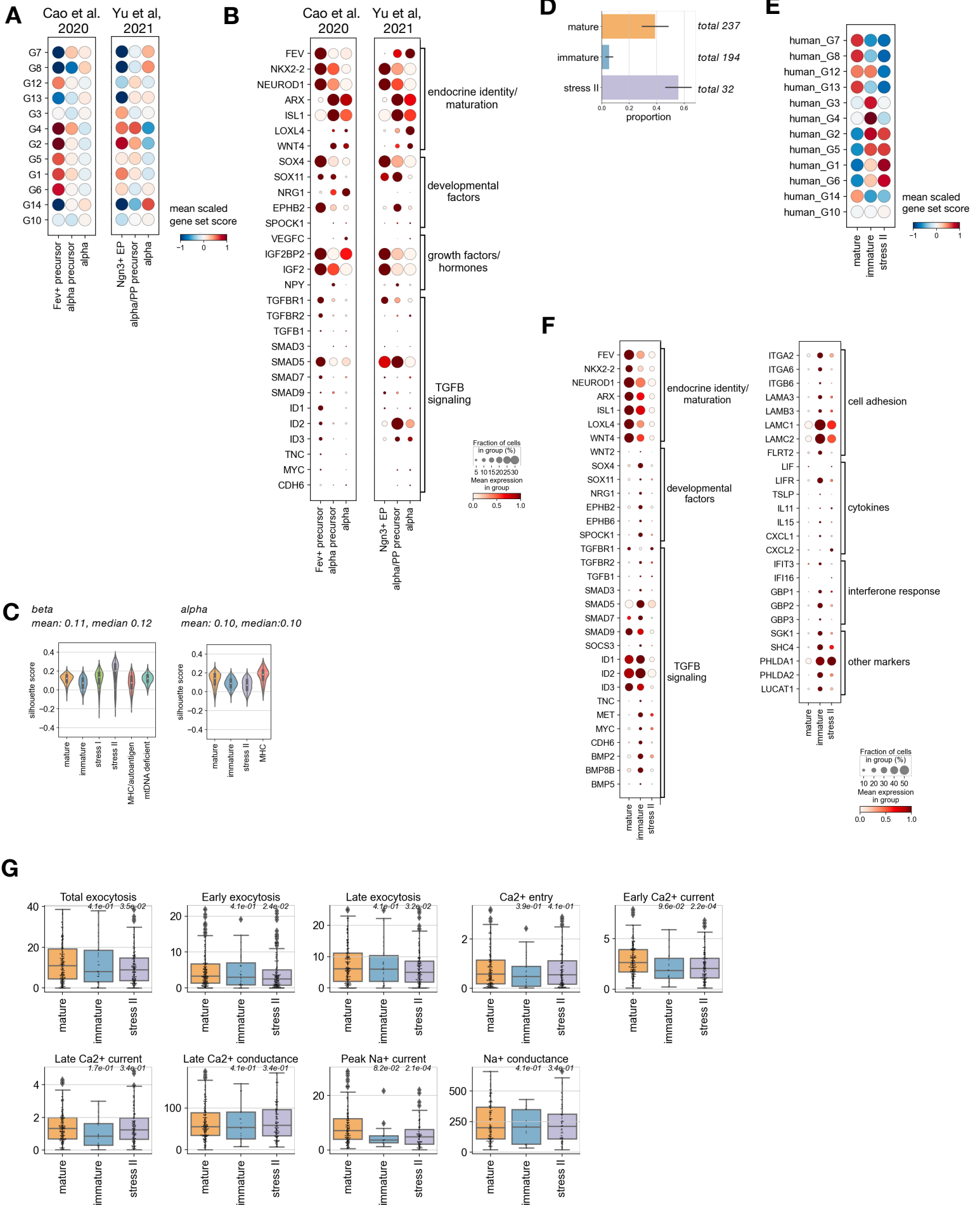
**D** Yu et al., 2021  
3774 cells



## Supplementary Figure 5 Maturation factor expression in human fetal $\beta$ -cell development of publicly available datasets

A-F) Comparison of identified immature  $\beta$ -cell cluster in adult islets to fetal  $\beta$ -cell development. A-C) Single cell sequencing data of fetal pancreata from [64]. D-F) Single cell sequencing data of fetal pancreata from [65]. A, D) UMAP plot of endocrine lineage cells isolated from fetal human pancreases. Colors indicate clusters of differentiation states from Ngn3<sup>+</sup> endocrine progenitors (EP) or Fev<sup>+</sup> precursors, respectively, to immature endocrine cells. B, E) Expression of known  $\beta$ -cell identity and maturity genes. C, F) Expression of genes driving inferred  $\beta$ -cell maturation dynamics (see Figure 3C). Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene.

# Supplementary Figure 6 Transcriptional profiling of human $\alpha$ -cell states



## Supplementary Figure 6 Transcriptional profiling of human $\alpha$ -cell states

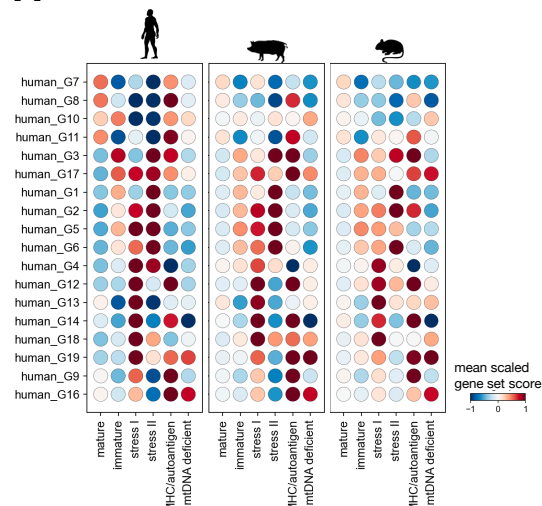
A, B) Comparison of adult  $\alpha$ -cell states to fetal  $\alpha$ -cell development from [64] (Cao et al 2022) and [65] (Yu et al 2021), see also Figure S5A. A) Activation of adult  $\alpha$ -cell gene sets (see Figure 4F) in fetal precursor and  $\alpha$ -cell clusters. Scaled mean scores for each gene set per  $\alpha$ -cell cluster are shown. B) Expression of  $\alpha$ -cell identity and maturation factors as well as developmental factors and genes of the TGF $\beta$  signaling pathway.

C) Silhouette scores [85] as a proxy of cluster similarity and homogeneity. Violinplots show distribution of silhouette scores per  $\beta$ -cell (left) and  $\alpha$ -cell (right) cluster. Silhouette scores were computed on the 50 top principal components using euclidean distance.

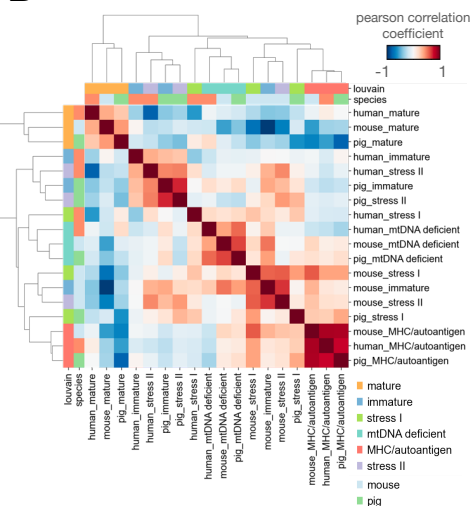
D-G) Excitability of  $\alpha$ -cell states measured in single-cell Patch-Seq data [4]. State labels were mapped in the  $\alpha$ -cell gene set representation using the Scanpy ingest functionality D) Barplot showing  $\alpha$ -cell state composition and total number of cells per state. Error bar indicating donor variation. E)  $\alpha$ -cell gene set activation in Patch-Seq cells. Scaled mean scores for each gene set per  $\alpha$ -cell cluster are shown. F) Expression of  $\alpha$ -cell identity and maturation factors as well as genes involved in pathways describing immature  $\alpha$ -cells. G) Boxplot showing distribution of different electrophysiological measurements per  $\alpha$ -cell state. Line indicates the median, values are FDR of differential test against mature state. Data were analyzed by a Mann-Whitney-U test and Benjamini-Hochberg correction for multiple testing per state comparison.

# Supplementary Figure 7 Conservation of human $\alpha$ - and $\beta$ -cell state signatures in pig and mouse

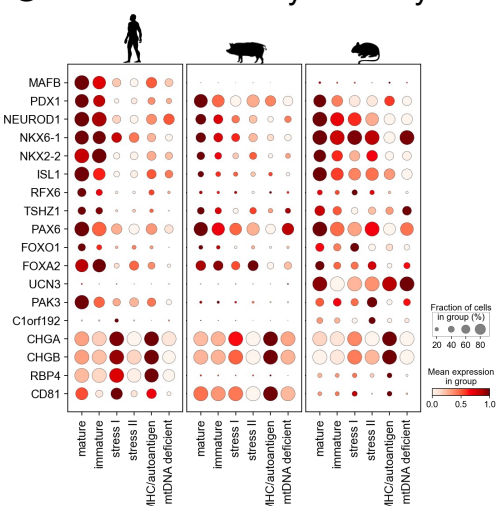
## A beta cell scores



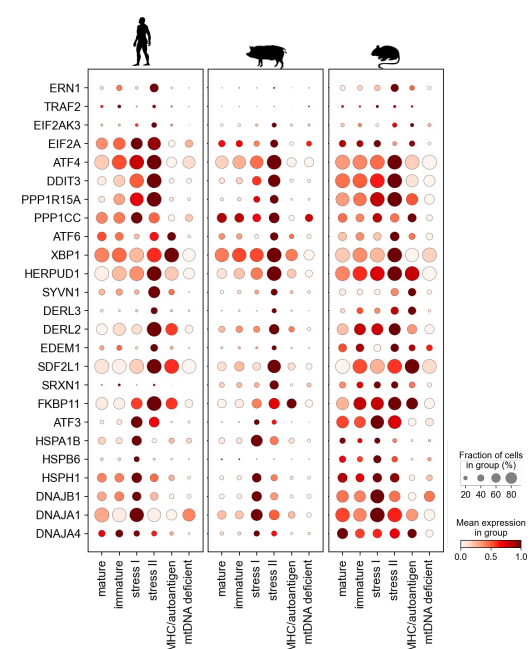
## B



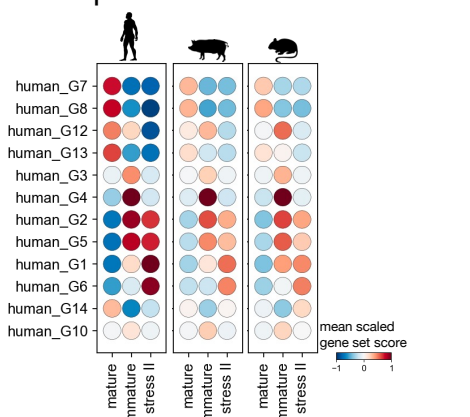
## C beta cell identity/maturity



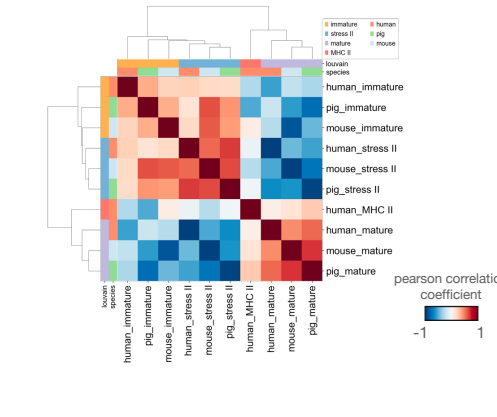
## D beta cell stress



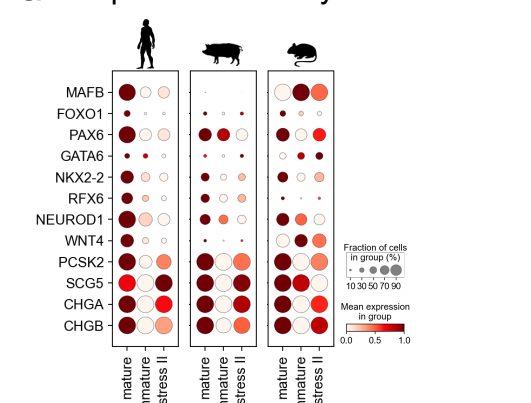
## E alpha cell scores



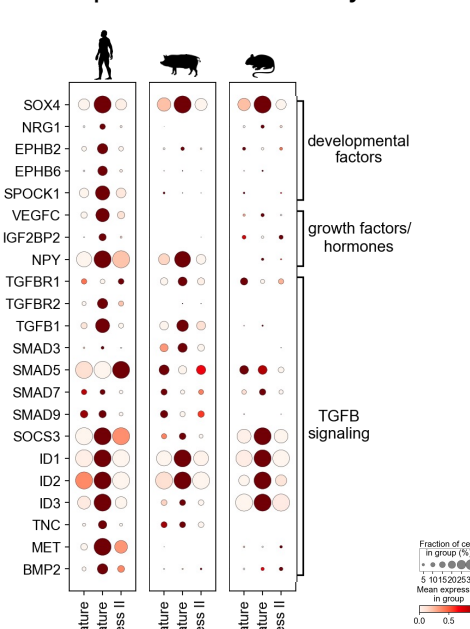
## F



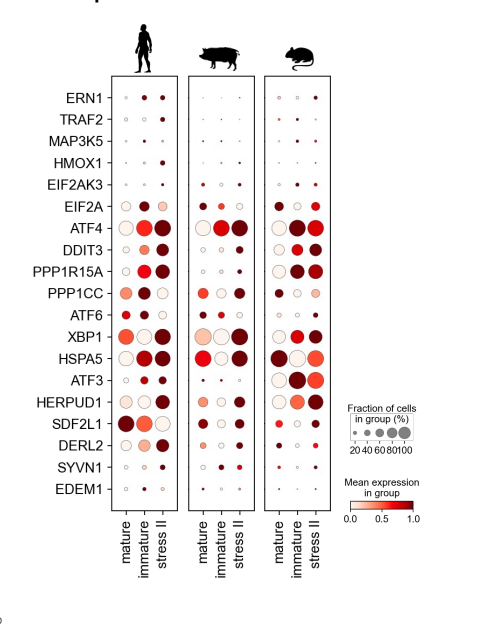
## G alpha cell identity



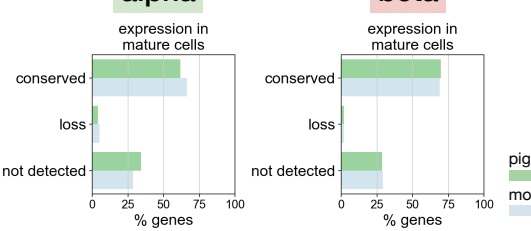
## H alpha cell immaturity



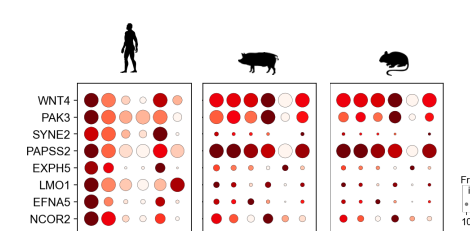
## I alpha cell stress



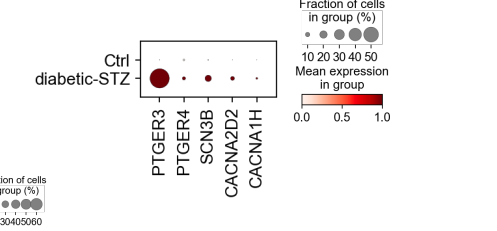
## J alpha



## K beta cell maturation



## L



## Supplementary Figure 7 Conservation of human $\alpha$ - and $\beta$ -cell state signatures in pig and mouse

A-D) Conservation of the human  $\beta$ -cell states A)  $\beta$ -cell gene set activation scores for  $\beta$ -cell clusters across species. B) Pearson correlation matrix of gene expression of  $\beta$ -cell clusters across species.  $\beta$ -cell clusters are grouped by hierarchical clustering. C,D) Expression of  $\beta$ -cell identity and maturity markers (C) and genes associated with a stress-response (D) in  $\beta$ -cell clusters across species. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

E-I) Conservation of the human  $\alpha$ -cell states. E)  $\alpha$ -cell gene set activation scores for  $\alpha$ -cell clusters across species. F) Pearson correlation matrix of gene expression of  $\alpha$ -cell clusters across species.  $\alpha$ -cell clusters are grouped by hierarchical clustering. G-H) Expression of  $\alpha$ -cell identity markers (G), genes describing immature human  $\alpha$ -cells (H) and stress-associated genes (I) in  $\alpha$ -cell clusters across species. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

J) Barplot indicating conservation of gene expression in mature  $\alpha$ - (left) and  $\beta$ - (right) cells from pig and mouse. Conservation of mappable genes within  $\alpha$ - or  $\beta$ -cell maturity gene sets is shown. Genes are considered expressed if detected in >5% of mature cells.

K) Expression of identified  $\beta$ -cell maturation markers in  $\beta$ -cell clusters across species. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.

L) Expression of selected genes in  $\beta$ -cells of scRNA-seq data from vehicle and STZ-treated diabetic mice [11]. Color intensity indicates mean expression in a cluster, dot size indicates the proportion of cells in a cluster expressing the gene. Expression is scaled per gene.



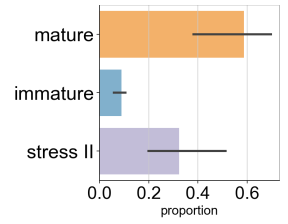
# Supplementary Figure 8 Cross-species mapping of human $\alpha$ - and $\beta$ -cell states using a publicly available mouse dataset

## A alpha

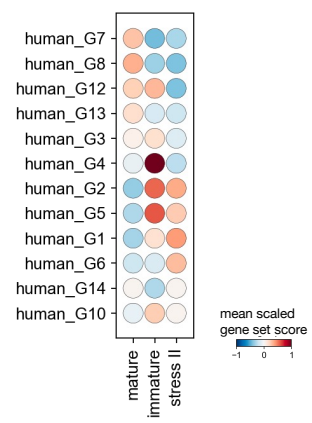
Pineros et al,  
 n=3, total cells=3851  
 median cells=1291, median genes=2467



● mature ● stress II  
 ● immature ● MHC/autoantigen



## B

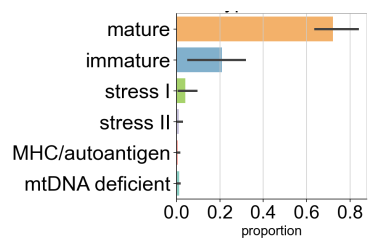


## C beta

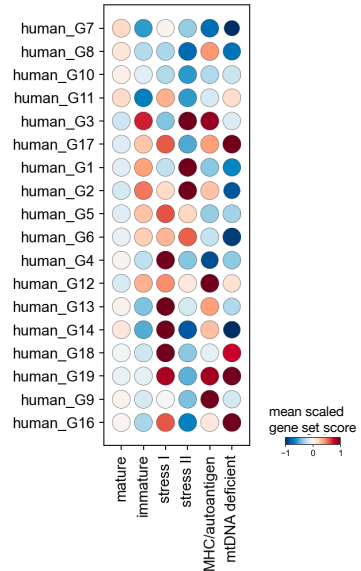
Pineros et al,  
 n=3, total cells=12980,  
 median cells=4203, median genes=3675



● mature ● stress II  
 ● immature ● MHC/autoantigen  
 ● stress I ● mtDNA deficient



## D



## **Supplementary Figure 8** Cross-species mapping of human $\alpha$ - and $\beta$ -cell states using a publicly available mouse dataset

A-D) Conservation of the human  $\alpha$ - and  $\beta$ -cell states in mouse cells of a publicly available mouse dataset [67] A,C) Mapping of mouse  $\alpha$ - (A) and  $\beta$ -cells (C) to the human reference UMAP. Cells were mapped through projecting on the reference gene set representation. Embedding and labels are mapped using the Scanpy ingest functionality (see Methods). The barplot indicates the frequencies of mapped clusters. Number of mice, total and median numbers of cells and genes per mouse are indicated. B,D)  $\alpha$ - (B) and  $\beta$ -cell (D) gene set activation in mapped  $\alpha$ - and  $\beta$ -cell states in [67].

## Appendix B

This is the published version of the article in Nature Metabolism following peer review. The article is published under a Springer License and here reproduced with permission from Springer Nature (License Number 5774150639900).

Stephan Sachs\*, Aimée Bastidas-Ponce\*, **Sophie Tritschler\***, Mostafa Bakhti, Anika Böttcher, Miguel A Sánchez-Garrido, Marta Tarquis-Medina, Maximilian Kleinert, Katrin Fischer, Sigrid Jall, Alexandra Harger, Erik Bader, Sara Roscioni, Siegfried Ussar, Annette Feuchtinger, Burcak Yesildag, Aparna Neelakandhan, Christine B Jensen, Marion Cornu, Bin Yang, Brian Finan, Richard D DiMarchi, Matthias H Tschöp, Fabian J Theis, Susanna M Hofmann, Timo D Müller, Heiko Lickert, **Targeted pharmacological therapy restores  $\beta$ -cell function for diabetes remission.** *Nature Metabolism*, 2, 192–209 (2020)



# Targeted pharmacological therapy restores $\beta$ -cell function for diabetes remission

Stephan Sachs<sup>1,2,3,4,19</sup>, Aimée Bastidas-Ponce<sup>1,4,5,6,19</sup>, Sophie Tritschler<sup>1,4,7,8,19</sup>, Mostafa Bakhti<sup>1,4,5</sup>, Anika Böttcher<sup>1,4,5</sup>, Miguel A. Sánchez-Garrido<sup>2</sup>, Marta Tarquis-Medina<sup>1,4,5,6</sup>, Maximilian Kleinert<sup>2,9</sup>, Katrin Fischer<sup>2,3</sup>, Sigrid Jall<sup>2,3</sup>, Alexandra Harger<sup>2</sup>, Erik Bader<sup>1</sup>, Sara Roscioni<sup>1</sup>, Siegfried Ussar<sup>1,4,6,10</sup>, Annette Feuchtinger<sup>11</sup>, Burcak Yesildag<sup>12</sup>, Aparna Neelakandhan<sup>12</sup>, Christine B. Jensen<sup>13</sup>, Marion Cornu<sup>13</sup>, Bin Yang<sup>14</sup>, Brian Finan<sup>1,4</sup>, Richard D. DiMarchi<sup>14,15</sup>, Matthias H. Tschöp<sup>1,2,3,4</sup>, Fabian J. Theis<sup>1,4,5,6</sup>✉, Susanna M. Hofmann<sup>1,4,17</sup>✉, Timo D. Müller<sup>1,4,18</sup>✉ and Heiko Lickert<sup>1,4,5,6</sup>✉

**Dedifferentiation of insulin-secreting  $\beta$  cells in the islets of Langerhans has been proposed to be a major mechanism of  $\beta$ -cell dysfunction. Whether dedifferentiated  $\beta$  cells can be targeted by pharmacological intervention for diabetes remission, and ways in which this could be accomplished, are unknown as yet. Here we report the use of streptozotocin-induced diabetes to study  $\beta$ -cell dedifferentiation in mice. Single-cell RNA sequencing (scRNA-seq) of islets identified markers and pathways associated with  $\beta$ -cell dedifferentiation and dysfunction. Single and combinatorial pharmacology further show that insulin treatment triggers insulin receptor pathway activation in  $\beta$  cells and restores maturation and function for diabetes remission. Additional  $\beta$ -cell selective delivery of oestrogen by Glucagon-like peptide-1 (GLP-1-oestrogen conjugate) decreases daily insulin requirements by 60%, triggers oestrogen-specific activation of the endoplasmic-reticulum-associated protein degradation system, and further increases  $\beta$ -cell survival and regeneration. GLP-1-oestrogen also protects human  $\beta$  cells against cytokine-induced dysfunction. This study not only describes mechanisms of  $\beta$ -cell dedifferentiation and regeneration, but also reveals pharmacological entry points to target dedifferentiated  $\beta$  cells for diabetes remission.**

The progressive loss or dysfunction of insulin-producing  $\beta$ -cell mass ultimately leads to type 1 diabetes (T1D) or type 2 diabetes (T2D), respectively<sup>1</sup>. Current pharmacological treatments do not stop the decline of  $\beta$ -cell function and number that leads to glucose excursions and, eventually, to devastating microvascular and macrovascular complications. Hence, the ideal treatment should be initiated when the first diabetic symptoms appear, to protect or regenerate glucose-sensing and insulin-secreting  $\beta$  cells for optimal blood glucose regulation and to prevent secondary complications. Recently, T1D progression has been halted by anti-CD3 immunotherapy for 2 years (ref. <sup>2</sup>), but it will be important to test whether additional  $\beta$ -cell regenerative therapy can further or permanently delay the onset of diabetes. Intensive insulin therapy at disease onset has been shown to partially restore  $\beta$ -cell function, which slows the disease progression in patients with T1D and T2D<sup>3–5</sup>. Despite both therapies providing similar glycaemic control, early intensive insulin treatment, compared with oral anti-diabetic drugs,

in patients with T2D better preserves  $\beta$ -cell function, which suggests additional glucose-independent beneficial effects of insulin therapy<sup>6,7</sup>. Thus, an understanding of the mechanisms of  $\beta$ -cell dysfunction and pharmacological replenishment is urgently required to stop or reverse diabetes progression and to improve the therapeutic options for patients. Dedifferentiation of  $\beta$  cells has been observed in genetic mouse models of T1D and T2D as well as in patients with diabetes, and is characterised by the loss of the expression of key maturation marker genes (for example, *Slc2a2* (also known as *Glut2*) and *Ucn3*) and by impaired insulin secretion, and thereby contributes to  $\beta$ -cell dysfunction and hyperglycaemia<sup>8–10</sup>. To investigate whether dysfunctional  $\beta$  cells under hyperglycaemic conditions can be targeted pharmacologically to restore  $\beta$ -cell function, we explored the multiple-low-dose model of streptozotocin-induced diabetes (mSTZ) in mice. STZ specifically ablates  $\beta$  cells, but when STZ is injected in multiple low doses, some residual  $\beta$  cells can survive<sup>11</sup>. Furthermore, the absence of genetic lesions and

<sup>1</sup>Institute of Diabetes and Regeneration Research, Helmholtz Diabetes Center, Helmholtz Center Munich, Neuherberg, Germany. <sup>2</sup>Institute of Diabetes and Obesity, Helmholtz Diabetes Center, Helmholtz Center Munich, Neuherberg, Germany. <sup>3</sup>Division of Metabolic Diseases, Department of Medicine, Technical University of Munich, Munich, Germany. <sup>4</sup>German Center for Diabetes Research (DZD), Neuherberg, Germany. <sup>5</sup>Institute of Stem Cell Research, Helmholtz Center Munich, Neuherberg, Germany. <sup>6</sup>Department of Medicine, Technical University of Munich, Munich, Germany. <sup>7</sup>Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany. <sup>8</sup>School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. <sup>9</sup>Section of Molecular Physiology, Department of Nutrition, Exercise and Sports, University of Copenhagen, Copenhagen, Denmark. <sup>10</sup>RG Adipocytes & Metabolism, Institute for Diabetes & Obesity, Helmholtz Diabetes Center, Helmholtz Center Munich, Neuherberg, Germany. <sup>11</sup>Research Unit Analytical Pathology, Helmholtz Center Munich, Neuherberg, Germany. <sup>12</sup>InSphero AG, Schlieren, Switzerland. <sup>13</sup>Global Drug Discovery, Novo Nordisk A/S, Maaaloev, Denmark. <sup>14</sup>Novo Nordisk Research Center Indianapolis, Indianapolis, IN, USA. <sup>15</sup>Department of Chemistry, Indiana University, Bloomington, IN, USA. <sup>16</sup>Department of Mathematics, Technical University of Munich, Munich, Germany. <sup>17</sup>Medical Clinic and Polyclinic IV, Ludwig Maximilian University of Munich, Munich, Germany. <sup>18</sup>Department of Pharmacology and Experimental Therapy, Institute of Experimental and Clinical Pharmacology and Toxicology, Eberhard Karls University Hospitals and Clinics, Tübingen, Germany. <sup>19</sup>These authors contributed equally: Stephan Sachs, Aimée Bastidas-Ponce, Sophie Tritschler. ✉e-mail: [fabian.theis@helmholtz-muenchen.de](mailto:fabian.theis@helmholtz-muenchen.de); [susanna.hofmann@helmholtz-muenchen.de](mailto:susanna.hofmann@helmholtz-muenchen.de); [timo.mueller@helmholtz-muenchen.de](mailto:timo.mueller@helmholtz-muenchen.de); [heiko.lickert@helmholtz-muenchen.de](mailto:heiko.lickert@helmholtz-muenchen.de)

autoimmunity in the mSTZ model permits the investigation of the fate of those remaining  $\beta$  cells and the effect of pharmacological treatment on  $\beta$ -cell protection and regeneration.

## Results

**Insulin restores  $\beta$ -cell function.** Ten days after the last mSTZ injection, mice were severely hyperglycaemic (Extended Data Fig. 1a) and showed an impaired islet architecture (Extended Data Fig. 1b) with markedly decreased  $\beta$ -cell numbers in comparison to those in control mice (Extended Data Fig. 1b,c). Proliferation of remaining  $\beta$  cells was unchanged (Extended Data Fig. 1d), whereas  $\beta$ -cell apoptosis was significantly increased in mSTZ-treated mice compared to control mice (Extended Data Fig. 1e) and was accompanied by a loss of identity and function (Extended Data Fig. 1f–h). Hence, at this time point (diabetes onset), when a fraction of dysfunctional  $\beta$  cells were still remaining, we initiated a permanent drug treatment over 100 d (Fig. 1a). Vehicle-treated mSTZ mice remained diabetic over the length of the study, which suggests that the residual  $\beta$ -cell functionality or endogenous  $\beta$ -cell regeneration<sup>12,13</sup> is insufficient to maintain or restore sufficient glucose homeostasis (Fig. 1b–f). To correct the insulin deficiency in mSTZ mice, we treated diabetic mice with a long-acting pegylated insulin analogue (PEG–insulin, once daily), which improved glycaemia (Fig. 1b), increased C-peptide levels (Fig. 1c), improved islet structure (Fig. 1d) and increased the number of insulin-positive cells (Fig. 1e,f). This shows functional  $\beta$ -cell recovery upon glycaemia normalisation, which extends findings from insulin treatment in genetic mouse models<sup>14,15</sup>. However, the risk of hypoglycaemia and unwanted weight gain are undesirable hallmarks of insulin therapy, and thus alternative pharmacological approaches are required to mitigate these possible effects.

**GLP-1–oestrogen and insulin polypharmacotherapy.** Oestrogen and glucagon-like peptide 1 (GLP-1) have been repeatedly implicated in the treatment of diabetes due to insulinotropic and  $\beta$ -cell protective effects in preclinical studies<sup>16,17</sup>. Chemically optimised GLP-1 analogues profoundly improve glucose and body weight management in obese people and individuals with T2D<sup>18</sup>. However, severe gynaecological, oncogenic and mitogenic side effects preclude chronic oestrogen use as a drug for diabetes and, as yet, GLP-1

analogues have failed to preserve  $\beta$ -cell function and mass in obese and diabetic humans<sup>17</sup>.

To circumvent the gynaecological, oncogenic and mitogenic actions of oestrogen, we recently designed and evaluated a stable GLP-1–oestrogen conjugate, which reversed the metabolic syndrome in diet-induced obese male and female mice<sup>19</sup>. Here, we used the GLP-1–oestrogen conjugate to test whether the specific delivery of oestrogen into the GLP-1 receptor protein (GLP-1R)-expressing  $\beta$  cells could restore  $\beta$ -cell functionality. GLP-1–oestrogen treatment for 100 d was more efficacious in decreasing fasting glucose (Fig. 1b) and increasing fasting C-peptide (Fig. 1c) and insulin levels (Extended Data Fig. 2a) than either of the monoagonists (oestrogen or GLP-1 alone). Moreover, only GLP-1–oestrogen treatment improved pancreatic islet architecture (Fig. 1d) and increased  $\beta$ -cell number, as compared with those in mSTZ-diabetic mice treated with vehicle (Fig. 1e,f). These effects were independent of body weight loss (Extended Data Fig. 2b).

Polypharmacotherapy holds the potential to simultaneously activate redundant or additive pathways to enhance efficacy, and to enable reduced dosing of the individual components and consequently reduce the risk of unwanted side effects<sup>20</sup>. We tested the combination of insulin and GLP-1–oestrogen to investigate a triple pharmacological approach to enhance the efficacy of both compounds, and particularly to lessen the amount of insulin required. The combination therapy normalised glycaemia (Fig. 1b) and increased C-peptide levels (Fig. 1c), as compared with those in mSTZ-diabetic mice treated with vehicle. Furthermore, the combination therapy displayed a superior effect compared to treatment with insulin alone to limit weight gain (Extended Data Fig. 2b), normalise glucose tolerance (Extended Data Fig. 2c), increase pancreatic insulin content (Extended Data Fig. 2d) and increase  $\beta$ -cell number (Fig. 1e,f). Importantly, we were able to reduce the insulin dose by 60% (10 nmol kg<sup>-1</sup>) compared with that in the insulin monotherapy (25 nmol kg<sup>-1</sup>) and still achieve superior therapeutic outcomes, which reduces the risk of hypoglycaemia and unintended weight gain.

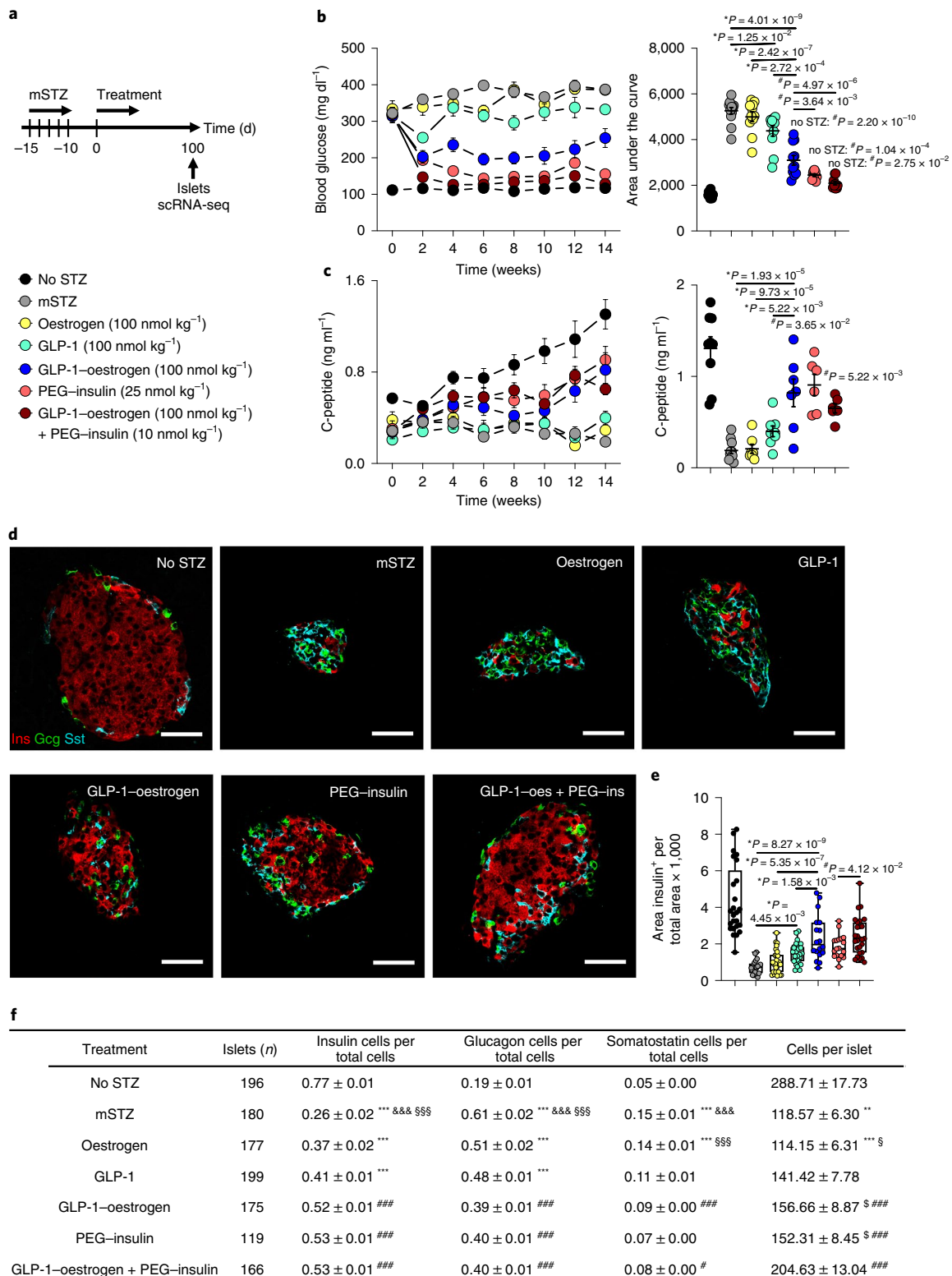
To test whether treatment-induced improvements on glucose and islet homeostasis are maintained, we switched a group of mice after 12 weeks of GLP-1–oestrogen treatment to two additional weeks of vehicle injections. Positive effects of GLP-1–oestrogen treatment to

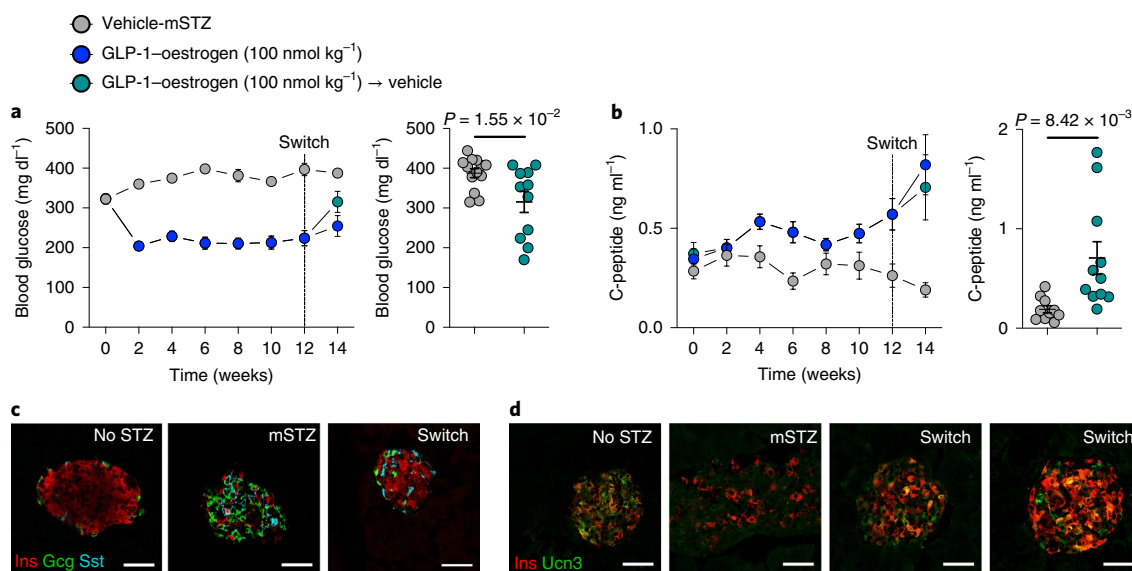
**Fig. 1 | GLP-1–oestrogen and PEG–insulin treatment regenerates functional  $\beta$ -cell mass.** **a**, Treatment scheme (no STZ-vehicle,  $n=12$ ; mSTZ-vehicle,  $n=13$ ; GLP-1,  $n=11$ ; oestrogen,  $n=11$ ; GLP-1–oestrogen,  $n=11$ ; PEG–insulin,  $n=9$ ; GLP-1–oestrogen and PEG–insulin,  $n=10$ ). Effects on fasting blood glucose (**b**) and fasting C-peptide (**c**). Data are mean  $\pm$  s.e.m. Data were analysed by a one-way ANOVA with Tukey post-hoc. \* $P$  among mSTZ, oestrogen, GLP-1 and GLP-1–oestrogen treatment:  $F_{(c)}$  (d.f.n, d.f.d) =  $F(3, 42) = 24.09$ ;  $F_{(c)}$  (3, 28) = 13.09. # $P$  among no STZ, GLP-1–oestrogen, PEG–insulin and cotreatment:  $F_{(c)}$  (3, 38) = 29.32;  $F_{(c)}$  (3, 24) = 5.45. **d**, Immunostaining of Ins, Gcg and Sst in pancreatic sections at the end of the study. Images are representative of mice treated with no STZ-vehicle ( $n=3$ ), mSTZ-vehicle ( $n=3$ ), oestrogen ( $n=3$ ), GLP-1 ( $n=3$ ), GLP-1–oestrogen ( $n=2$ ) or PEG–insulin ( $n=2$ ), and of mice cotreated with GLP-1–oestrogen and PEG–insulin ( $n=3$ ). Scale bars, 50  $\mu$ m. Images are representative of data set shown in **f**. **e**, Total insulin area in pancreatic sections (no STZ-vehicle: 26 sections of  $n=3$  mice; mSTZ-vehicle: 21,  $n=3$ ; oestrogen: 27,  $n=3$ ; GLP-1: 26,  $n=3$ ; GLP-1–oestrogen: 18,  $n=2$ ; PEG–insulin: 18,  $n=2$ ; GLP-1–oestrogen (GLP-1–oes) and PEG–insulin (PEG–ins): 27,  $n=3$ ). \* $P$  among mSTZ, oestrogen, GLP-1 and GLP-1–oestrogen treatment; data were analysed by a one-way ANOVA with Tukey post-hoc,  $F(3, 88) = 17.66$ . # $P$  between PEG–insulin and cotreatment; data were analysed by an unpaired two-sided  $t$ -test,  $t = 2.11$ , d.f. = 43. **f**, Endocrine islet cell composition at study end (no STZ-vehicle: 196 islets of  $n=3$  mice; mSTZ-vehicle: 180,  $n=3$ ; oestrogen: 177,  $n=3$ ; GLP-1: 199,  $n=3$ ; GLP-1–oestrogen: 175,  $n=2$ ; PEG–insulin: 119,  $n=2$ ; GLP-1–oestrogen and PEG–insulin: 166,  $n=3$ ). Data are mean  $\pm$  s.e.m. Data were analysed by a one-way ANOVA with Tukey post-hoc.  $\beta$  cells: mSTZ versus GLP-1–oestrogen (\*\*\* $P < 1 \times 10^{-15}$ ), mSTZ versus oestrogen (<sup>666</sup> $P = 4.17 \times 10^{-9}$ ), mSTZ versus GLP-1 (<sup>666</sup> $P < 1 \times 10^{-15}$ ), oestrogen versus GLP-1–oestrogen (\*\*\* $P < 1 \times 10^{-15}$ ), GLP-1 versus GLP-1–oestrogen (\*\*\* $P < 1 \times 10^{-15}$ ), GLP-1–oestrogen versus no STZ (### $P < 1 \times 10^{-15}$ ), PEG–insulin versus no STZ (#### $P < 1 \times 10^{-15}$ ), GLP-1–oestrogen and PEG–insulin versus no STZ (### $P < 1 \times 10^{-15}$ );  $F_{\beta\text{-cells}}$  (6, 1,187) = 175.6.  $\alpha$ -cells: mSTZ versus GLP-1–oestrogen (\*\*\* $P < 1 \times 10^{-15}$ ), mSTZ versus oestrogen (<sup>666</sup> $P = 1.01 \times 10^{-8}$ ), mSTZ versus GLP-1 (<sup>666</sup> $P < 1 \times 10^{-15}$ ), oestrogen versus GLP-1–oestrogen (\*\*\* $P = 6.30 \times 10^{-10}$ ), GLP-1 versus GLP-1–oestrogen (\*\*\* $P = 2.48 \times 10^{-6}$ ), GLP-1–oestrogen versus no STZ (### $P < 1 \times 10^{-15}$ ), PEG–insulin versus no STZ (#### $P < 1 \times 10^{-15}$ ), GLP-1–oestrogen and PEG–insulin versus no STZ (### $P < 1 \times 10^{-15}$ );  $F_{\alpha\text{-cells}}$  (6, 1,202) = 124.8.  $\delta$ -cells: mSTZ versus GLP-1–oestrogen (\*\*\* $P = 4.00 \times 10^{-12}$ ), mSTZ versus GLP-1 (<sup>666</sup> $P = 5.91 \times 10^{-6}$ ), GLP-1–oestrogen versus no STZ (### $P = 9.89 \times 10^{-7}$ ), GLP-1–oestrogen and PEG–insulin versus no STZ (<sup>6</sup> $P = 1.91 \times 10^{-2}$ );  $F_{\delta\text{-cells}}$  (6, 1,144) = 45.87. Cells per islet (mSTZ, oestrogen, GLP-1 and GLP-1–oestrogen): mSTZ versus GLP-1–oestrogen (\*\* $P = 2.07 \times 10^{-3}$ ), oestrogen versus GLP-1–oestrogen (\*\*\* $P = 4.43 \times 10^{-4}$ ), oestrogen versus GLP-1 (<sup>6</sup> $P = 4.28 \times 10^{-2}$ ),  $F(3, 727) = 7.049$ . Cells per islet (all treatments): PEG–insulin versus GLP-1–oestrogen and PEG–insulin (<sup>6</sup> $P = 4.05 \times 10^{-2}$ ), GLP-1–oestrogen versus GLP-1–oestrogen and PEG–insulin (<sup>6</sup> $P = 3.49 \times 10^{-2}$ ), GLP-1–oestrogen versus no STZ (### $P < 1 \times 10^{-15}$ ), PEG–insulin versus no STZ (#### $P < 1 \times 10^{-15}$ ), GLP-1–oestrogen and PEG–insulin versus no STZ (### $P = 7.89 \times 10^{-7}$ );  $F(6, 1,205) = 33.94$ . Additional information is available in Source data.

reduce fasting glycaemia (Fig. 2a), increase fasting C-peptide levels (Fig. 2b), and enhance the  $\beta$ -cell maturation state (Fig. 2c,d) were sustained after these two weeks, which supports the notion of preserved islet cell function even after treatment cessation.

**GLP-1-oestrogen targets  $\beta$  cells.** We next wanted to confirm the absence of systemic toxicity that was related to the oestrogen

component of the GLP-1-oestrogen conjugate, a pre-requisite for clinical use. To that end, we investigated whether GLP-1-oestrogen (doses up to 10 $\times$  higher than are generally used in mouse experiments, and at least 1,000 $\times$  the plasma oestradiol exposure as compared with that in women on hormone replacement therapy (see Methods)) increased uterus weight in ovariectomised (OVX) rats after two weeks of treatment (Extended Data Fig. 3a and Methods).





**Fig. 2 | Sustained effects of GLP-1-oestrogen to ameliorate mSTZ diabetes in mice.** A group of GLP-1-oestrogen-treated mice were switched from daily GLP-1-oestrogen treatment for 12 weeks to vehicle treatment for another two weeks (mSTZ-vehicle,  $n=13$ ; GLP-1-oestrogen,  $n=11$ ; GLP-1-oestrogen to vehicle (switch),  $n=11$ ). **a**, Fasting blood glucose at week 14. Data were analysed by an unpaired two-sided  $t$ -test,  $t=2.623$ , d.f. = 22 (mSTZ-vehicle,  $n=13$ ; GLP-1-oestrogen to vehicle,  $n=11$ ). Data are mean  $\pm$  s.e.m. **b**, Fasting C-peptide levels at week 14. Data were analysed by an unpaired two-sided  $t$ -test,  $t=2.939$ , d.f. = 19 (mSTZ-vehicle,  $n=10$ ; GLP-1-oestrogen to vehicle,  $n=11$ ). Data are mean  $\pm$  s.e.m. **c**, Ins, Gcg and Sst immunohistochemistry of representative pancreatic islets of the data set plotted in **b** at the end of the study. **d**, Ucn3 immunohistochemistry of representative pancreatic islets of the data set plotted in **b** at the end of the study. In **c** and **d**, images are representative of mice treated with no STZ-vehicle ( $n=3$ ), mSTZ-vehicle ( $n=3$ ) and GLP-1-oestrogen to vehicle ( $n=3$ ). Scale bars, 50  $\mu$ m. Additional information is available in Source data.

In contrast to treatment with oestrogen alone, no treatment-related effect was observed with the GLP-1-oestrogen conjugate (Extended Data Fig. 3b), which is consistent with previously reported results in OVX mice<sup>19</sup>.

To confirm  $\beta$ -cell-specific targeting of the GLP-1-oestrogen conjugate, we used a double knock-in fluorescent reporter mouse model (Foxa2-Venus Fusion (FVF)  $\times$  Pdx1-BFP (blue fluorescent protein) fusion (PBF); FVFPBF<sup>DHom</sup>)<sup>21</sup>, which allows  $\alpha$ - and  $\beta$ -cell sorting (Extended Data Fig. 3c,d). Male FVFPBF<sup>DHom</sup> mice develop maturity onset diabetes of the young owing to reduced Pdx1 levels in islets, accompanied by hyperglycaemia, reduced  $\beta$ -cell number and impaired islet architecture at weaning age<sup>21</sup>. In this genetic diabetes model, none of the therapies used in this study improved glycaemia after four weeks of treatment (Extended Data Fig. 3e,f). These results suggest that  $\beta$ -cell function cannot be restored pharmacologically in the presence of the genetically induced  $\beta$ -cell lesions in the FVFPBF<sup>DHom</sup> mice. However, the GLP-1-oestrogen conjugate, but not the monoagonists, specifically increased  $\beta$ -cell granularity of FVFPBF<sup>DHom</sup> mice in comparison to control mice, which shows that the conjugate selectively targets  $\beta$  cells (Extended Data Fig. 3g).

**GLP-1-oestrogen improves human  $\beta$ -cell function.** To provide human relevance to the findings, we tested GLP-1-oestrogen and the monoagonists in human micro-islets in the absence or presence of  $\beta$ -cell stressors (cytokine cocktail, see Methods). After acute compound exposure, GLP-1-oestrogen was more potent than either of the individual components in increasing glucose-stimulated insulin secretion (GSIS) from human micro-islets in comparison to that in untreated and monoagonist-treated micro-islets (Fig. 3a). We next exposed the human micro-islets to cytokines to determine whether the beneficial effects of GLP-1-oestrogen protect against stress-induced impairment of  $\beta$ -cell functionality (Fig. 3b). Seven-day treatment with GLP-1-oestrogen enhanced GSIS, and exceeded the effects that were seen with either of the individual components (Fig. 3b). Moreover, only GLP-1-oestrogen treatment increased the total insulin

content of cytokine-exposed human micro-islets (Fig. 3c). This was independent of changes in the total ATP content (Extended Data Fig. 4a) and in caspase luciferase activity (Extended Data Fig. 4b), which suggests that the compound treatment improved functionality, but did not improve cell survival of human micro-islets. These results show that GLP-1-oestrogen is superior to both of the monoagonists in improving  $\beta$ -cell function in homeostasis, and in acting upon cytokine stress in both mice and humans.

**$\beta$ -cell heterogeneity in homeostatic and healthy mice.** To elucidate the molecular mechanisms that underlie  $\beta$ -cell failure in mSTZ diabetes, and  $\beta$ -cell recovery after the different therapeutic approaches, we performed scRNA-seq of isolated islets from mice that responded to treatment (Extended Data Fig. 5).

In normal islet homeostasis, we identified the four main endocrine cell subtypes,  $\alpha$  cells,  $\beta$  cells,  $\delta$  cells and pancreatic polypeptide (PP) cells, by unbiased graph-based clustering. Clusters were annotated on the basis of predominant endocrine hormone expression of glucagon (Gcg), insulin (Ins), somatostatin (Sst) and PP (Extended Data Fig. 6a,b and Methods). For each of the four endocrine subtypes we identified a specific marker gene signature (Supplementary Table 1 and Methods). Refined clustering of insulin-positive cells revealed the presence of two main  $\beta$ -cell subpopulations,  $\beta$ 1 and  $\beta$ 2 (Extended Data Fig. 6c and Methods). Single-cell trajectory inference suggests a continuum of transcriptional states, rather than discrete phenotypes within  $\beta$  cells and a transition between  $\beta$ 1 and  $\beta$ 2 subpopulations (Extended Data Fig. 6d and Methods). We found a progressive increase in expression of  $\beta$ -cell maturation marker genes (for example *Ins1*, *Ins2* and *Ucn3* (ref. <sup>22</sup>)), and genes of the secretion machinery (*G6pc2*, *Sytl4* and *Slc2a2*), as well as a concomitant decrease in expression of the  $\beta$ -cell immaturity gene *Mafb*<sup>23</sup>, and pan-endocrine lineage marker genes (*Chga* and *Chgb*) along the pseudotime trajectory from  $\beta$ 2- to  $\beta$ 1-cells (Extended Data Fig. 6d and Supplementary Table 2). The expression of transcription factor genes that are associated with  $\beta$ -cell identity (for example *Pdx1*,

*Nkx6.1* and *NeuroD1*) was unchanged (Extended Data Fig. 6d). The  $\beta_2$ -cell cluster was characterised by the downregulation of genes that are involved in insulin secretion, oxidative phosphorylation and cell-cycle inhibition, as well as by the upregulation of genes that are involved in cAMP and WNT signalling (Extended Data Fig. 6e), both of which observations are suggestive of a more immature and/or proliferative state for  $\beta_2$  cells<sup>24</sup>. We consistently observed upregulation of cell-cycle-associated genes, such as *Ki67* and *Cdk1* in the immature  $\beta_2$  subpopulation and, in accordance with this result, 16 out of 403 of the  $\beta_2$ -cells, but only 2 out of 5,319 of the mature  $\beta_1$ -cells, were classified as being involved in cell cycling (Extended Data Fig. 6f (see Methods)). Taken together, these results confirm the co-existence of mature ( $\beta_1$ ) and immature and/or proliferative  $\beta$  cells ( $\beta_2$ ) in healthy mouse islets<sup>24,25</sup>. In addition, we found subpopulations of polyhormonal cells that could be distinguished from doublets and ambient RNA, both of which are common problems with the scRNA-seq technology (Extended Data Fig. 6g,h and Methods).

**$\beta$ -cell dedifferentiation at the single-cell level.** To obtain a deeper understanding of the cell autonomous and non-cell autonomous effects that underlie chemical  $\beta$ -cell ablation in the islet cell niche, we performed a scRNA-seq survey of diabetic islets after 100 d of persistent hyperglycaemia. Unsupervised clustering and embedding of the scRNA-seq data revealed altered endocrine subtype composition and cell-intrinsic gene expression profiles that were indicated by a shifted cell cluster location in the Uniform Manifold Approximation and Projection (UMAP) space of mSTZ-diabetic compared to healthy mice (Fig. 4a and Supplementary Table 3). In particular, there was a threefold decrease in the proportion of  $\beta$  cells from mSTZ-treated mice ( $\beta$ -mSTZ), and  $\beta$ -mSTZ formed a cluster that was clearly distinct from healthy  $\beta$  cells (Fig. 4a). In contrast to the results for  $\beta$  cells, few transcriptional changes were detected for  $\alpha$  cells,  $\delta$  cells and PP cells (Extended Data Fig. 7a-f and Supplementary Table 3). We next sought to describe the progression from healthy to dysfunctional  $\beta$  cells and the associated gene expression changes, using single-cell trajectory inference. Cells were ordered on the basis of a cell-to-cell distance metric that was calculated using the concept of diffusion pseudotime (see Methods). We identified a cellular trajectory in which cells transitioned from mature to immature to  $\beta$ -mSTZ cells (Fig. 4b). Remaining  $\beta$ -mSTZ cells expressed low *Ins1* and/or *Ins2* messenger RNA and also showed sustained low expression of  $\beta$ -cell identity transcription factor genes, such as *Pdx1*, *Nkx2.2*, *Nkx6.1*, *Pax6*, *Isl1* and *NeuroD1* (refs. 26–31) (Fig. 4b). Pathways that are associated with  $\beta$ -cell maturity and functionality included genes with down-regulated expression, whereas genes in ER stress and oxidative phosphorylation pathways were upregulated in the remaining

$\beta$ -mSTZ cells; together these observations indicate an ER stress response and  $\beta$ -cell dysfunction (Fig. 4c and Supplementary Table 3). Along this trajectory, expression of key markers of  $\beta$ -cell maturity and functionality gradually decreased concomitantly with an increase in the very few known markers of  $\beta$ -cell immaturity and dedifferentiation (for example *Aldh1a3* (ref. 32) and *gastrin*<sup>33</sup>) (Fig. 4b,d). Strikingly, our single-cell analysis uncovered a large number of upregulated genes and pathways in  $\beta$ -mSTZ cells that are not expressed at all or are expressed only subtly in mature, functional murine  $\beta$  cells (Fig. 4e). We confirmed the increased expression of, for example, *Cck* and *Slc5a10* proteins by immunohistochemistry in mSTZ diabetic mice (Fig. 4f). These identified targets may serve as biomarkers for dysfunctional  $\beta$  cells, and may have the potential to be part of druggable pathways to restore  $\beta$ -cell function. Some of these were also identified recently in  $\beta$  cells and pancreata of T1D and T2D human specimens<sup>34,35</sup> (Extended Data Fig. 7g,h). There is some debate as to whether  $\beta$ -cell dedifferentiation resembles reversal to a pluripotent (*Oct3/4*, *Nanog*, *Sox2*) or endocrine progenitor state (*Neurogenin 3* or *Neurog 3*, hereafter called *Ngn3*), whether it is part of normal phenotypic variation described as  $\beta$ -cell heterogeneity, or whether it resembles a glucotoxic-induced reversible state<sup>36</sup>. Although dedifferentiated  $\beta$  cells had been characterised previously by upregulation of pluripotency or endocrine transcription factors<sup>8,9,15</sup>, in our study expression levels of *Sox9*, *Pou5f1* (*Oct3/4*), *Myc* and *Ngn3* genes were unaltered in mSTZ-treated  $\beta$  cells (Extended Data Fig. 7i).

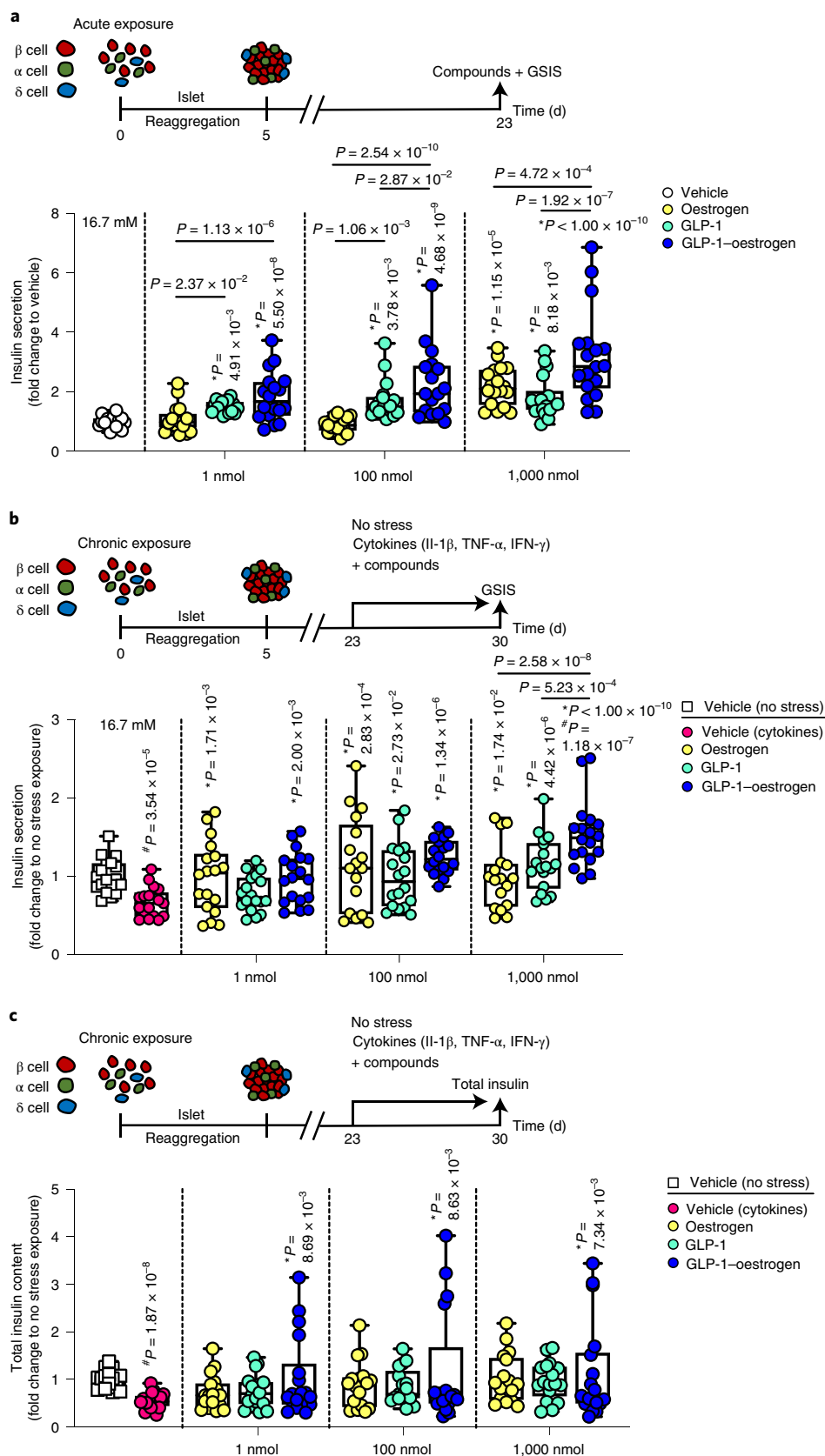
To further characterise the maturation state of mSTZ-derived  $\beta$  cells, we compared our data set to  $\beta$ -cell expression profiles during embryonic (E17.5) to postnatal development (P60)<sup>37</sup>. We assessed transcriptional similarity of  $\beta$ -cell subpopulations using Partition-based graph abstraction (PAGA) after data integration and computation of a common embedding. The enforcement of integration using only genes associated to  $\beta$ -cell maturation in the reference data allowed us to match maturation states independent of differences in other biological processes between the two data sets (see Methods). PAGA uses a statistical model to measure the relatedness of groups of single cells (see Methods). We found that dedifferentiated  $\beta$ -mSTZ cells were more strongly connected to early time points of the maturation data set, whereas  $\beta_2$  cells clustered to intermediate time points and  $\beta_1$  cells to late time points (Fig. 5a). Remarkably, the inferred cellular trajectory from  $\beta$ -mSTZ to  $\beta_2$  to  $\beta_1$  cells aligned with the trajectory of  $\beta$ -cell maturation from embryonic (E17.5) to mature  $\beta$  cells (P60) of the reference data set (Fig. 5b). An increase in known  $\beta$ -cell maturation and a decrease of immaturity markers along this trajectory further indicated that the cells follow a similar differentiation or dedifferentiation programme (Fig. 5c,d). To validate these findings, we scored each  $\beta$  cell using the gene sets that are characteristic for the start (E17.5/P0) and end

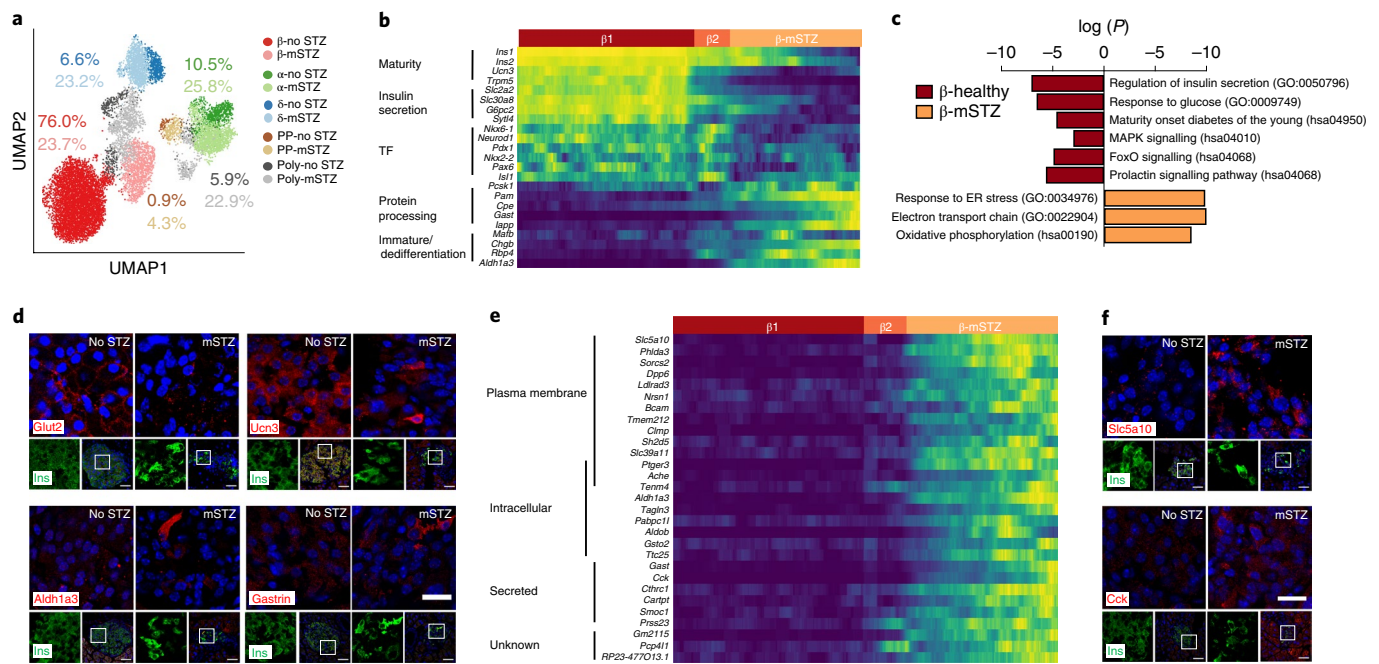
**Fig. 3 | GLP-1-oestrogen improves function of human micro-islets.** **a**, Insulin secretion of human micro-islets after acute exposure to 16.7 mM glucose and vehicle, oestrogen, GLP-1 or the GLP-1-oestrogen treatment.  $n=5-6$  micro-islets of  $n=3$  human donors for each condition. Secretion (mean  $\pm$  s.e.m.) of donor 1 =  $0.36 \pm 0.01$  ng ml<sup>-1</sup>, donor 2 =  $0.37 \pm 0.02$  ng ml<sup>-1</sup> and donor 3 =  $0.29 \pm 0.03$  ng ml<sup>-1</sup> after vehicle exposure. Box plot of all data points. Line indicates the median. Data were analysed by a one-way ANOVA with donor as random effect followed by Tukey post-hoc ( $F_{\text{low dose}}(3, 65) = 14.04$ ;  $F_{\text{medium dose}}(3, 64) = 18.59$ ;  $F_{\text{high dose}}(3, 66) = 25.50$ ). \* $P$  indicates  $P$  value to vehicle. **b**, Insulin secretion at 16.7 mM glucose of human micro-islets after 7-d exposure to cytokine stress and effect of oestrogen, GLP-1 or GLP-1-oestrogen treatment.  $n=6$  micro-islets of  $n=3$  human donors for each condition. Secretion (mean  $\pm$  s.e.m.) of donor 1 =  $0.13 \pm 0.02$  ng ml<sup>-1</sup>, donor 2 =  $0.56 \pm 0.04$  ng ml<sup>-1</sup> and donor 3 =  $0.51 \pm 0.06$  ng ml<sup>-1</sup> after chronic vehicle (no stress) exposure. Box plot of all data points. Line indicates the median. \* $P$  indicates  $P$  value to vehicle under the no stress condition. \* $P$  indicates  $P$  value to vehicle cytokine exposure. Data from no stress versus cytokine stress were analysed by an unpaired two-sided  $t$ -test ( $t = 4.776$ , d.f. = 33). Otherwise, data were analysed by a one-way ANOVA with donor as random effect followed by Tukey post-hoc ( $F_{\text{low dose}}(4, 82) = 6.35$ ;  $F_{\text{medium dose}}(4, 81) = 7.65$ ;  $F_{\text{high dose}}(4, 81) = 21.91$ ). **c**, Total insulin content of human micro-islets after 7-d exposure to cytokine stress and effect of oestrogen, GLP-1 or GLP-1-oestrogen treatment.  $n=6$  micro-islets of  $n=3$  human donors for each condition. Insulin content (mean  $\pm$  s.e.m.) of donor 1 =  $41.11 \pm 3.73$  ng per islet, donor 2 =  $30.86 \pm 3.36$  ng per islet and donor 3 =  $82.73 \pm 3.99$  ng per islet after chronic vehicle (no stress) exposure. Box plot of all data points. Line indicates the median. \* $P$  indicates  $P$  value to vehicle no stress condition. \* $P$  indicates  $P$  value to vehicle cytokine exposure. Data from no stress versus cytokine stress were analysed by an unpaired two-sided  $t$ -test ( $t = 7.429$ , d.f. = 32). Otherwise, data were analysed by a one-way ANOVA with donor as random effect followed by Tukey post-hoc ( $F_{\text{low dose}}(4, 80) = 4.12$ ;  $F_{\text{medium dose}}(4, 78) = 3.01$ ;  $F_{\text{high dose}}(4, 79) = 3.31$ ). Additional information is available in Source data.



(P60) point of the developmental trajectory (Supplementary Table 4 and Methods). Healthy mature  $\beta$  cells ( $\beta$ 1) scored high for maturity genes, whereas healthy immature cells ( $\beta$ 2) and dedifferentiated cells ( $\beta$ -mSTZ) scored higher for the embryonic immaturity gene

set (Fig. 5e). Taken together, these results imply that during the transition from healthy  $\beta$ 1 to  $\beta$ 2 to dedifferentiated  $\beta$ -mSTZ,  $\beta$  cells revert, at least in part, back to a more immature and, further, to an embryonic-like state.





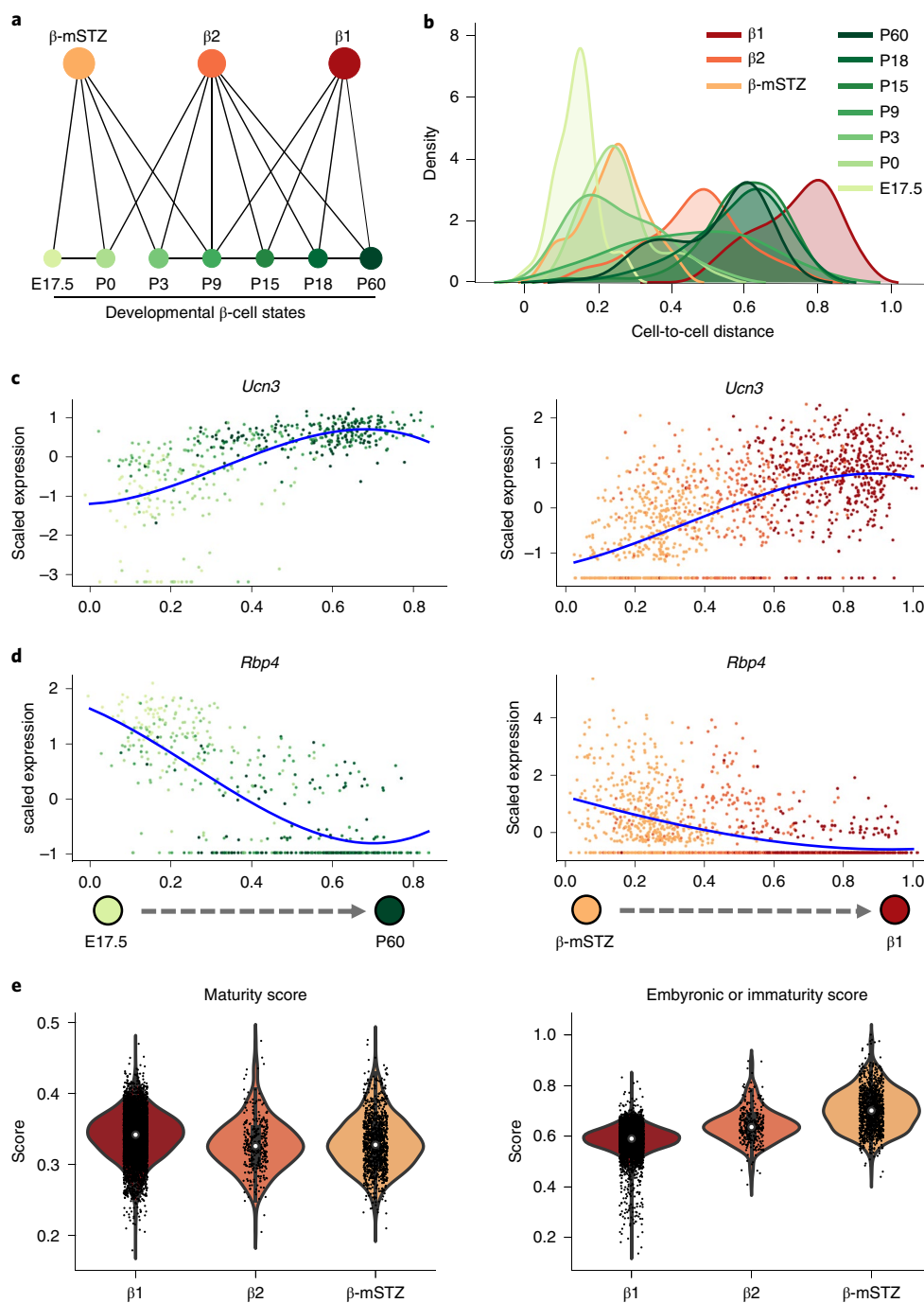
**Fig. 4 |  $\beta$ -cell dedifferentiation in mSTZ-diabetic mice.** **a**, UMAP plot of 12,430 cells from healthy and STZ-treated mice. Colours highlight endocrine cell clusters, and colour tone distinguishes no STZ (dark colours) or mSTZ treatment (light colours). Values indicate the proportions of each cell cluster in no STZ or mSTZ treatment. Poly, polyhormonal cells. **b**, Gene expression changes of representative markers of  $\beta$ -cell identity, maturation and function along the trajectory from mature ( $\beta$ 1) to immature ( $\beta$ 2) to  $\beta$ -mSTZ indicating a continuous transition. TF, transcription factor. Cells are ordered on the basis of a random-walk-based cell-to-cell distance metric. Expression is shown as the running average along the inferred trajectory scaled to the maximum observed level per gene.  $\beta$ 1-cells were downsampled to 1,500 cells for better visualisation. **c**, Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of significantly up- and downregulated genes in  $\beta$  cells of mSTZ-treated mice compared to  $\beta$  cells of healthy mice (absolute  $\log(\text{fold change}) > 0.25$ ,  $\text{FDR} < 0.01$ ; selective pathways are depicted; see also Supplementary Table 3). Cells were pooled from mice treated with no STZ-vehicle ( $n = 3$ ) and mSTZ-vehicle ( $n = 3$ ). We used limma-trend to find differentially expressed genes (see Methods). Gene enrichment was done with EnrichR using Fisher's exact test to identify regulated ontologies or pathways (see Methods). **d**, Immunohistochemical analysis of Glut2, Ucn3, Aldh1a3 and gastrin in  $\beta$  cells of mSTZ mice and healthy mice at study end. Images are representative of mice treated with no STZ-vehicle ( $n = 3$ ) and mSTZ-vehicle ( $n = 3$ ). Scale bars, 50  $\mu\text{m}$ ; scale bar zoom-in, 20  $\mu\text{m}$ . **e**, Gene expression along the trajectory from  $\beta$ 1 to  $\beta$ -mSTZ (as in **c**) of 29 genes specifically expressed in  $\beta$  cells from mSTZ-treated mice (expression in  $< 5\%$  of no-STZ- $\beta$  cells and in  $> 25\%$  of  $\beta$ -mSTZ cells, see Methods). Cellular locations of associated proteins are indicated. **f**, Immunohistochemical analysis of Slc5a10 and Cck in  $\beta$  cells of mSTZ and healthy mice at study end. Images are representative of mice treated with no STZ-vehicle ( $n = 3$ ) and mSTZ-vehicle ( $n = 3$ ). Scale bars, 50  $\mu\text{m}$ ; scale bar zoom-in, 20  $\mu\text{m}$ .

To separate the altered maturation state from other processes induced in dedifferentiated  $\beta$  cells, we compared differentially regulated gene ontologies and pathways between embryonic (E17.5/P0) and mature (P60)  $\beta$  cells and between  $\beta$ -mSTZ cells and healthy control  $\beta$  cells (see Methods). According to the trajectory and PAGA analysis, these  $\beta$ -cell states of the reference data set correspond best to dedifferentiated  $\beta$ -mSTZ cells and mature, healthy  $\beta$  cells, respectively. Embryonic and mSTZ-diabetic  $\beta$  cells shared downregulation of molecular processes connected to  $\beta$ -cell function and maturity (for example, insulin secretion and FoxO signalling) compared to healthy mature  $\beta$  cells (Extended Data Fig. 8a and Supplementary Table 4), whereas genes involved in oxidative phosphorylation and gene and protein transcription were upregulated (Extended Data Fig. 8b and Supplementary Table 4). Specific to embryonic  $\beta$  cells was a downregulation of lipid and carbohydrate metabolism and an upregulation of WNT signalling (Extended Data Fig. 8a,b). This corresponds to known mechanisms of  $\beta$ -cell maturation during embryogenesis<sup>37</sup>. Interestingly, in mSTZ-diabetic  $\beta$  cells, but not the embryonic  $\beta$  cells, we found an upregulation of pathways and ontologies associated with ER stress and a decreased expression of genes involved in insulin and MAPK signalling in comparison to healthy mature  $\beta$  cells (Extended Data Fig. 8a,b).

Thus,  $\beta$ -cell dedifferentiation involves partial reversal to an embryonic or immature  $\beta$ -cell programme and upregulation of an

ER stress response and altered signalling state. These results suggest that surviving  $\beta$  cells are dedifferentiated, which shows that mSTZ-induced diabetes is a good model in which to study mechanisms of  $\beta$ -cell dedifferentiation and redifferentiation in the absence of genetic lesions.

**Mechanisms of  $\beta$ -cell redifferentiation.** In line with the pharmacological data, single-cell analysis of the different treatments revealed that  $\beta$  cells of mice treated with vehicle (Extended Data Fig. 9a), oestrogen (Extended Data Fig. 9b) and GLP-1 (Extended Data Fig. 9c) remained dedifferentiated. By contrast, we observed an increased fraction of immature  $\beta$ 2-cells from GLP-1-oestrogen-treated mice (Extended Data Fig. 9d). In PEG-insulin-treated mice (Extended Data Fig. 9e) and GLP-1-oestrogen plus PEG-insulin-cotreated mice (Extended Data Fig. 9f), almost no dedifferentiated  $\beta$  cells remained and most cells clustered with immature  $\beta$ 2-cells. To further assess the transcriptional state of  $\beta$  cells from the treated mice, we calculated a cell-to-cell distance so that cells could be ordered along the cellular trajectory from dedifferentiated to healthy  $\beta$  cells (see Methods). On this one-dimensional axis,  $\beta$  cells of mice treated with PEG-insulin or with the combination of PEG-insulin and GLP-1-oestrogen were located closest to  $\beta$  cells from healthy mice (Fig. 6a–c). This transcriptional similarity to healthy  $\beta$  cells was further supported by PAGA (see Methods). In the PAGA graph,



**Fig. 5 | mSTZ-derived  $\beta$  cells are similar to immature  $\beta$  cells from embryonic or immature islets.** **a**, Transcriptional similarity of  $\beta$ -cell subpopulations and  $\beta$  cells sampled between E17.5 and P60 (according to ref. <sup>37</sup>) inferred on the basis of a measure for cluster connectivity using PAGA. Edge weights indicate significance. Edge weights  $> 0.7$  are shown. **b**, Cell density of  $\beta$ -cell clusters along a cellular trajectory reflecting  $\beta$ -cell maturation. Cells are ordered on the basis of a random-walk-based cell-to-cell distance metric. Data sets were integrated by computing a common single-cell neighbourhood graph (see Methods). **c,d**, Expression of a  $\beta$ -cell maturity (**c**) and immaturity (**d**) marker along the inferred trajectory in **b**. Left panels show  $\beta$ -cell subpopulations sampled between E17.5 and P60, right panels show  $\beta$ -cell subpopulations from healthy and mSTZ-treated mice. Dots show expression levels of individual cells coloured by  $\beta$ -cell subpopulations. Red lines approximate expression along the trajectory by polynomial regression fits. **e**, Violin plots show the distribution of cell scores characteristic of  $\beta$ -cell maturity (left) and embryonic-like or immaturity (right). Scores were calculated on the basis of differentially expressed genes between clusters of mature (P60) and embryonic or immature (E17.5 or P0) cells. Violin shows the distribution as a kernel density estimate fit. Points in the violin interior show individual data points. Box plot in the violin interior shows median, quartile and whisker values.

$\beta$  cells of PEG-insulin-treated and GLP-1-oestrogen plus PEG-insulin-cotreated mice showed the strongest connection to healthy  $\beta$  cells (Fig. 6d). The observed overall re-establishment of the healthy  $\beta$ -cell expression profiles was substantiated by an increased

expression of  $\beta$ -cell maturity markers and decreased expression of immaturity and dedifferentiation markers along the inferred trajectory (Fig. 6e-g). Moreover, *Ucn3* expression recovered during the pharmacological treatment (Extended Data Fig. 10a). This shows

that the maturation state before treatment was different from that achieved after treatment. Hence, upon PEG–insulin or PEG–insulin plus GLP-1–oestrogen treatment,  $\beta$  cells adopt a molecular immature yet functional phenotype that is sufficient for blood glucose normalisation and diabetes remission.

$\beta$  cells from mice treated with PEG–insulin or the combination of PEG–insulin and GLP-1–oestrogen were grouped in distinct  $\beta$ -cell subpopulations, albeit at a similar maturation state (Fig. 6b). This implies the existence of a compound-specific mechanism-of-action (MOA) that underlies the recovery of  $\beta$ -cell function. To investigate the distinct MOAs of the different treatments, we identified the  $\beta$ -cell-specific transcriptional signature of treated and mSTZ-derived  $\beta$  cells (Supplementary Table 5). An increased expression of genes involved in  $\beta$ -cell functionality and maturation (Fig. 6h) was common to both treatments and may result from improved blood glucose levels and/or stimulation of shared pathways of insulin and GLP-1–oestrogen signalling. GLP-1, oestrogen and insulin receptor activation regulate MAPK and FoxO signalling<sup>38–40</sup>, both of which were increased after PEG–insulin, GLP-1–oestrogen and PEG–insulin and GLP-1–oestrogen co-therapy (Fig. 6h). Although we cannot dissect the signalling contribution of each individual receptor, we believe that the polypharmacological approach might potentiate the simultaneous activation of commonly regulated pathways.

Unexpectedly, treatment with PEG–insulin elicited a  $\beta$ -cell-specific stimulation of the insulin signalling cascade as well as stimulation of the recently characterised RNA polymerase II mediated pathway<sup>41</sup> (Fig. 6h). Hence, our data suggest that direct effects of insulin on  $\beta$  cells contribute to the improvement of  $\beta$ -cell function and recovery, as was proposed for T2D<sup>42,43</sup>.

Our goal was to use GLP-1–oestrogen to selectively deliver oestrogen into  $\beta$  cells. We consistently found  $\beta$ -cell-specific induction of the ER-associated degradation (ERAD) pathway and induction of transfer RNA signalling in PEG–insulin and GLP-1–oestrogen cotreated mice (Fig. 6h). ERAD mitigates ER stress, which, when unresolved, contributes to functional  $\beta$ -cell mass loss in T1D and T2D<sup>44</sup>. Chemical and genetic disturbances of ERAD impair  $\beta$ -cell function<sup>44</sup>. We found an increased proinsulin-to-C-peptide ratio in mSTZ-diabetic mice, which is used as an ER-stress surrogate in diabetes<sup>45</sup> (Extended Data Fig. 10b). GLP-1–oestrogen, PEG–insulin and GLP-1–oestrogen and PEG–insulin co-therapy normalised this ratio (Extended Data Fig. 10b). Recently, it has been reported that oestrogen, via nuclear oestrogen receptor alpha signalling, stabilizes the ERAD proteins Sel1l and Hrd1 in  $\beta$  cells, an effect associated with diabetes amelioration in Akita mice<sup>46</sup>. We observed increased co-staining for insulin and Sel1l in GLP-1–oestrogen and PEG–insulin cotreated islets only 25 d after treatment initiation (Extended Data Fig. 10c). In addition to upregulation of *Sel1l* and *Hrd1* (also known as *Syvn1*), ERAD-associated gene expression (for example,

expression of *Sdf2l1*, *Herpud1*, *Dnajb11*, *Dnajb9*, *Derl3* and *Hspa5* genes) was specifically increased in  $\beta$  cells of GLP-1–oestrogen and PEG–insulin cotreated mice (Extended Data Fig. 10d). The *Sdf2l1*, *Herpud1* and *Hspa5* genes encode ERAD-proteins that have beneficial effects on  $\beta$  cells that allow correct insulin folding and/or function<sup>47–49</sup>. *Dnajb9* and *Dnajb11* are chaperone proteins that might aid correct insulin protein folding. *Derl3* is required for ER-associated degradation<sup>50</sup>. The specific role of *Derl3* in  $\beta$  cells is unknown, but interestingly, *Derl3* expression was shown to protect cardiomyocytes against ER-stress-induced death by enhancing ERAD<sup>50</sup>. These results support a role for ERAD activation by GLP-1–oestrogen and PEG–insulin cotreatment that induces a treatment-specific molecular profile for the protection and regeneration of  $\beta$  cells (Fig. 6b).

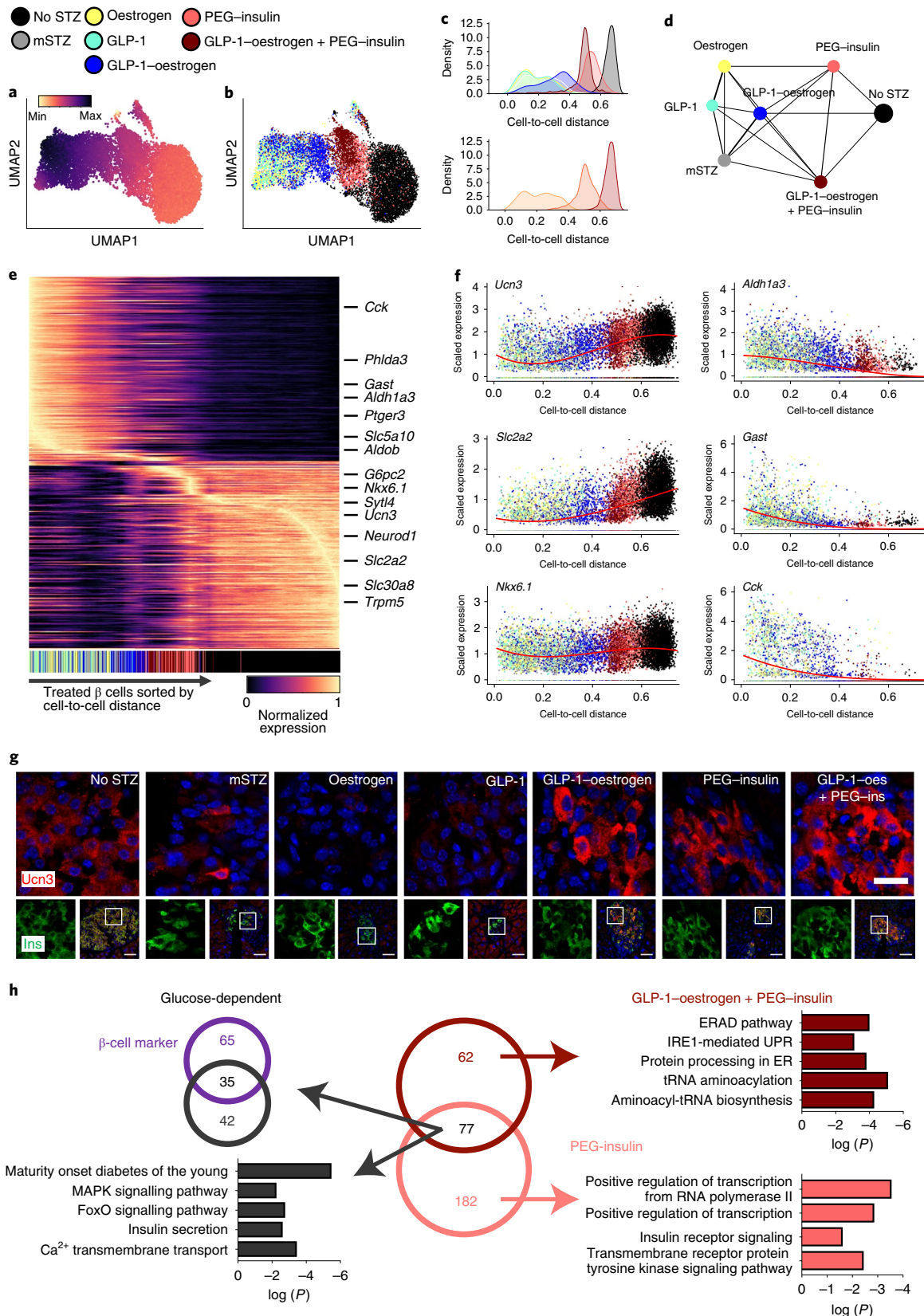
In a similar manner, tRNA signalling is known to be an intracellular target of oestrogen, and an increased abundance of tRNA has been associated with proliferating cells<sup>51,52</sup>. Indeed, from the single-cell data, we observed the highest fraction of proliferative  $\beta$  cells in the PEG–insulin and GLP-1–oestrogen conjugate cotreated mice (Fig. 7a). Moreover, we found increased  $\beta$ -cell proliferation in GLP-1–oestrogen and PEG–insulin cotreated mice that was already evident after 25 d of treatment, and that, importantly, was not evident in the single-treatment groups (Fig. 7b). Stressed  $\beta$  cells, such as those under chronic hyperglycaemic conditions, lack an adequate response to GLP-1 therapy probably as a result of decreased expression of GLP-1R<sup>53–55</sup>. Here, we reasoned that the restoration of glycaemia in mSTZ mice, notably through additional chronic insulin therapy, may increase the expression of GLP-1R. Indeed, we found progressively increased levels of GLP-1R on the surfaces of  $\beta$  cells of mice with improved glucose levels (Fig. 7c). This may have facilitated the enhanced delivery, uptake and action of oestrogen and GLP-1 in  $\beta$  cells, especially for those cells cotreated with GLP-1–oestrogen and insulin.

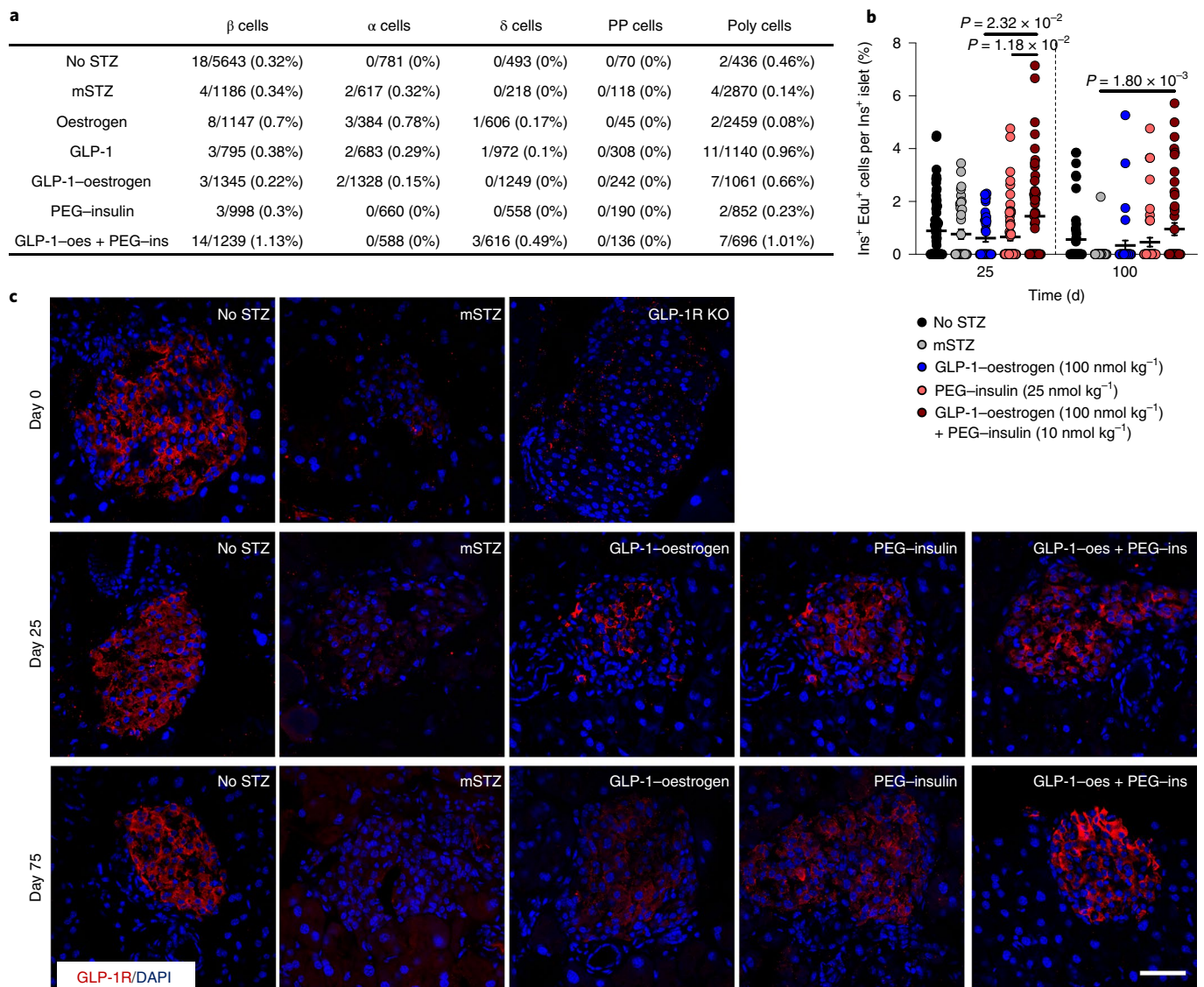
To examine whether other endocrine cells have contributed to the regeneration of functional  $\beta$  cells, we explored cluster relations and possible cellular transitions using PAGA and RNA velocity estimation (see Methods). For this we included  $\beta$ -mSTZ cells as the origin (starting point) of treated cells and we investigated where cells moved from that baseline (Fig. 8a). We found no direct connection or cell movement from other (non- $\beta$ ) cell populations towards redifferentiated  $\beta$  cells. We next examined the RNA velocity and potential fate of treated endocrine cells (Fig. 8b). The velocity of some of the immature  $\beta$  cells observed after GLP-1–oestrogen, PEG–insulin and combined treatment pointed towards mature  $\beta$  cells of healthy mice, thus further substantiating  $\beta$ -cell redifferentiation. Moreover, the scRNA-seq data suggested that neogenesis was not increased after 100 d of treatment, as the expression levels of *Ngn3* mRNA remained unchanged in endocrine cell subtypes (Fig. 8c). We also found no indication, from *Ngn3* immunostaining

**Fig. 6 |  $\beta$ -cell redifferentiation upon insulin and GLP-1–oestrogen treatment.** **a, b**, UMAP plot of  $\beta$  cells from all treatment groups. Colour indicates random-walk-based cell-to-cell distance (**a**) and treatment groups (**b**). **c**, Cell density of treatment groups and  $\beta$ -cell subpopulations along a cell trajectory from dedifferentiated cells to mature  $\beta$  cells that indicate the redifferentiation state. Cells are ordered according to a random-walk-based cell-to-cell distance as shown in **a**. **d**, Abstracted graph of transcriptomic similarity of  $\beta$  cells between treatment groups inferred on the basis of a measure for cluster connectivity using PAGA. Edge weight indicates link significance. **e**, Gene expression changes of the top 200 upregulated and downregulated genes in  $\beta$  cells of mSTZ-treated mice along the cell trajectory from a dedifferentiated to mature state as in **c** and **d**. Expression is shown as the running average scaled to the maximum observed level per gene. Genes are ordered by their maximum expression. The bar at the bottom shows the location of individual cells coloured by treatment group. **f**, Gene expression of selected  $\beta$ -cell maturity and dedifferentiation markers along the cell-to-cell distance. Dots show expression levels of individual cells coloured by treatment group. Superimposed red lines are polynomial regression fits. **g**, Immunohistochemical analysis of the  $\beta$ -cell maturity marker *Ucn3* at the end of the study. Images are representative of mice treated with no STZ-vehicle ( $n=3$ ), mSTZ-vehicle ( $n=3$ ), oestrogen ( $n=3$ ), GLP-1 ( $n=3$ ), GLP-1–oestrogen ( $n=2$ ) and PEG–insulin ( $n=2$ ), and of mice cotreated with GLP-1–oestrogen and PEG–insulin ( $n=3$ ). Scale bars, 50  $\mu\text{m}$ . Scale bar zoom-in, 20  $\mu\text{m}$ . **h**, Venn diagram shows the number of upregulated genes ( $\log(\text{fold change}) > 0.25$ ,  $\text{FDR} < 0.01$ ) specifically in  $\beta$  cells of PEG–insulin-treated mice and GLP-1–oestrogen plus PEG–insulin-cotreated mice compared to mSTZ-treated mice (left). Selected GO terms and KEGG pathways are depicted (all pathways are listed in Supplementary Table 5). We used *limma-trend* to find differentially expressed genes (see Methods). Gene enrichment was done with *EnrichR* using Fisher's exact test to identify regulated ontologies or pathways (see Methods). Cells of  $n=3$  mice per treatment were pooled.

in tissue sections of earlier time points, that overt neogenesis contributed to  $\beta$ -cell regeneration (Fig. 8d). Together, these results suggest that redifferentiation of  $\beta$  cells along dedifferentiation and redifferentiation trajectories is the main mechanism that

underlies the re-establishment of functional  $\beta$  cells in the mSTZ model by treatment with GLP-1-oestrogen or PEG-insulin, or by their cotreatment. By using a combination of low dose insulin with GLP-1-oestrogen treatment, we were able to trigger a





**Fig. 7 | Treatment specific effects of  $\beta$ -cell regeneration. a, b.** Contribution of  $\beta$ -cell proliferation to  $\beta$ -cell regeneration after drug treatment of mSTZ mice. **a**, ScRNA-seq of endocrine cells after 100 d of treatment suggests an increased proliferation of specifically  $\beta$  cells after GLP-1–oestrogen and PEG–insulin cotreatment. **b**, Increased rate of proliferating  $\beta$  cells as indicated by EdU<sup>+</sup>  $\beta$  cells per islet in mice treated with no STZ, mSTZ, GLP-1–oestrogen and PEG–insulin, and GLP-1–oestrogen plus PEG–insulin-cotreated mice after 25 d and 100 d of treatment. Day 25: No STZ, 73 islets of  $n=3$  mice; mSTZ, 36 islets,  $n=3$ ; GLP-1–oestrogen, 37 islets,  $n=3$ ; PEG–insulin, 61 islets,  $n=3$ ; GLP-1–oestrogen and PEG–insulin, 50 islets,  $n=3$ . Day 100: No STZ, 47 islets of  $n=3$  mice; mSTZ, 47 islets,  $n=3$ ; GLP-1–oestrogen, 36 islets,  $n=2$ ; PEG–insulin, 30 islets,  $n=2$ ; GLP-1–oestrogen and PEG–insulin, 47 islets,  $n=3$ . Data are mean  $\pm$  s.e.m. One-way ANOVA followed by Tukey post-hoc comparison at day 25 or day 100 ( $F_{D25}(4, 247) = 3.413$ ;  $F_{D100}(4, 198) = 3.814$ ). **c**, Immunohistochemical analysis of GLP-1R expression. A GLP-1R knockout (GLP-1R KO) mouse was used to show specificity of the GLP-1R antibody within the islets of Langerhans. Scale bar, 50  $\mu$ m. Day 0: Images are representative of mice treated with no STZ-vehicle ( $n=3$ ) and mSTZ-vehicle ( $n=3$ ). Day 25: Images are representative of mice treated with no STZ-vehicle ( $n=3$ ), mSTZ-vehicle ( $n=3$ ), PEG–insulin ( $n=3$ ) and GLP-1–oestrogen ( $n=3$ ), and of mice cotreated with GLP-1–oestrogen and PEG–insulin ( $n=3$ ). Day 75: Images are representative of mice treated with no STZ-vehicle ( $n=2$ ), mSTZ-vehicle ( $n=2$ ), PEG–insulin ( $n=1$ ) and GLP-1–oestrogen ( $n=3$ ), and of mice cotreated with GLP-1–oestrogen and PEG–insulin ( $n=3$ ). Additional information is available in Source data.

$\beta$ -cell-specific transcriptional response that was characterised by increased  $\beta$ -cell proliferation and enhanced functionality.

## Discussion

Herein we have established the mSTZ model of diabetes as a model to study  $\beta$ -cell dysfunction and dedifferentiation. Single-cell profiling of remaining  $\beta$  cells identified many markers of  $\beta$ -cell dedifferentiation that were undescribed previously and that code for surface molecules, receptors and secreted proteins. These may be used as

biomarkers or may allow the detection, isolation and characterisation of dedifferentiated  $\beta$  cells. This could reveal pathomechanisms of T1D and T2D and may have the potential to identify unique diagnostic markers and therapeutic targets. Using scRNA-seq we were able to delineate a  $\beta$ -cell fate trajectory in which cells transitioned from mature to immature to dedifferentiated  $\beta$  cells, which implies that  $\beta$  cells can be characterised by a continuum of transcriptional states that reflect discrete phenotypes. Inference of cell transitions using the RNA velocity concept further suggested that there was

no ongoing transdifferentiation from other non- $\beta$  and non-endocrine cells towards dedifferentiated  $\beta$  cells. Upregulation of the endocrine master regulator Ngn3 might depend on the severity of hyperglycaemia; high glucose levels ( $>33$  mM) have been shown to induce Ngn3 expression<sup>9,15</sup>, whereas lower levels ( $<25$  mM) had no effect<sup>14,56</sup>. We show that  $\beta$ -cell dedifferentiation in mSTZ-diabetic mice is independent of induction of Ngn3<sup>+</sup> endocrine progenitors and that the transcriptional state of dedifferentiated  $\beta$  cells is more similar to late embryonic or early postnatal  $\beta$  cells.

Recently, the Kushner laboratory has provided evidence that some of the remaining insulin in the blood stream of patients with long-term T1D<sup>57</sup> originates from dedifferentiated  $\beta$  cells and/or from polyhormonal non- $\beta$  cells that function as 'insulin microsecretors'<sup>58</sup>. Similarly, the Korsgren laboratory found histological evidence for  $\beta$ -cell dedifferentiation at T1D onset<sup>59</sup>. Therefore, triggering the redifferentiation of dedifferentiated  $\beta$  cells seems to be an intuitive approach for the treatment of diabetes that does not involve  $\beta$ -cell proliferation or neogenesis per se<sup>60</sup>. Preclinical as well as clinical findings from patients with type 1 and 2 diabetes suggest that a transient recovery of  $\beta$ -cell dysfunction occurs upon glycaemia normalisation by intensive insulin treatment by either  $\beta$ -cell rest or redifferentiation<sup>3,15</sup>. By using scRNA-seq we can dissect endocrine subtype-specific treatment responses and show that insulin treatment triggers transcriptional changes in  $\beta$  cells that are connected to insulin and/or IRS signalling. This supports the idea that in addition to lowering the glucotoxic stress on  $\beta$  cells, direct insulin- or IGF-signalling improves  $\beta$ -cell health and performance and can redifferentiate  $\beta$ -cell mass in diabetic models<sup>13</sup>. Importantly, the redifferentiated  $\beta$  cells that were induced by insulin therapy were functional and responded to physiological stimuli, as indicated by increased fasting plasma C-peptide levels.

Moreover, we show that targeted delivery of oestrogen using GLP-1 as a peptide carrier and intensive insulin co-therapy by a distinct MOA alleviates hyperglycaemia, increases fasting C-peptide levels and redifferentiates  $\beta$  cells, whereas it reduces daily insulin requirements by 60% and limits weight gain in mice. The enhanced restoration of GLP-1R expression in dedifferentiated  $\beta$  cells by the GLP-1-oestrogen and insulin cotreatment renders them susceptible to targeted delivery of oestrogen. As has been proposed previously in the Akita mouse model<sup>16</sup>, we observed that stimulating the ERAD pathway by GLP-1-oestrogen beneficially influences  $\beta$ -cell physiology in rodent models of diabetes. In future studies, it might be of specific interest to test GLP-1-oestrogen with and without PEG-insulin co-therapy in genetically perturbed mouse models of ERAD. Finan et al. showed previously that peptide-based targeting prevented adverse side effects of oestrogen, such as uterus and tumour growth<sup>19</sup>. There was also no measurable oestrogen-induced increase in bone content owing to the effect of the absence or limited expression of GLP-1R on off-target tissues and cells<sup>19</sup>. Here, we extended the safety profile of GLP-1-oestrogen and demonstrated that GLP-1-oestrogen did not stimulate uterine tissue growth in OVX rats. This study further verified that there is insufficient free,

systemic oestrogen to drive toxicity as well as the selectivity and specificity of GLP-1-mediated oestrogen targeting. The strategy to use GLP-1 as a carrier may be adopted to selectively target any other small molecule or biologic to  $\beta$  cells. The prerequisite for the transport of the molecule of interest into the target cell (that is, stressed and dedifferentiated  $\beta$  cells) is adequate GLP-1R expression. Under hyperglycaemic conditions, adjunctive treatments that reduce the glycaemic burden, such as chronic insulin therapy as demonstrated here, can facilitate the restoration of GLP-1R expression in stressed  $\beta$  cells. Notably, chronic PEG-insulin treatment also increased functional  $\beta$ -cell number and the scRNA-seq data suggest a direct effect of insulin on  $\beta$  cells. Thus, combinatorial pharmacological treatments that include insulin might have additional beneficial effects on  $\beta$ -cell survival, protection, proliferation and function. In summary, our work has identified mechanisms and pathways of  $\beta$ -cell dedifferentiation and opens avenues for pharmacological targeting of these dedifferentiated cells for diabetes remission.

## Methods

**mSTZ treatment.** mSTZ (Sigma-Aldrich cat. no. S0130) was injected intraperitoneally in 8-week-old male C57BLJ/6 mice ( $n = 125$ ) at 50 mg kg<sup>-1</sup> for five consecutive days following the mSTZ model to induce diabetes. A subset of age-matched male mice was injected with ice-cold citrate buffer (pH 4.5) as control animals ( $n = 20$ ). C57BLJ/6 mice were obtained from Janvier Labs. Ten days after the last mSTZ injection, fasting blood glucose was taken, as well as fasting plasma, to determine fasting insulin and C-peptide levels. We included hyperglycaemic mice with fasting blood glucose levels of  $>190$  mg dl<sup>-1</sup> ( $n = 116$ ). We estimated  $\beta$ -cell function and mass of mSTZ-treated mice by combining fasting blood glucose levels, the homeostatic model assessment (HOMA)- $\beta$ -score and the ratio of fasting C-peptide to blood glucose levels. Among mSTZ-treated mice, animals with fasting glucose levels of  $>25$ th percentile, a HOMA- $\beta$ -score of  $<25$ th percentile and a C-peptide to blood glucose ratio of  $<25$ th percentile were excluded from the study ( $n = 9$ ).

**Pharmacological study in mSTZ mice.** mSTZ diabetic mice were randomised and evenly distributed to different treatments according to fasting blood glucose levels. Ten days after the last STZ injection, mice were allocated to different treatments of daily subcutaneous injection with vehicle (PBS;  $n = 17$ , not mSTZ-treated), vehicle (PBS;  $n = 17$ , mSTZ-treated), a GLP-1 analogue ( $n = 16$ ), oestrogen ( $n = 14$ ), GLP-1-oestrogen ( $n = 28$ , of which  $n = 11$  mice were switched to vehicle (PBS) treatment after 12 weeks of GLP-1-oestrogen treatment), PEG-insulin ( $n = 13$ ), or GLP-1-oestrogen and PEG-insulin ( $n = 16$ ) at the doses indicated in Fig. 1a for 100 d. Mice were housed up to four per cage on a 12:12-h light-dark cycle at 22 °C and given free access to a normal chow diet (Altromin 1314) and water. Compounds were administered in a vehicle of PBS (Gibco) and were given by daily subcutaneous injections at the indicated doses at a volume of 5  $\mu$ l g<sup>-1</sup> body weight for the indicated durations as indicated in the figure legends. The investigators were not blinded to group allocation during the in vivo experiments or to the assessment of longitudinal endpoints. All rodent studies were approved by and performed according to the guidelines of the Animal Use and Care Committee of Bavaria, Germany.

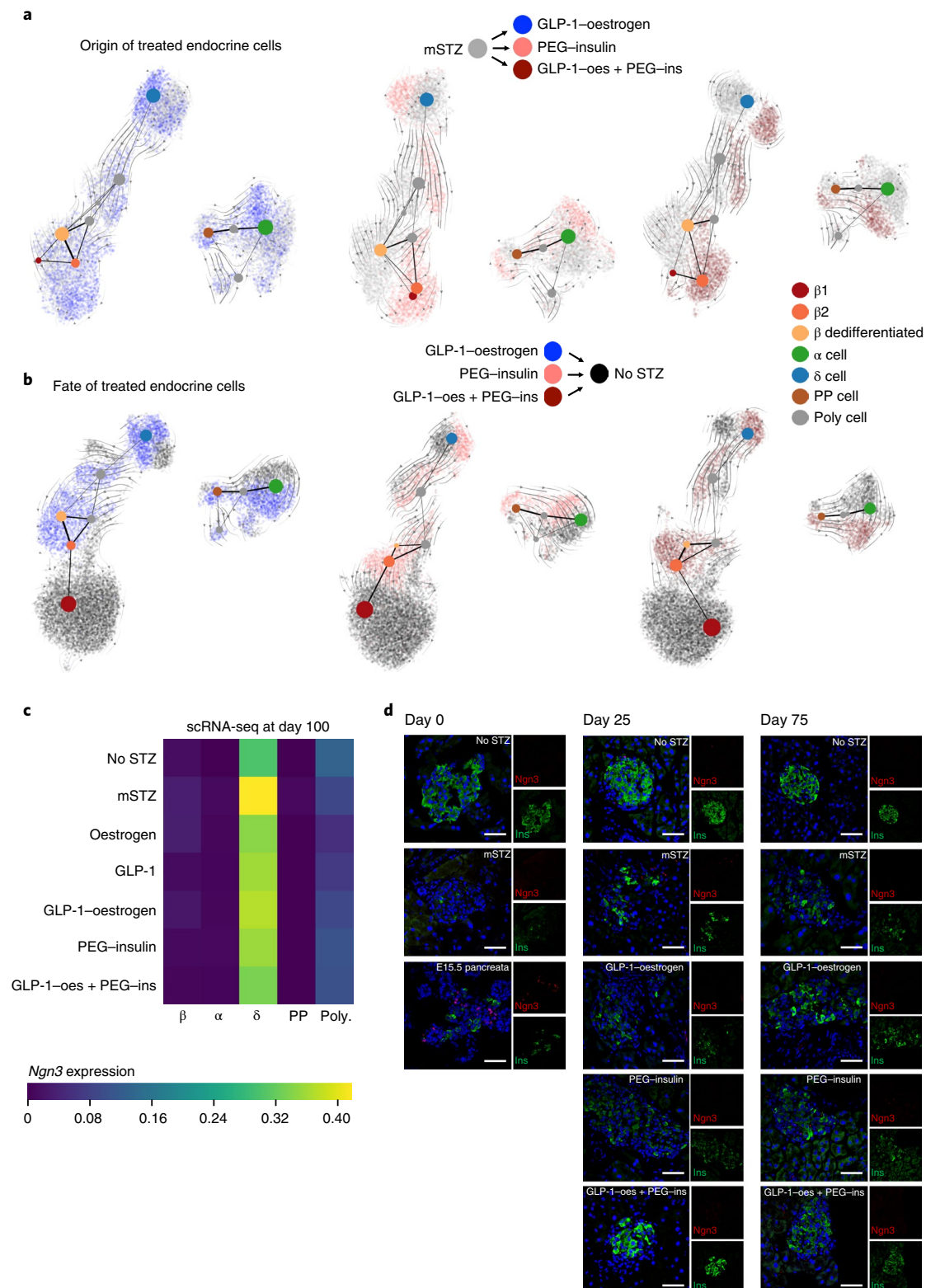
**Study in FVFPBF<sup>D<sup>fl</sup>om</sup> mice.** Male 8-week old FVFPBF<sup>D<sup>fl</sup>om</sup> mice with fasting blood glucose  $>250$  mg dl<sup>-1</sup> were randomised to vehicle ( $n = 7$ ), oestrogen ( $n = 5$ ), GLP-1 ( $n = 9$ ) or GLP-1-oestrogen ( $n = 11$ ) treatment according to their fasting blood glucose levels. FVFPBF<sup>D<sup>fl</sup>om</sup> mice were obtained from breeding. Mice were treated daily for four weeks with subcutaneous injections. Fasting blood glucose was measured after a 6-h fast. We single- or group-housed the mice on a 12-h light,

**Fig. 8 | Origin and fate of treated endocrine cells.** **a, b**, Cluster relationships and cell transitions to indicate the origin and fate of treated endocrine cells. Graphs of lineage relationships are derived from cluster connectivity using PAGA. Paths in the graph signify potential lineage transitions and are weighted by significance. Cell transitions are inferred from estimated RNA velocities and the direction of movement plotted as streamlines on the UMAP. **a**, Plots show endocrine cells from mice treated with mSTZ and GLP-1-oestrogen (left two), PEG-insulin (middle two) or GLP-1-oestrogen plus PEG-insulin (right two), showing movement from mSTZ (origin or starting point) towards treated cells. **b**, Plots show endocrine cells from healthy mice and and those treated with GLP-1-oestrogen (left two), PEG-insulin (middle two) or GLP-1-oestrogen plus PEG-insulin (right two), showing a potential movement of the treated cells towards healthy cells (fate). **c, d**, Expression of the endocrine progenitor marker Ngn3 to assess the contribution of  $\beta$ -cell neogenesis to  $\beta$ -cell regeneration by GLP-1-oestrogen, PEG-insulin and GLP-1-oestrogen plus PEG-insulin treatment. **c**, ScRNA-seq shows no increase in *Ngn3* gene expression in endocrine cells after 100 d of treatment. **d**, Immunohistochemistry of Ngn3 expression during the course of the study. Mouse E15.5 pancreas was used as a positive control and shows nuclear staining for Ngn3. Scale bar, 50  $\mu$ m. Day 0: Images are representative of mice treated with no STZ-vehicle ( $n = 3$ ) and mSTZ-vehicle ( $n = 3$ ). Day 25: Images are representative of mice treated with no STZ-vehicle ( $n = 3$ ), mSTZ-vehicle ( $n = 3$ ), PEG-insulin ( $n = 3$ ) and GLP-1-oestrogen ( $n = 3$ ), and of mice cotreated with GLP-1-oestrogen and PEG-insulin ( $n = 3$ ). Day 75: Images are representative of mice treated with no STZ-vehicle ( $n = 2$ ), mSTZ-vehicle ( $n = 2$ ), PEG-insulin ( $n = 1$ ) and GLP-1-oestrogen ( $n = 3$ ), and of mice cotreated with GLP-1-oestrogen and PEG-insulin ( $n = 3$ ).

12-h dark cycle at 22 °C and provided free access to food and water. This study was approved by and performed according to the guidelines of the Animal Use and Care Committee of Bavaria, Germany.

**Uterotrophic assessment in OVX rats.** The study was designed in accordance with the Endocrine Disruptor Screening Programme Test Guidelines OPPTS 890.1600: Uterotrophic Assay, a standardised *in vivo* screening test intended to evaluate the ability of a chemical to elicit biological activities consistent with agonists of natural oestrogens (for example, 17 $\beta$ -oestradiol). The test measures the increase

in uterine weight or uterotrophic response in comparison with non-treated controls<sup>61</sup>. The study was further designed according to accepted pharmacological principles and followed Good Laboratory Practice (conducted by Envigo). A total of 44 OVX female Sprague-Dawley rats (Charles River) were supplied for the study, of which 40 animals were allocated to treatment groups (randomised by body weight to ensure equal group mean starting body weight) and the remaining 4 animals were allocated as spares. On the day of dosing (following 14–22 d of acclimatisation), rats were approximately 9–12 weeks of age, and weighed 217 g to 348 g. Four groups of 8 rats each were treated for 14 consecutive days with once-





daily subcutaneous administration of GLP-1-oestrogen at doses of 400  $\mu\text{g kg}^{-1}$ , 2,200  $\mu\text{g kg}^{-1}$  and 4,000  $\mu\text{g kg}^{-1}$  per day or with volume-matched vehicle (PBS). An additional group of animals received once-daily subcutaneous injections of vehicle for the first 10 d followed by once-daily subcutaneous injection of 17  $\alpha$ -ethynyl oestradiol (positive control) at 300  $\mu\text{g kg}^{-1} \text{d}^{-1}$  on days 11 to 14. Animals were weighed daily from day 1 (prior to dosing) until the day of necropsy, and food consumption was recorded on days -4, 1, 4, 8, 11 and 15 (day of necropsy). Standard toxicological organ weight measurement was carried out at necropsy, including the weighing of wet and dry (blotted) uterine tissue. The uterus was sampled and weighed according to OPPTS 890.1600.

**Administration of EdU.** To investigate cell proliferation, we used the modified uracil analogue 5'-ethynyl-2'-desoxyuridine (EdU). EdU was injected intraperitoneal (i.p.) at 50  $\mu\text{g kg body weight}^{-1}$  24 h prior to sacrifice.

**Compound formulation and peptide synthesis and cleavage.** *GLP-1 and GLP-1-oestrogen.* Peptides were synthesised by solid-phase peptide synthesis methods using *in situ* neutralisation for fluorenylmethoxycarbonyl (Fmoc)-based chemistries. For Fmoc-based neutralisation peptide synthesis, we used 0.1 mmol Rink MBHA resin (0.5–0.7 mmol  $\text{g}^{-1}$ , 100–200 mesh, Novabiochem) on a Symphony peptide synthesiser by standard Fmoc methods using diisopropylcarbodiimide/1-hydroxybenzotriazole chloride (DIC/HOBt-Cl, both from AAPPTec) in dichloromethane (DMF, VWR Chemicals) for coupling and 20% piperidine in DMF for deprotection of N-terminal amines. Completed peptidyl resins were treated with trifluoroacetic acid (TFA, VWR Chemicals), triisopropylsilane (TIS, Sigma-Aldrich) and water (9.0:0.5:0.5 (v/v/v)) for 2 h with agitation. We confirmed the peptide molecular weights and character by liquid chromatography-mass spectrometry on an Agilent 1260 Infinity/6120 Quadrupole instrument with a Kinetex C8 column and a gradient of 10%–80% eluent B. Eluent A is water with 0.05% TFA and eluent B is 10% water, 90% MeCN, 0.05% TFA. Preparative HPLC purifications were performed on a Waters instrument with a Kinetex 5  $\mu\text{m}$  C8.

To synthesise the derivatized oestrogen for construction of the conjugate, we reacted oestradiol 17-acetate (Sigma) and a tenfold excess of ethyl 2-bromoacetate in dioxane in the presence of  $\text{K}_2\text{CO}_3$  under reflux conditions and agitation. After the removal of dioxane *in vacuo*, we resuspended the intermediate product in dioxane with 1 N NaOH with heat, followed by a re-acidification by HCl in dichloromethane. After the removal of DCM *in vacuo*, the crude product was resuspended in aqueous solvent containing at least 20% MeOH, 20% acetonitrile (ACN, VWR Chemicals) and 1% AcOH. The crude extract was purified by reversed-phase HPLC.

The derivatized oestrogen was added covalently to the side-chain amine on the C-terminal lysine amide residue. We used a C-terminal N-methyltrityl-L-lysine (Lys(Mtt)-OH) residue, whose side chain was orthogonally deprotected with four 10-min treatments with 1.5% TFA, 2% TIS and 1% anisole in DMF. The oestrogen was attached through this side-chain amine of the C-terminal lysine after treatment with a threefold excess of the purified derivatized oestrogen and DIC/HOBt-Cl in N-methyl-2-pyrrolidone. After TFA cleavage as described above, the crude extract was resuspended in aqueous buffer containing 20% ACN and 0.1 M  $\text{NH}_4\text{HCO}_3$ . The conjugate was purified as described above.

**PEG-insulin.** Pegylated insulin (PEG-insulin) was prepared by insulin N-terminal amine reductive amination with 20 K methoxy PEG propionaldehyde. In brief, human insulin was dissolved in 50 mM sodium acetate buffer (pH 5.0) and 50% ACN. A 30-fold excess of sodium cyanoborohydride and a 1.5-fold excess of methoxy PEG propionaldehyde (M-ALD-20K, JenKem Technology) was added to the buffer containing insulin for 3 h at room temperature with stirring. Purification by reverse phase chromatography on a C-8 column in 0.1% TFA acetonitrile solvents yielded the final product at greater than 95% purity.

Oestrogen (17 $\beta$ -oestradiol, Sigma-Aldrich) was dissolved in 100% ethanol (Sigma-Aldrich) at a concentration of 1 mg  $\text{ml}^{-1}$  and was diluted with PBS to the required concentration.

**Blood parameters.** Blood was collected from tail veins after a 4-h fast, using EDTA-coated microvette tubes (Sarstedt). Blood was immediately chilled on ice. Plasma was separated by centrifugation at 5,000g at 4°C for 10 min using a micro centrifuge and was stored at -20°C until further usage. Plasma insulin and C-peptide (Crystal Chem) and Proinsulin (Alpco) were quantified by ELISA (enzyme-linked immunosorbent assay) following the manufacturer's instructions. Four-hour fasting blood glucose levels were determined using a handheld glucometer (FreeStyle).

**Pancreas dissection.** Adult pancreata were dissected and fixed in 4% PFA in PBS for 24 h at 4°C. The tissues were cryoprotected in a sequential gradient of 7.5%, 15% and 30% sucrose-PBS solutions at room temperature (2 h incubation for each solution). Next, the pancreata were incubated in 30% sucrose and tissue embedding medium (Leica) (1:1) at 4°C overnight. Afterwards, they were embedded in a cryoblock using tissue-freezing medium (Leica), frozen in dry ice and stored at -80°C. Sections of 20  $\mu\text{m}$  thickness were cut from each sample, mounted on a glass slide (Thermo Fisher Scientific) and dried for 10 min at room temperature before use or storage at -20°C.

**Immunostaining of sections.** The cryosections were rehydrated by three washes with 1 $\times$  PBS, and were permeabilised with 0.2–0.15% Triton X-100 in  $\text{H}_2\text{O}$  for 30 min. Permeabilisation was not performed for staining with GLP-1R. Then, the samples were blocked in blocking solution (PBS, 0.1% Tween-20, 1% donkey serum, 5% FCS) for 1 h. The following primary antibodies were used: guinea pig polyclonal anti-insulin (1:300, Thermo Fisher Scientific), goat polyclonal anti-Glut2 (1:500, Abcam), goat polyclonal anti-Nkx6.1 (1:200, R&D Systems), goat polyclonal anti-somatostatin (1:500, Santa Cruz), rat monoclonal anti-somatostatin (1:300, Invitrogen), rabbit polyclonal anti-urocortin 3 (1:300, Phoenix Pharmaceuticals), rabbit monoclonal anti-insulin (1:300, Cell Signaling), guinea pig polyclonal anti-glucagon (1:500, Takara), guinea pig polyclonal anti-insulin (1:300, ABD Serotec), rabbit polyclonal cleaved caspase-3 (Asp 175) (1:300, Cell Signaling), rabbit monoclonal anti-ki67 (1:300, Abcam), rabbit polyclonal anti-Aldh1a3 (1:300, Abcam), rabbit monoclonal anti-GLP-1R (1  $\mu\text{g ml}^{-1}$ , Novo Nordisk), rabbit polyclonal anti-gastrin (1:100, Abcam), rabbit polyclonal anti-cholecystokinin (1:100, ENZO), goat polyclonal anti-Sel11 (1:300, Novus Biologicals) and rabbit anti-Ngn3 (1:800, donated by H. Edlund). Dilutions were prepared in blocking solution and sections were incubated overnight at 4°C. Thereafter, sections were rinsed 3 $\times$  and washed 3 $\times$  with 1 $\times$  PBS. All secondary antibodies were used at a 1:800 dilution prepared in blocking buffer. We used the following secondary antibodies: donkey anti-goat IgG (H+L) secondary antibody (Alexa Fluor 633, Invitrogen A-2108), donkey anti-rabbit IgG (H+L) secondary antibody (Alexa Fluor 555, Invitrogen A-31572); donkey anti-rabbit IgG (H+L) secondary antibody (Alexa Fluor 488, Invitrogen A-21206); donkey anti-guinea pig IgG (H+L) secondary antibody (DyLight 649, Dianova, 706–495–148); donkey anti-rat IgG (H+L) secondary antibody (DyLight 647 Dianova 711–605–152); donkey anti-rat IgG (H+L) secondary antibody (Cy3, Dianova 712–165–153); donkey anti-guinea pig (H+L) secondary antibody (Alexa Fluor 488, Dianova 706–545–148). After 4–5 h of incubation, pancreatic sections were stained with DAPI (1:500 in 1 $\times$  PBS) for 30 min, rinsed and washed 3 $\times$  with 1 $\times$  PBS, and subsequently mounted. All images were obtained with a Leica microscope of the type DMI 6000 using LAS AF software (Leica). Images were analysed using the LAS AF and/or ImageJ software programme.

**Automatic tissue analysis.** Stained tissue sections were scanned with an AxioScan.Z1 digital slide scanner (Zeiss) equipped with a  $\times 20$  magnification objective. We scanned three sections per animal. Images were evaluated using the commercially available image analysis software Definiens Developer XD 2 (Definiens). First, regions of interest were annotated manually to select islets of Langerhans for analysis. A specific rule set was then defined to detect and quantify the cells within each defined region on the basis of the fluorescence intensity of DAPI, morphology, size and neighbourhood. The Ins-, Gcg- or Sst-expressing cells were classified automatically using the fluorescence intensity of each hormone.

**EdU detection protocol.** EdU staining was carried out according to the EdU imaging kit manual (Life Technologies) after staining with the secondary antibody. DAPI staining and mounting was performed as described above.

**Pancreatic insulin content.** Pancreatic insulin content was determined by an acid ethanol extraction. The pancreas was dissected, washed in 1 $\times$  PBS and homogenised in an acid-ethanol solution (5 ml 1.5% HCl in 70% ethanol) followed by incubation at -20°C for 24 h. After 2 rounds of acid-ethanol precipitation, the tissue was centrifuged (2000 r.p.m., 15 min, 4°C) and the supernatant was neutralised with 1 M Tris pH 7.5. Insulin was measured using a mouse insulin ELISA (Crystal Chem) and was normalised over the protein concentration that was determined by BCA protein assay.

**Islet isolation.** Islet isolation was performed by collagenase P (Roche) digestion of the adult pancreas. In brief, 3 ml of collagenase P (1 mg  $\text{ml}^{-1}$ ) was injected into the bile duct and the perfused pancreas was consequently dissected and placed into another 3 ml of collagenase P for 15 min at 37°C. Then, 10 ml of G-solution (HBSS (Lonza) + 1% BSA (Sigma-Aldrich)) was added to the samples, and this was followed by centrifugation at 1,600 r.p.m. at 4°C. After another washing step with G-solution, the pellets were re-suspended in 5.5 ml of gradient preparation (5 ml 10% RPMI (Lonza) + 3 ml of 40% Optiprep (Sigma-Aldrich) per sample), and placed on top of 2.5 ml of the same solution. To form a 3-layer gradient, 6 ml of G-solution was added on the top. Samples were then incubated for 10 min at room temperature before centrifugation at 1,700 r.p.m. Finally, the interphase between the upper and the middle layers of the gradient was harvested and was filtered through a 70  $\mu\text{m}$  Nylon filter then washed with G-solution. Islets were hand picked under the microscope.

**Single-cell suspension.** To achieve a single-cell suspension of islets, islets were hand picked into a 1.5 ml Eppendorf tube, pelleted (800 r.p.m., 1 min) washed with PBS (minus Mg or Ca, Gibco) and digested with 0.25% trypsin with EDTA (Gibco) at 37°C for 8 min. Mechanical disaggregation every 2–3 min was required. The digestive reaction was then stopped and cells were pelleted (1200 r.p.m., 5 min).

**Single-cell sequencing.** Single-cell libraries were generated using the Chromium Single-cell 3' library and gel bead kit v2 (PN 120237) from 10x Genomics. In brief, to reach a target cell number of 10,000 cells per sample 16,000 cells per sample

were loaded onto a channel of the 10x chip to produce Gel Bead-in-Emulsions (GEMs). This underwent reverse transcription to barcode RNA before cleanup and cDNA amplification followed by enzymatic fragmentation and 5' adaptor and sample index attachment. Libraries were sequenced on the HiSeq4000 (Illumina) with 150 bp paired-end sequencing of read2.

**FACS sorting.** FACS sorting of endocrine cells was performed using the FACS-Aria III (BD Biosciences). Single cells were gated according to their FSC-A (front scatter area) and SSC-A (side scatter area). Singlets were gated dependent on the FSC-W (front scatter width) and FSC-H (front scatter height) and dead cells were excluded using the marker 7AAD (eBioscience). The FVF endocrine populations were discriminated according to their Venus fluorescence emission at 488 nm and the  $\beta$ - and  $\alpha$ -lineages were discriminated according to their BFP emission at 405 nm (positive and negative respectively). To enrich for  $\beta$  cells the distinct SSC-A high populations were gated. FACS sorted cells were sorted directly into Qiazol (Qiagen) for RNA isolation.

**RNA isolation and cDNA preparation.** The mRNA isolation was performed using the miRNeasy Micro Kit (Qiagen) according to the manufacturer's instructions. On-column DNase I treatment was applied to degrade DNA. For cDNA preparation the SuperScript Vilo cDNA Synthesis Kit (Thermo Fisher Scientific) and the GoScript Reverse Transcript System Kit (Promega) were used. The cDNA synthesis was carried out according to both manufacturers' instructions.

**Quantitative PCR.** Quantitative PCR (qPCR) was carried out using the Vii7 Real Time PCR System (Thermo Fisher Scientific) and the following TaqMan probes (Life Technologies): *Ins 1* (Mm01950294\_s1), *Ins 2* (Mm00731595\_Gh), *Gcg* (Mm01269055\_m1), *Sst* (Mm00436671\_m1), *Ppy* (Mm01250509\_g1), *Ghrelin* (Mm00445450\_m1), *Pecam1* (Mm01242584\_m1), *Gapdh* (Mm99999915\_g1) and *18S* (Mm03928990\_g1). Each reaction contained 25 ng of cDNA. For analysis, the  $C_t$ -values were transformed to the linear expression values and were normalised to the reference genes (*Gapdh* and *18S*) and to the control samples.

**Reaggregated human micro-islets.** All primary human islets were obtained through Prodo Laboratories with no information on the identity of the donor for ethical and privacy reasons (donor 1: male, BMI 32.38, age 48, HbA1c 5.6%; donor 2: male, BMI 33.2, age 46, HbA1c 5.4%; donor 3: male, BMI 28.65, age 34, HbA1c 5.2%). For all donors, consent was obtained from next of kin. For each preparation of InSphero 3D InSight human islet microtissues, 10,000–20,000 islet equivalents were dispersed in dissociation solution (1 $\times$  TrypLE Express solution (Thermo Fisher Scientific 12604013) with 40  $\mu\text{g ml}^{-1}$  DNase I (Sigma-Aldrich 10104159001)) by gentle pipetting at 37°C. Remaining cell clumps were removed by filtering the cell suspension through a cell strainer (70  $\mu\text{m}$  pore size). Islet microtissues were produced by hanging-drop-based scaffold-free reaggregation of 2,500 cells in each well of the InSphero 96-well Hanging Drop System for 5 d. The primary aggregates were then transferred to the Akura 96 well-plate to further mature for at least another 8 d before the start of the experiments. All experiments were performed within 30 d after the start of the aggregation. Islet microtissues were maintained in 3D InSight Human Islet Maintenance Medium (InSphero).

**Compound, cytokine treatments and GSIS with human micro-islets.** Dilution series of compounds were performed in 3D InSight Human Islet Maintenance Medium. Each of the assessed compounds was added to the culture medium at the concentration indicated below, one day prior to the start of the cytokine treatment. The cytokine cocktail contained: tumour necrosis factor alpha (TNF $\alpha$ , 10 ng  $\text{ml}^{-1}$ , Thermo Fisher Scientific PHC3016), interferon gamma (IFN $\gamma$ , 10 ng  $\text{ml}^{-1}$ , Sigma-Aldrich I3265) and interleukin-1beta (IL-1 $\beta$ , 2 ng  $\text{ml}^{-1}$ , Sigma-Aldrich I17001), was prepared in PBS containing 0.1% BSA (Sigma-Aldrich A7888). The same concentration of PBS-BSA solution was maintained in each experimental condition. Regular redosings with cytokines and compounds were performed every 2–3 d. Prior to GSIS, culture medium was removed and islet microtissues were washed twice with Krebs Ringer Hepes Buffer (KRHB; 131 mM NaCl, 4.8 mM KCl, 1.3 mM CaCl $_2$ , 25 mM Hepes, 1.2 mM KH $_2$ PO $_4$ , 1.2 mM MgSO $_4$ , 0.5% BSA) containing 2.8 mM glucose and equilibrated for 1 h in the same solution. GSIS was performed in KRHB containing indicated glucose concentrations for 2 h. Following GSIS, the tissues were lysed using the CellTiter-Glo Luminescent Cell Viability Assay (Promega G9241) with protease inhibitor cocktail (Promega G6521) and the luminescence was recorded with a microplate reader (Infinite M1000, TECAN) for the analysis of total ATP content. The lysates were then used for assessment of total insulin content. After the correct dilutions in KRHB were performed, total and secreted insulin was quantified using the Stellux Chemi Human Insulin ELISA (Alpco, 80-INSHU-CH10). The Caspase-Glo 3/7 Assay (Promega, G8090) was used to assess caspase-3 and caspase-7 activity in the islet microtissues following compound treatment.

**Statistical analysis not including scRNA-seq data.** Preliminary data processing and calculations during ongoing studies were carried out using Microsoft Excel 2016. All further statistical analyses were performed using GraphPad Prism 8. We used the one-way analysis of variances (ANOVA) followed by Tukey's post hoc analysis to determine significance among the different treatment groups. In the case of only

two groups, the unpaired Student two-tailed *t*-test was used to detect significant differences. The human micro-islets were derived from three different donors, each of whom naturally varied in their GSIS. To compare the treatment effects among all donors, we used one-way ANOVA with the different donors as random effect followed by Tukey's post hoc analysis. This analysis was performed in R. Grubbs' test ( $\alpha < 0.05$ ) was used to detect significant outliers, which were then excluded from subsequent statistical analysis and figure drawing.  $P < 0.05$  was considered to be statistically significant. All results are mean  $\pm$  s.e.m. unless otherwise indicated.

**Preprocessing of droplet-based scRNA-seq data.** Demultiplexing of binary base call (BCL) files, alignment, read filtering, barcode and unique molecular identifier (UMI) counting were performed using the Cell Ranger analysis pipeline (v2.0.0) provided by 10x Genomics. High quality barcodes were selected on the basis of the overall distribution of total UMI counts per cell using the standard Cell Ranger cell detection algorithm. All further analyses were run with python3 using the scanpy package<sup>62</sup> (v1.0.4+92.g9a754bb, <https://github.com/theislab/scanpy>) except stated otherwise. Genes with expression in less than ten cells were excluded. Furthermore, as is also applied as standard preprocessing steps in scanpy tutorials, low quality or outlier cells were removed if they: (1) had a high fraction of counts from mitochondrial genes (40% or more); (2) expressed more than 7,000 genes; or (3) had more than 100,000 UMI counts. Cell-by-gene count matrices of all samples were then concatenated to a single matrix. To account for differences in sequencing depth, UMI counts of each cell were normalised by total counts of that cell (pp.normalize\_per\_cell with mean = TRUE) and values were log-transformed. Highly variable genes ( $n = 1,625$ ) were selected on the basis of normalised dispersion using the setting the lower cutoffs for the mean to 0.0125 and for the dispersion to 0.5. This matrix was used as input for all further analyses unless otherwise indicated.

**Embedding, clustering and cell type annotation.** Clustering was performed on the full data set to reduce systematic biases such as batch effects as was recommended in ref. <sup>63</sup>. A single-cell neighbourhood graph (kNN-graph) was computed on the 50 first principal components using 15 neighbours. To minimise condition effects and to facilitate clustering we recomputed the kNN-graph using the first 15 diffusion components of the PCA-based graph as suggested in ref. <sup>64</sup>. For clustering and cell type annotation, Louvain-based clustering<sup>65</sup> was used as implemented in louvain-igraph (v0.6.1 <https://github.com/vtraag/louvain-igraph>) and adopted by scanpy (tl.louvain). The resolution parameter was varied in different parts of the data manifold to account for strong changes in resolution (for details, see Data availability). Clusters were annotated on the basis of the mRNA expression of the four main hormone genes *Ins1* and *Ins2*, *Gcg*, *Sst* and *Ppy* (endocrine cells) and other known marker genes (non-endocrine cells) and were merged if they reflected heterogeneity only in a cell type outside the focus of this study.

Ductal cells (that express *Krt19*), acinar cells (that express *Prss2*), endothelial cells (that express *Plvap*), stellate cells (that express *Col1a2*) and small clusters of potential doublet-like cells that co-express endocrine and non-endocrine markers were removed from further analysis. Immune cells (that express *Cd74*) that infiltrate the islets were finely subclustered into macrophages (that express *Cd86*, *Adgre1* and *Cd14*), dendritic cells (that express *Cd86*, *Ilgax* and *Iftim3*), B cells (that express *Cd79a* and *Cd79b*) and T cells (that express *Cd8a* and *Cd3d*).

The hormone genes *Ins1* and *Ins2*, *Gcg*, *Sst* and *Ppy* were expressed at very high levels and showed background level expression in all other endocrine subtypes and non-endocrine cell types. Such background expression is a common phenomenon in droplet-based scRNA-seq data. It is commonly said to be due to free-floating mRNA in the single-cell solution that comes from lysed cells and that is incorporated into all droplets. For annotation, only hormone expression that was well above the background level in non-endocrine cells, such as ductal, immune and endothelial cells, was considered.

For the identification of  $\beta$ -cell subtypes a new kNN-graph on the first 50 principal components was calculated and put into Louvain-based clustering of both *Ins* monohormonal and the connected *Ins-PP* cells. Similarly, *Ins-Sst* cells were subclustered from *Ins-Sst-PP* cells after recalculating the kNN-graph on the first 50 principal components. *Ins-Gcg-Sst* cells were assigned using a manual threshold for all three hormones that was well above ambient levels (threshold = 6 for normalised data).

For visualisation, UMAP was run as recommended in ref. <sup>66</sup>. For each UMAP-plot the UMAP was newly calculated by recomputing the kNN-graph on the represented cell subset using the first 50 principal components.

**Identification of polyhormonal singlets and doublet-like endocrine cell clusters.** Polyhormonal cells have previously been reported to exist in pancreatic islets<sup>67–69</sup>. However, the expression of multiple hormones in the same droplet can also be an indication of a doublet. It can therefore be difficult to distinguish polyhormonal singlets from doublets. A doublet rate of ~8–10% was measured in experiments with the same concentration of cells using the 10x technology<sup>70</sup>. This rate includes doublets with contributions from two different cell types (here, polyhormonal doublets) and from the same cell type (here, monohormonal doublets). The monohormonal doublets resemble monohormonal singlets and do not affect subsequent analyses<sup>71</sup>. We calculated the expected doublet frequency

of polyhormonal doublets for a doublet rate of 10% using the frequency of monohormonal cell types that contribute to the doublet (doublet contributors) and assuming that doublets are generated by sampling singlet cells uniformly at random<sup>71</sup>. In every sample the proportion of observed polyhormonal cells clearly exceeded the expected polyhormonal doublet frequency. Therefore, in our data set it is unlikely that all detected polyhormonal cells are doublets. Application of doublet cell detection tools Scrublet<sup>71</sup> (v0.1, <https://github.com/AllonKleinLab/scrublet>) and DoubletDetection (<https://github.com/JonathanShor/DoubletDetection>) failed to resolve which clusters represent doublets and which clusters represent polyhormonal singlets. Predictions of the tools disagreed with each other and the doublet rate was consistently overestimated. We therefore used the following criteria to evaluate polyhormonal cell clusters and to distinguish between singlets and doublets:

- (1) Doublets should not express unique genes. All genes should also be expressed in at least one doublet contributor.
- (2) Doublets should express marker genes or lineage-determining transcription factors of the doublet contributors. Downregulation of these genes indicates singlet populations.
- (3) Previous reporting of polyhormonal singlet cells in the literature.
- (4) Clusters of polyhormonal cells with a higher frequency than expected by our doublet simulation indicate polyhormonal singlet clusters (Extended Data Fig. 6g).
- (5) Clusters with Scrublet doublet score distributions that are comparable to monohormonal singlet clusters indicate polyhormonal singlets (Extended Data Fig. 6h).

On the basis of these criteria we found sufficient evidence for Ins-PP, Ins-Sst-PP, Gcg-PP (low) and Gcg-PP (high) cells to be polyhormonal singlets, but we excluded Ins-PP-Gcg, Ins-Gcg, Ins-Gcg-Sst, Gcg-Sst-PP and Sst-PP(high) cells as probable doublets.

**Cell-cycle classification.** To classify cells into cycling and non-cycling cells, first, a score was assigned to each cell for a set of S and G2/M phase genes<sup>72</sup> as proposed<sup>73</sup>, and second, all cells with an S-score or a G2/M-score > 0.25 were classified as cycling. The threshold was chosen on the basis of the score distribution. The score for a given gene set was computed as described<sup>74</sup> and implemented in scanpy (tl.score\_genes\_cell\_cycle).

**Marker genes of the main endocrine cell types.** For the characterisation of the four endocrine cell types, specific marker genes were identified by comparing the gene expression profile of each cell type against all cells of the other three cell types using a test with overestimated variance as implemented in scanpy (tl.rank\_genes\_groups). All genes that ranked within the top 300 genes, had a test score of >8 and were unique markers for one cell type were considered as marker genes.

**Differential expression testing to describe subpopulations and treatment responses.** Differential expression testing between treatments and for the characterisation of immature  $\beta$  cells and polyhormonal subpopulations was performed on quantile-normalised (quantile threshold = 0.95) and log-transformed data to account for extremely highly expressed genes (for example, the main hormones in endocrine cells) that may incorrectly alleviate the expression of other genes in a cell when total count normalisation is applied. By quantile normalisation, each cell is normalised by the total UMI count in the cell, of genes that account for less than 5% of the total UMI counts across all cells. Thus, very highly expressed genes are not considered for normalisation. For differential expression testing we used limma-trend<sup>75,76</sup>, as implemented in the Bioconductor package limma (v3.28.10), through an rpy2 (v2.9.1) interface as recommended in ref. <sup>77</sup>. In each test only genes expressed in >1% of cells in any of the two subsets tested were considered. Gene set enrichment was performed using EnrichR<sup>78</sup> through its web interface. Genes with a false discovery rate (FDR) of <0.01 and an estimated log(fold change) (output from limma model, not the actual log(fold change), as log-transformed data were the input) of >0.25 were used as input. Notably, the hormone genes *Ins1* and *Ins2*, *Gcg*, *Sst* and *Ppy*, as well as other known cell type marker genes *Pyy*, *Iapp*, *Trt*, *Gpx3*, *Ctrb1* and *Try5* were also differentially expressed in other cell-types in which they are expressed only at background levels (free-floating mRNA, see Embedding, clustering and annotation). These genes are indicated in the Supplementary Tables and were excluded from plotting.

**Identification of specific  $\beta$ -cell dedifferentiation markers.** Genes specific for dedifferentiated  $\beta$  cells ( $\beta$ -mSTZ) were extracted from the list of all significantly upregulated genes (FDR < 0.05, estimated log(fold change) > 0.25) in  $\beta$  cells from mSTZ-treated mice compared to  $\beta$  cells of healthy mice by two filtering steps. First, non-specific genes that were also differentially expressed in any of the other monohormonal endocrine cell types ( $\alpha$ ,  $\delta$  and PP) were excluded. Second, only genes that were expressed in at least 25% of  $\beta$  cells from mSTZ-treated mice and in less than 5% of  $\beta$  cells from healthy mice were considered. Location was extracted from the GeneCards database (<https://www.genecards.org>).

**Inference of  $\beta$ -cell maturation, dedifferentiation and regeneration trajectories.** Pseudotime of  $\beta$ -cell maturation in healthy islets and dedifferentiation upon mSTZ treatment was calculated using diffusion pseudotime (dpt)<sup>79,80</sup> as implemented in scanpy (tl.dpt), selecting a random root cell within the starting population. The

choice of root cell did not affect the inferred pseudotemporal ordering strongly. Similarly, the dpt approach was used to model  $\beta$ -cell regeneration and to estimate the location of treated  $\beta$  cells along the path from dedifferentiated to mature  $\beta$  cells. Here, dpt was used as a cell-to-cell distance metric across samples. Cycling cells as well as a small subpopulation of  $\beta$  cells were excluded from visualisation as they were not part of the linear trajectory and they showed very high pseudotime values.

**Comparison of  $\beta$ -cell dedifferentiation trajectory to embryonic and postnatal maturation.** To compare the dedifferentiation trajectory to embryonic and postnatal maturation we used a publicly available single-cell RNA-seq data set as a reference that contained cells that were sorted from Gcg-Venus and Ins-GFP reporter mice at 6 different time points (E17.5, P0, P3, P9, P18 and P60)<sup>37</sup>. The filtered and annotated raw count matrix was downloaded from the Gene Expression Omnibus (GEO) (GEO accession number: [GSE87375](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87375)). The analysis was run using the updated scanpy package v1.4.4, as only this version includes the necessary data integration methods. ERCC RNA spike-in and genes with expression in fewer than three cells were excluded. The data were normalised to total counts per cell using the *pp.normalize\_total* function in scanpy with default parameters and excluding highly expressed genes, and were log-transformed (*pp.log1p*). For the scope of this manuscript we used a subset of the data that contained only  $\beta$  cells (*Ins1*-positive cells). Therefore, we computed a kNN-graph on the 50 first principal components using 15 neighbours and performed a first round of Louvain-based clustering (tl.louvain). As input, data were subset to the 3,000 top ranking highly variable genes (*pp.highly\_variable\_genes*). We excluded all clusters that showed high expression of *Gcg*, which is indicative of  $\alpha$  cells, or that showed high expression of *Mki67*, which is indicative of proliferative cells. In addition, we filtered cells that showed high expression of the  $\delta$  cell markers *Sst* or *Hhex*.

We integrated this reference data set with the data from our study, which was subset to  $\beta$  cells from Ctrl and mSTZ-Vehicle treatment, using BBKNN<sup>81</sup> from the scanpy external package (sce), which allowed us to compute an embedding and trajectory that included both data sets. To reduce noise, we excluded genes that were expressed in fewer than 15 cells in each data set. In addition, cycling cells (*Mki67* > 1, 33 reference cells and 30 of the cells from this study) were excluded, as their expression profile is dominated by the expression of cell-cycle genes and these cells therefore formed a separate cluster which was not part of the linear maturation trajectory. To ensure that  $\beta$ -cell maturation dominates the gene expression variation we first considered only the 2,000 top-ranked highly variable genes (*pp.highly\_variable\_genes*) of the reference data set that were also expressed in our data, and, second, we reduced the contribution of heterogeneity within the  $\beta$ 1 and  $\beta$ -mSTZ cluster to gene expression variation by randomly subsetting both clusters to 500 cells. We then scaled and zero centred each data set separately (*pp.scale*) and concatenated the two data sets, which resulted in a 1,788 cells by 1,654 genes count matrix. We computed a common kNN-graph on the first ten principal components using the *sce.pp.bbkn* function with default parameters and *k* = 5 within batch neighbours. As the data sets did not show a strong batch effect even without integration, assessed by visual inspection of the first principal components and diffusion components, and thus transitions between cells from the two data sets also showed a high probability, we were able to use diffusion pseudotime (tl.dpt) to infer the maturation trajectory with the common kNN-graph as input. The trajectory was calculated by selecting a random root cell from the embryonic cells that were sampled at E17.5. The choice of root cell did not affect the inferred ordering strongly. The ordering of the cells from each data set was largely consistent with the ordering obtained prior to integration, assessed by the distribution along the trajectory of the time points or  $\beta$ -cell subgroups. To quantify the cluster similarity, PAGA was applied to the common kNN-graph (tl.paga).

To compute an embryonic or immaturity cell score and a maturity cell score we extracted the gene signatures from the reference data set. The reference data was subset to the 2,000 top ranked highly variable genes and Louvain-based clustering was performed on the kNN-graph that was computed on the first 50 principal components with 15 neighbours. The cluster that consisted of cells that were sampled at E17.5 and P0 was annotated as 'embryonic or neonatal', whereas the cluster consisting mainly of cells that were sampled at P60 was annotated as 'mature'. Differentially expressed genes between these two clusters were used for scoring. Genes that were upregulated in the mature cluster or in the embryonic or neonatal cluster were used as a gene set for maturity or for embryonic or immaturity score, respectively. For differential expression testing the *t*-test with overestimated variance implemented in the *tl.rank\_genes\_groups* function of scanpy was used. The top 500 ranked genes with a log(foldchange) > 0.25 and an adjusted *P* value < 0.01 were considered. Cell scores were computed using the *tl.score\_genes* function in scanpy.

**Inference of cluster-to-cluster distances, lineage relations and cell movement.** PAGA<sup>64</sup> was performed to infer cluster and lineage relations using the *tl.paga* function of scanpy with a threshold on edge significance of 0.05. In a PAGA graph, paths represent cluster connections or relations that indicate potential routes of cell transitions. Edge weights represent the confidence of a connection calculated on the basis of a measure of cluster connectivity.

To infer the direction of possible transitions<sup>82</sup> and cell movements we estimated RNA velocity using a stochastic version implemented in the scVelo python package

(v0.1.16.dev13 + c1a6dad, <https://github.com/theislab/scvelo> with scanpy v1.3.2). Splicing information of reads was extracted using the velocity pipeline (v0.17.7, <http://velocity.org>). We then followed the recommended steps described in scVelo to estimate RNA velocities and RNA velocity force fields. First, data were preprocessed by filtering genes with less than 30 spliced or 30 unspliced counts and both unspliced and spliced counts were normalised by total counts. Then the first- and second-order moments for each cell were computed across its 15 nearest neighbours of the kNN graph in PC space (50 PCs). Next, RNA velocities were estimated using a stochastic model of transcriptional dynamics. To obtain a more conservative estimate a 95% quantile fit was used. Finally, to project the velocity vector of each cell into the low-dimensional UMAP embedding for visualisation and interpretation, the expected mean direction given all potential cell transitions on the kNN graph was computed. Each potential cell transition is assigned a probability corresponding to the correlation to the predicted transition by the velocity vector (velocity graph). For example, a high probability corresponds to a high correlation with the velocity vector. The projection results in a low dimensional map of RNA velocity which indicates the predicted cell state transitions. For computation of the velocity graph and embedding only genes with an  $r^2 > 0.1$  of the velocity fit were considered.

The velocities of each gene were calculated over all treatments except for healthy  $\beta$  cells, for which only healthy cells were used. A treatment can here be considered as a process by which cells move from the diseased cells potentially towards healthy cells, as for the pseudotime inference described above. During this process genes are induced and/or repressed, as approximated by RNA velocity. Therefore, to also take into account these intermediate gene states, all treatments were included for model fitting and velocity estimation. Both PAGA and the RNA velocity graph and projection were instead only computed on the represented cell subset. For this, the kNN-graph was recalculated for the cell subset using the first 50 principal components and the highly variable genes as initially defined.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Custom python scripts written for performing scRNA-seq data analysis are available in a github repository ([https://github.com/theislab/pancreas-targeted\\_pharmacology](https://github.com/theislab/pancreas-targeted_pharmacology)). Versions of packages that might influence numerical results are indicated in the scripts. Raw data and gene expression matrices of scRNA-seq are deposited in GEO under the accession number [GSE128565](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128565). Source data for Figs. 1–3 and 7 and Extended Data Figs. 1–5 and 10 are provided with the paper.

Received: 31 July 2019; Accepted: 15 January 2020;

Published online: 20 February 2020

## References

- Matveyenko, A. V. & Butler, P. C. Relationship between  $\beta$ -cell mass and diabetes onset. *Diabetes Obes. Metab.* **10**, 23–31 (2008).
- Herold, K. C. et al. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. *N. Engl. J. Med.* **381**, 603–613 (2019).
- Harrison, L. B., Adams-Huet, B., Raskin, P. & Lingvay, I.  $\beta$ -cell function preservation after 3.5 years of intensive diabetes therapy. *Diabetes Care* **35**, 1406–1412 (2012).
- Chen, H.-S. et al. Beneficial effects of insulin on glycemic control and beta-cell function in newly diagnosed type 2 diabetes with severe hyperglycemia after short-term intensive insulin therapy. *Diabetes Care* **31**, 1927–1932 (2008).
- The Diabetes Control and Complications Trial Research Group. Effect of intensive therapy on residual  $\beta$ -cell function in patients with type 1 diabetes in the diabetes control and complications trial: a randomized, controlled trial. *Ann. Intern. Med.* **128**, 517–523.
- Weng, J. et al. Effect of intensive insulin therapy on  $\beta$ -cell function and glycaemic control in patients with newly diagnosed type 2 diabetes: a multicentre randomised parallel-group trial. *The Lancet* **371**, 1753–1760 (2008).
- Alvarsson, M. et al. Beneficial effects of insulin versus sulphonylurea on insulin secretion and metabolic control in recently diagnosed type 2 diabetic patients. *Diabetes Care* **26**, 2231–2237 (2003).
- Rui, J. et al.  $\beta$  cells that resist immunological attack develop during progression of autoimmune diabetes in NOD mice. *Cell Metab.* **25**, 727–738 (2017).
- Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic  $\beta$  cell dedifferentiation as a mechanism of diabetic  $\beta$  cell failure. *Cell* **150**, 1223–1234 (2012).
- Cinti, F. et al. Evidence of  $\beta$ -cell dedifferentiation in human type 2 diabetes. *J. Clin. Endocrinol. Metab.* **101**, 1044–1054 (2016).
- Like, A. A. & Rossini, A. A. Streptozotocin-induced pancreatic insulinitis: new model of diabetes mellitus. *Science* **193**, 415–417 (1976).
- Thorel, F. et al. Conversion of adult pancreatic  $\alpha$ -cells to  $\beta$ -cells after extreme  $\beta$ -cell loss. *Nature* **464**, 1149–1154 (2010).
- Chera, S. et al. Diabetes recovery by age-dependent conversion of pancreatic  $\delta$ -cells into insulin producers. *Nature* **514**, 503–507 (2014).
- Brereton, M. F. et al. Reversible changes in pancreatic islet structure and function produced by elevated blood glucose. *Nat. Commun.* **5**, 4639 (2014).
- Wang, Z., York, N. W., Nichols, C. G. & Remedi, M. S. Pancreatic  $\beta$  cell dedifferentiation in diabetes and redifferentiation following insulin therapy. *Cell Metab.* **19**, 872–882 (2014).
- Tiano, J. P. & Mauvais-Jarvis, F. Importance of oestrogen receptors to preserve functional  $\beta$ -cell mass in diabetes. *Nat. Rev. Endocrinol.* **8**, 342–351 (2012).
- Chon, S. & Gautier, J.-F. An update on the effect of incretin-based therapies on  $\beta$ -cell function and mass. *Diabetes Metab. J.* **40**, 99–114 (2016).
- Marso, S. P. et al. Semaglutin and cardiovascular outcomes in patients with type 2 diabetes. *N. Engl. J. Med.* **375**, 1834–1844 (2016).
- Finan, B. et al. Targeted estrogen delivery reverses the metabolic syndrome. *Nat. Med.* **18**, 1847–1856 (2012).
- Clemmensen, C. et al. Emerging hormonal-based combination pharmacotherapies for the treatment of metabolic diseases. *Nat. Rev. Endocrinol.* **15**, 90–104 (2019).
- Bastidas-Ponce, A. et al. Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic  $\beta$ -cells. *Mol. Metab.* **6**, 524–534 (2017).
- Blum, B. et al. Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. *Nat. Biotechnol.* **30**, 261–264 (2012).
- Nishimura, W. et al. A switch from MafB to MafA expression accompanies differentiation to pancreatic beta-cells. *Dev. Biol.* **293**, 526–539 (2006).
- Bader, E. et al. Identification of proliferative and mature  $\beta$ -cells in the islets of Langerhans. *Nature* **535**, 430–434 (2016).
- Roscioni, S. S., Migliorini, A., Gegg, M. & Lickert, H. Impact of islet architecture on  $\beta$ -cell heterogeneity, plasticity and function. *Nat. Rev. Endocrinol.* **12**, 695–709 (2016).
- Ediger, B. N. et al. Islet-1 is essential for pancreatic  $\beta$ -cell function. *Diabetes* **63**, 4206–4217 (2014).
- Gao, T. et al. Pdx1 maintains  $\beta$  cell identity and function by repressing an  $\alpha$  cell program. *Cell Metab.* **19**, 259–271 (2014).
- Gu, C. et al. Pancreatic  $\beta$  cells require NeuroD to achieve and maintain functional maturity. *Cell Metab.* **11**, 298–310 (2010).
- Gutiérrez, G. D. et al. Pancreatic  $\beta$  cell identity requires continual repression of non- $\beta$  cell programs. *J. Clin. Invest.* **127**, 244–259 (2016).
- Swisa, A. et al. PAX6 maintains  $\beta$  cell identity by repressing genes of alternative islet cell types. *J. Clin. Invest.* **127**, 230–243 (2016).
- Taylor, B. L., Liu, F.-F. & Sander, M. Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. *Cell Rep.* **4**, 1262–1275 (2013).
- Kim-Muller, J. Y. et al. Aldehyde dehydrogenase 1a3 defines a subset of failing pancreatic  $\beta$  cells in diabetic mice. *Nat. Commun.* **7**, 12631 (2016).
- Dahan, T. et al. Pancreatic  $\beta$ -cells express the fetal islet hormone gastrin in rodent and human diabetes. *Diabetes* **66**, 426–436 (2017).
- Solimena, M. et al. Systems biology of the IMIDIA biobank from organ donors and pancreatectomised patients defines a novel transcriptomic signature of islets from individuals with type 2 diabetes. *Diabetologia* **61**, 641–657 (2018).
- Camunas-Soler, J. et al. Pancreas patch-seq links physiologic dysfunction in diabetes to single-cell transcriptomic phenotypes. Preprint at *bioRxiv* <https://doi.org/10.1101/555110> (2019).
- Weir, G. C. & Bonner-Weir, S. Islet  $\beta$  cell mass in diabetes and how it relates to function, birth, and death: islet  $\beta$  cell mass in diabetes. *Ann. N. Y. Acad. Sci.* **1281**, 92–105 (2013).
- Qiu, W.-L. et al. Deciphering pancreatic islet  $\beta$  cell and  $\alpha$  cell maturation pathways and characteristic features at the single-cell level. *Cell Metab.* **25**, 1194–1205.e4 (2017).
- Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nat. Rev. Mol. Cell Biol.* **7**, 85–96 (2006).
- Rowlands, J., Heng, J., Newsholme, P. & Carlessi, R. Pleiotropic effects of GLP-1 and analogs on cell signaling, metabolism, and function. *Front. Endocrinol.* **9**, 672 (2018).
- Segars, J. H. & Driggers, P. H. Estrogen action and cytoplasmic signaling cascades. Part I: membrane-associated signaling complexes. *Trends Endocrinol. Metab.* **13**, 349–354 (2002).
- Hancock, M. L. et al. Insulin receptor associates with promoters genome-wide and regulates gene expression. *Cell* **177**, 722–736.e22 (2019).
- Kulkarni, R. N. et al. Altered function of insulin receptor substrate-1-deficient mouse islets and cultured  $\beta$ -cell lines. *J. Clin. Invest.* **104**, R69–R75 (1999).
- Ueki, K. et al. Total insulin and IGF-I resistance in pancreatic  $\beta$  cells causes overt diabetes. *Nat. Genet.* **38**, 583–588 (2006).
- Fonseca, S. G., Gromada, J. & Urano, F. Endoplasmic reticulum stress and pancreatic  $\beta$ -cell death. *Trends Endocrinol. Metab.* **22**, 266–274 (2011).
- Sims, E. K. et al. Elevations in the fasting serum proinsulin-to-C-peptide ratio precede the onset of type 1 diabetes. *Diabetes Care* **39**, 1519–1526 (2016).
- Xu, B. et al. Estrogens promote misfolded proinsulin degradation to protect insulin production and delay diabetes. *Cell Rep.* **24**, 181–196 (2018).

47. Tiwari, A. et al. SDF2L1 interacts with the ER-associated degradation machinery and retards the degradation of mutant proinsulin in pancreatic  $\beta$ -cells. *J. Cell Sci.* **126**, 1962–1968 (2013).
48. Ho, D. V. & Chan, J. Y. Induction of Herpud1 expression by ER stress is regulated by Nrf1. *FEBS Lett.* **589**, 615–620 (2015).
49. Wong, N., Morahan, G., Stathopoulos, M., Proietto, J. & Andrikopoulos, S. A novel mechanism regulating insulin secretion involving Herpud1 in mice. *Diabetologia* **56**, 1569–1576 (2013).
50. Belmont, P. J. et al. Roles for endoplasmic reticulum-associated degradation and the novel endoplasmic reticulum stress response gene Derlin-3 in the ischemic heart. *Circ. Res.* **106**, 307–316 (2010).
51. Zhu, D. et al. Single-cell transcriptome analysis reveals estrogen signaling coordinately augments one-carbon, polyamine, and purine synthesis in breast cancer. *Cell Rep.* **25**, 2285–2298.e4 (2018).
52. Torrent, M., Chalancon, G., de Groot, N. S., Wuster, A. & Madan Babu, M. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci. Signal.* **11**, eaat6409 (2018).
53. Xu, G. et al. Downregulation of GLP-1 and GIP receptor expression by hyperglycemia: possible contribution to impaired incretin effects in diabetes. *Diabetes* **56**, 1551–1558 (2007).
54. Fritsche, A., Stefan, N., Hardt, E., Häring, H. & Stumvoll, M. Characterisation of beta-cell dysfunction of impaired glucose tolerance: evidence for impairment of incretin-induced insulin secretion. *Diabetologia* **43**, 852–858 (2000).
55. Kjemis, L. L., Holst, J. J., Vølund, A. & Madsbad, S. The influence of GLP-1 on glucose-stimulated insulin secretion: effects on beta-cell sensitivity in type 2 and nondiabetic subjects. *Diabetes* **52**, 380–386 (2003).
56. Jonas, J. C. et al. Chronic hyperglycemia triggers loss of pancreatic beta cell differentiation in an animal model of diabetes. *J. Biol. Chem.* **274**, 14112–14121 (1999).
57. Keenan, H. A. et al. Residual insulin production and pancreatic  $\beta$ -cell turnover after 50 years of diabetes: Joslin Medalist Study. *Diabetes* **59**, 2846–2853 (2010).
58. Lam, C. J., Chatterjee, A., Shen, E., Cox, A. R. & Kushner, J. A. Low-level insulin content within abundant non- $\alpha$  islet endocrine cells in long-standing type 1 diabetes. *Diabetes* **68**, 598–608 (2019).
59. Seiron, P. et al. Characterisation of the endocrine pancreas in type 1 diabetes: islet size is maintained but islet number is markedly reduced. *J. Pathol. Clin. Res.* **5**, 248–255 (2019).
60. Zhou, Q. & Melton, D. A. Pancreas regeneration. *Nature* **557**, 351–358 (2018).
61. Waaseth, M. et al. Hormone replacement therapy use and plasma levels of sex hormones in the Norwegian Women and Cancer postgenome cohort: a cross-sectional analysis. *BMC Womens Health* **8**, 1 (2008).
62. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
63. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019).
64. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
65. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
66. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).
67. Chiang, M.-K. & Melton, D. A. Single-cell transcript analysis of pancreas development. *Dev. Cell* **4**, 383–393 (2003).
68. Katsuta, H. et al. Single pancreatic beta cells co-express multiple islet hormone genes in mice. *Diabetologia* **53**, 128–138 (2010).
69. Alpert, S., Hanahan, D. & Teitelman, G. Hybrid insulin genes reveal a developmental lineage for pancreatic endocrine cells and imply a relationship with neurons. *Cell* **53**, 295–308 (1988).
70. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
71. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
72. Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
73. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
74. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
75. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
76. Ritchie, M. E. et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
77. Sonesson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).
78. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
79. Haghverdi, L., Büttner, M., Wolf, F. A., Büttner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
80. Tritschler, S. et al. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, dev170506 (2019).
81. Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2019).
82. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

## Acknowledgements

We thank L. Müller, L. Seher, E. Malogajski and M. Kilian from the Helmholtz Diabetes Center in Munich for excellent assistance with in vivo mouse experiments. We thank J. Jaki, C. Salinno, F. Volta, J. Beckenbauer, A. Savoca and R. Fimmen for excellent assistance with in vitro experiments. We thank C. Pyke and P. Gottrup Mortensen for providing the GLP-1R antibody. We thank V. Bergen, M. Lücken, L. Simon and D. Fischer for fruitful discussions on the computational analysis. This work was supported in part by funding to M.H.T. from the Alexander von Humboldt Foundation and the Initiative and Networking Fund of the Helmholtz Association and funding by the European Research Council (AdG HypoFlam grant no. 695054). In addition, this work was supported by funds from the Helmholtz future topic ‘Aging and Metabolic programming’, the Helmholtz Alliance ICEMED, the Helmholtz Initiative on Personalised Medicine, iMed, by the Helmholtz Association, and the Helmholtz cross-programme topic ‘Metabolic Dysfunction’. M.A.S.-G. was funded by a Marie Skłodowska-Curie individual Fellowship (grant no. 706965, GCG-T3 Dyslipidemia). F.J.T. acknowledges support by the BMBF (grant no. 01IS18036A and grant no. 01IS18053A), by the German Research Foundation (DFG) within the Collaborative Research Centre 1243, Subproject A17, by the Helmholtz Association (Incubator grant sparse2big, grant no. ZT-I-0007) and by the Chan Zuckerberg Initiative DAF (advised fund of Silicon Valley Community Foundation, 182835).

## Author contributions

S.S. performed in vivo and ex vivo rodent experiments, pancreas histology, analysed and interpreted all data, interpreted scRNA-seq data, and wrote the manuscript. A.B.-P. performed ex vivo rodent experiments, pancreas histology, analysed and interpreted data, and cowrote the manuscript. S.T. analysed and interpreted scRNA-seq data and cowrote the manuscript. M.B. performed ex vivo experiments and helped to draft the manuscript. A.B. performed ex vivo experiments and helped to prepare the single-cell suspensions for scRNA-seq. M.A.S.-G. performed in vivo experiments. M.T.-M. performed ex vivo experiments. M.K., K.F., S.J., and A.H. performed in vivo experiments and helped to interpret data. E.B. and S.R. performed ex vivo experiments and helped to interpret data. S.U. helped to interpret data. A.F. conducted and analysed automatic pancreatic histology. B.Y. and A.N. performed, analysed, and interpreted human micro-islet experiments. C.B.J. designed, analysed, interpreted and supervised the rat study, interpreted in vivo data and helped to write the manuscript. M.C. designed and oversaw human micro-islet experiments and helped to interpret data. B.Y. synthesised and characterised compounds. B.F. designed the in vivo rodent experiment, synthesised and characterised compounds, interpreted the data, and helped to write the manuscript. R.D.D. and M.H.T. conceptualised and interpreted all studies and helped to write the manuscript. F.J.T. conceptualised, supervised, and interpreted the scRNA-seq analysis and helped to write the manuscript. S.M.H., T.D.M., and H.L. conceptualised, designed, supervised, and interpreted all studies and wrote the manuscript.

## Competing interests

C.B.J., M.C., B.Y., B.F., and R.D.D. are current employees of Novo Nordisk. Novo Nordisk has licensed from Indiana University intellectual property pertaining to this report. M.H.T. serves as a scientific advisory board member of ERX Pharmaceuticals, Inc., Cambridge, MA. F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH and Cellarity Inc., and ownership interest in Cellarity, Inc. and Dermagnostix. S.T. reports receiving consulting fees from Cellarity, Inc. The Institute for Diabetes and Obesity receives research support from Novo Nordisk. All other authors declare no conflict of interest.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42255-020-0171-3>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42255-020-0171-3>.

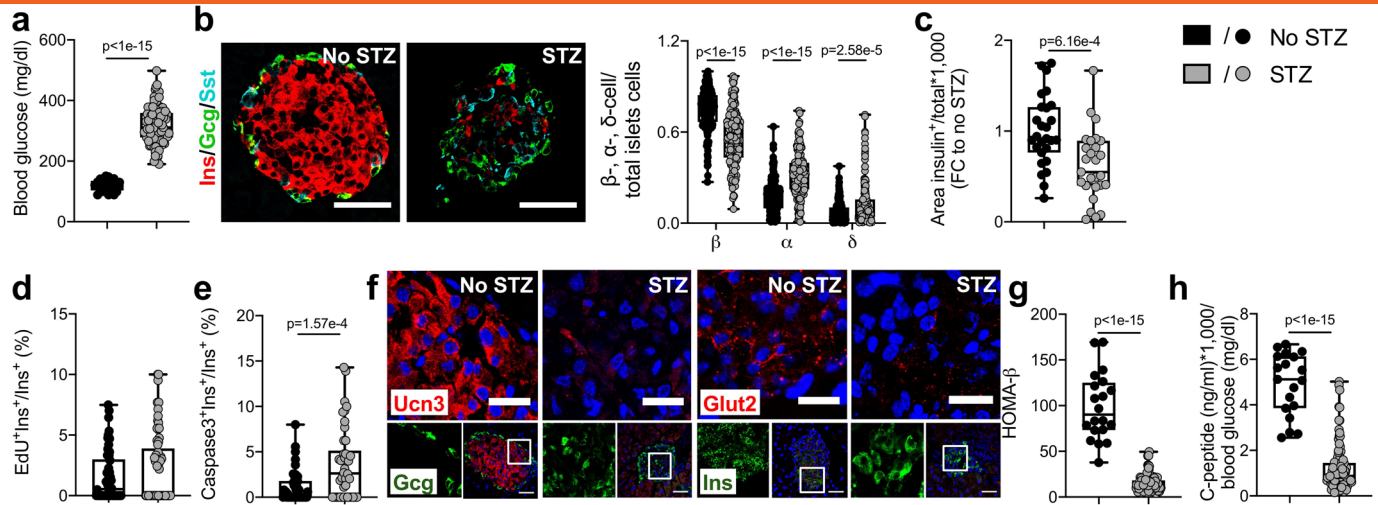
**Correspondence and requests for materials** should be addressed to F.J.T., S.M.H., T.D.M. or H.L.

**Peer review information** Primary Handling Editor: Elena Bellafante.

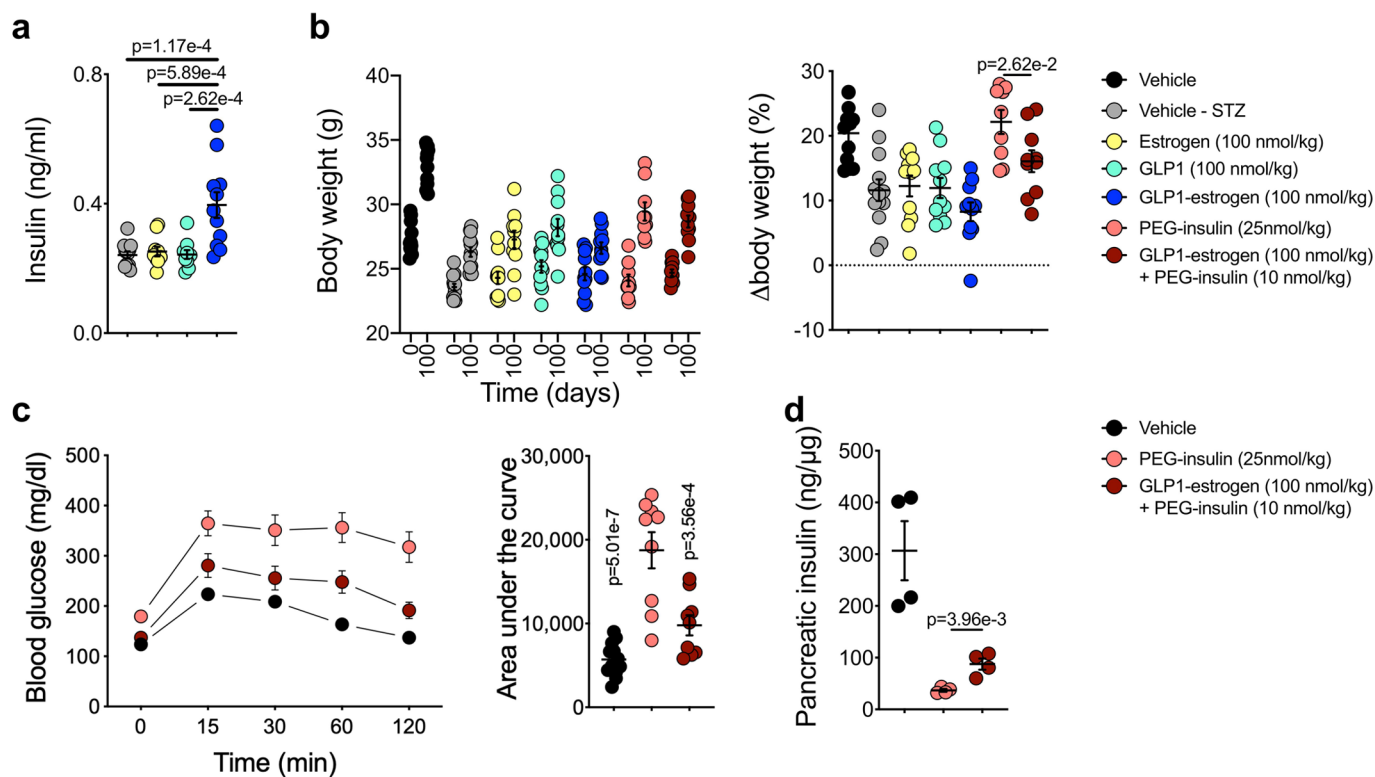
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

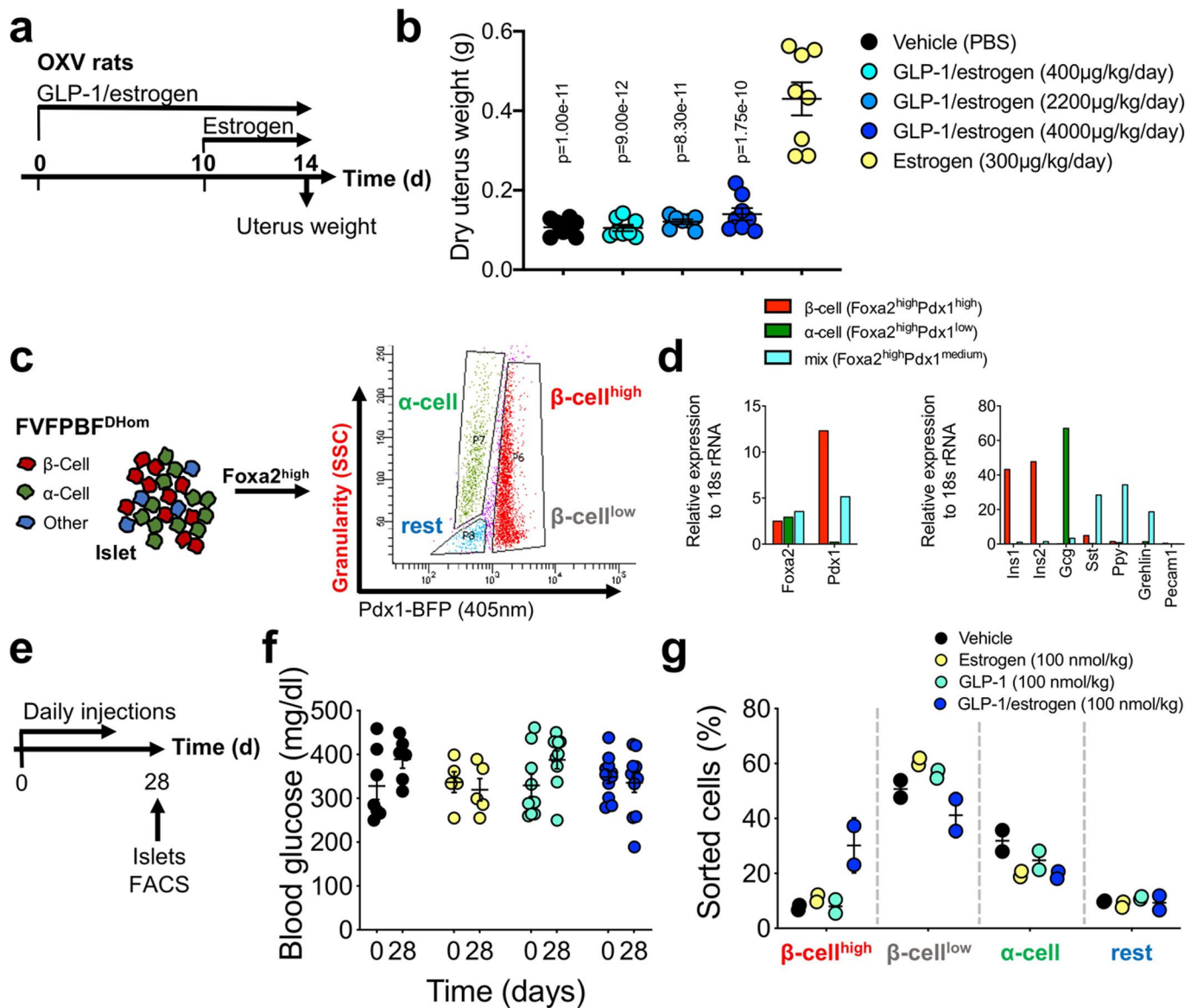
© The Author(s), under exclusive licence to Springer Nature Limited 2020



**Extended Data Fig. 1 | Remaining  $\beta$  cells lose cell identity 10 days after last STZ injection.** Effects of either vehicle or the mSTZ treatment on **a**, fasting blood glucose (No STZ:  $n=20$ , mSTZ:  $n=107$ ; unpaired two-sided t-test;  $t=14.64$ ,  $df=125$ ), **b**, pancreatic islets histology (No STZ: 179 islets of  $n=3$  mice, mSTZ: 182,  $n=3$ ; unpaired two-sided t-test;  $\beta$ :  $t=11.44$ ,  $df=358$ ;  $\alpha$ :  $t=10.98$ ,  $df=356$ ;  $\delta$ :  $t=4.27$ ,  $df=338$ ; images are representative from no STZ  $n=3$  and mSTZ  $n=3$  mice), **c**, the insulin positive area within pancreatic sections (No STZ: 27 sections of  $n=3$  mice; STZ: 27,  $n=3$ ; unpaired two-sided t-test;  $t=3.646$ ,  $df=52$ ), **d**, the proliferation (No STZ: 58 islets of  $n=3$  mice; STZ: 69,  $n=3$ ) and **e**, apoptosis rate in  $\beta$  cells (No STZ: 46 islets of  $n=3$  mice; STZ: 42,  $n=3$ ; unpaired two-sided t-test,  $t=3.955$ ,  $df=86$ ), **f**, the expression of  $\beta$ -cell functional marker Ucn3 and Glut2 (images are representative of dataset plotted in **b** from no STZ  $n=3$  and mSTZ  $n=3$  mice), **g**, the homeostatic model assessment of  $\beta$ -cell function (HOMA- $\beta$ ) (No STZ:  $n=20$ , STZ:  $n=107$ ; unpaired two-sided t-test;  $t=20.65$ ,  $df=124$ ) and **h**, the ratio of fasting C-peptide to fasting blood glucose (No STZ:  $n=20$ , STZ:  $n=106$ ; unpaired two-sided t-test;  $t=14.03$ ,  $df=122$ ). Boxplots covering all data points are depicted. Line indicates the median. Scale bar, 50  $\mu\text{m}$ . Scale bar zoom-in, 20  $\mu\text{m}$ .

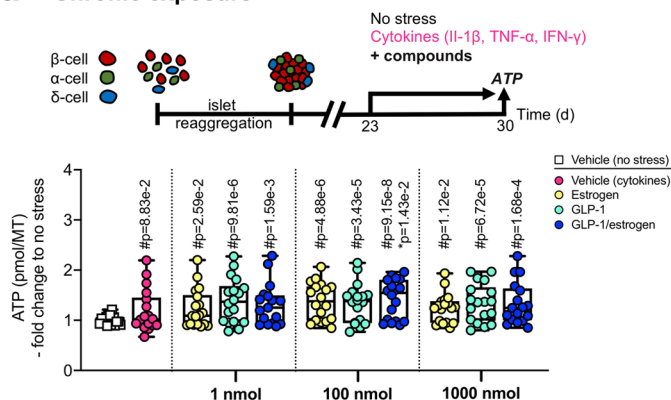
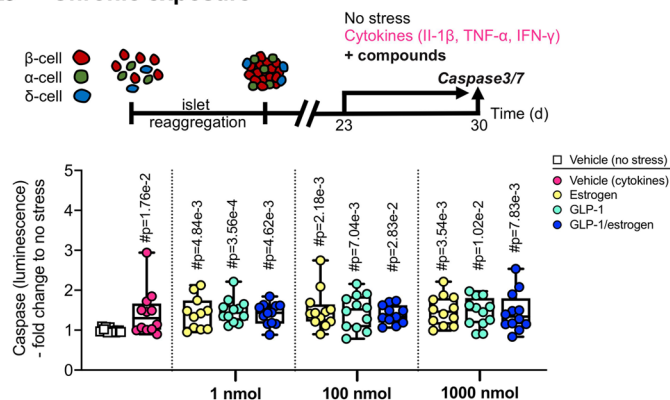


**Extended Data Fig. 2 | Benefits of polypharmacotherapy to ameliorate mSTZ diabetes in mice.** Effect of treatment with indicated compounds and doses on **a**, fasting plasma insulin levels at week 12 of treatment (mSTZ-vehicle,  $n=12$ ; oestrogen,  $n=10$ ; GLP-1,  $n=11$ ; GLP-1/oestrogen,  $n=11$ ; one-way ANOVA with Tukey post-hoc;  $F(3, 39)=10.66$ ) and **b**, body weight in the end of the study (no STZ-vehicle,  $n=12$ ; mSTZ-vehicle,  $n=13$ ; GLP-1,  $n=11$ ; oestrogen,  $n=11$ ; GLP-1/oestrogen,  $n=11$ ; PEG-insulin,  $n=9$ ; GLP-1/oestrogen and PEG-insulin,  $n=10$ ; unpaired two-sided t-test;  $t=2.436$ ,  $df=17$ ). (**c**, **d**) Comparison of PEG-insulin and GLP-1/oestrogen plus PEG-insulin co-treated mice. **c**, Blood glucose after intraperitoneal glucose (0.5 g/kg) at week 12 (no STZ-vehicle,  $n=12$ ; PEG-insulin,  $n=9$ ; GLP-1/oestrogen and PEG-insulin,  $n=10$ ; one-way ANOVA with Tukey post-hoc ( $F(2, 27)=24.71$ )). **d**, Pancreatic insulin content in the end of the study (no STZ-vehicle,  $n=4$ ; PEG-insulin,  $n=4$ ; GLP-1/oestrogen and PEG-insulin,  $n=4$ ; unpaired two-sided t-test;  $t=4.534$ ,  $df=6$ ). All data are mean  $\pm$  SEM.



**Extended Data Fig. 3 | Tissue specificity and  $\beta$ -cell selectivity of the GLP-1/oestrogen conjugate.** (a, b) Treatment of female OVX Sprague-Dawley rats. a, Study scheme. b, Dry uterus weight. Data are mean  $\pm$  SEM. N=8 female rats per group. One-way ANOVA with Tukey post-hoc ( $F(4, 34) = 44.89$ ). (c-g) Treatment of male FVFPBFDHOM mice. (c) FACS gating strategy of dispersed endocrine cells based on granularity (Side Scatter Cell (SSC)) and PBF (405 nm) intensity. d, qPCR analysis confirmed sorting strategy of endocrine cells. Data are values of sorted cells from n=2 mice. e, Study scheme; FVFPBFDHOM male mice were treated with vehicle (n=7), oestrogen (n=5), GLP-1 (n=9), or GLP-1/oestrogen (n=11) at the indicated doses for four weeks. f, Fasting blood glucose. Data are mean  $\pm$  SEM. g, Sorted endocrine cell populations after treatment (vehicle (n=4, cells of n=2 mice each were pooled), oestrogen (n=4, cells of n=2 mice each were pooled), GLP-1 (n=5, cells of n=2 and n=3 mice were pooled), or GLP-1/oestrogen (n=6, islets of n=3 mice each were pooled)). Data are mean  $\pm$  SD.



**a Chronic exposure****b Chronic exposure**

**Extended Data Fig. 4 | Viability and cell death of human micro-islets.** Measurement of human micro-islet viability and cell death with and without cytokine exposure and in the present of different compounds at the indicated doses. **(a)** ATP content and **(b)** Caspase 3/7 activity of human micro-islets.

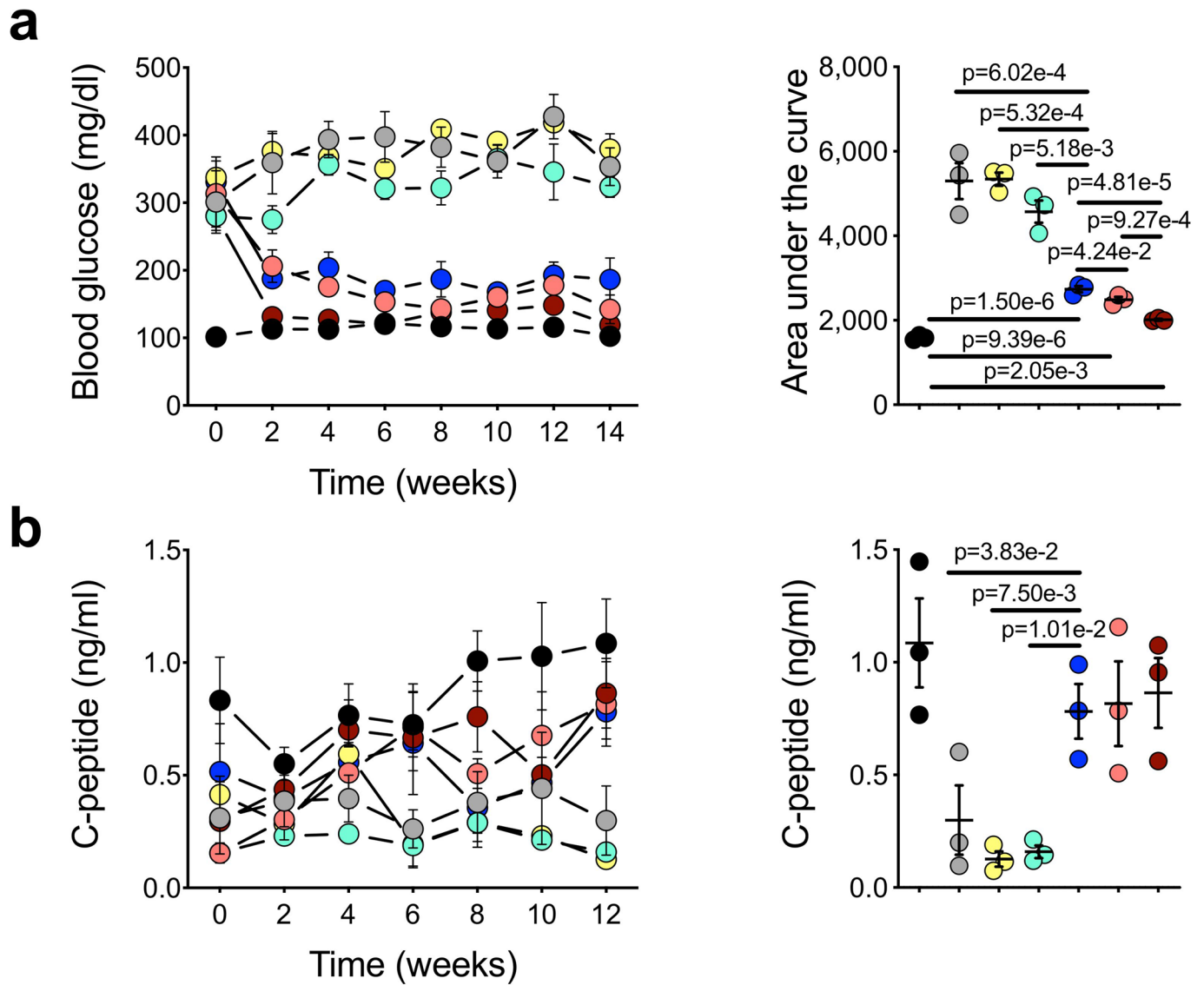
**a**, N=6 micro-islets of n=3 human donors for each condition. Boxplot of all data points. Line indicates the median. #P indicates P-value to vehicle no stress condition. \*P indicates P-value to vehicle cytokine exposure. No stress versus cytokines stress by unpaired two-sided t-test ( $t=1.756$ ,  $df=33$ ).

Otherwise one-way ANOVA with donor as random effect followed by Tukey post-hoc ( $F_{\text{low dose}}(4, 81)=6.68$ ;  $F_{\text{medium dose}}(4, 82)=10.68$ ;  $F_{\text{high dose}}(4, 81)=6.30$ ).

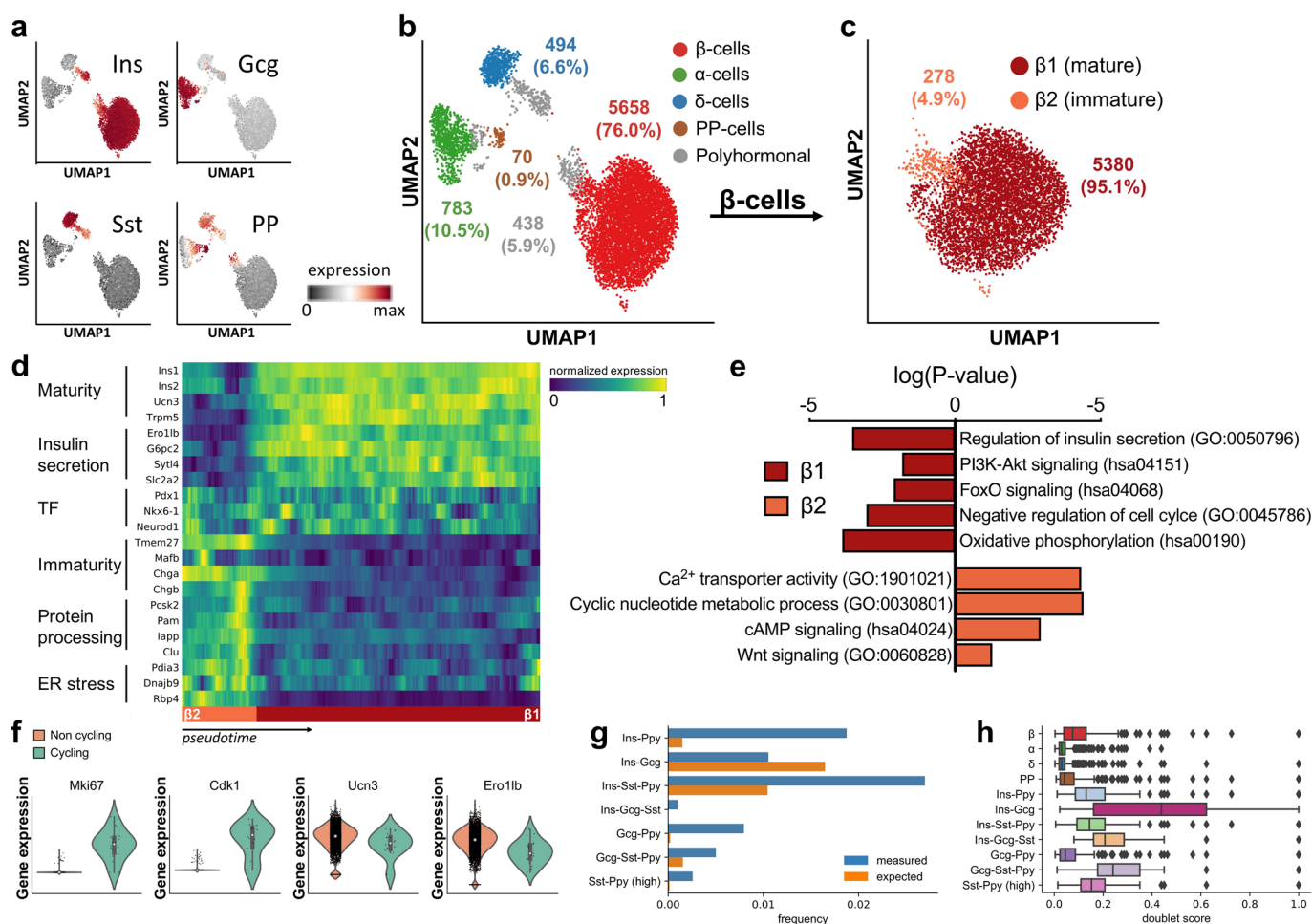
**b**, N=3-4 micro-islets of n=3 human donors for each condition. Boxplot of all data points. Line indicates the median. #P indicates P-value to vehicle no stress condition. \*P indicates P-value to vehicle cytokine exposure. No stress versus cytokines stress by unpaired two-sided t-test ( $t=2.567$ ,  $df=22$ ).

Otherwise one-way ANOVA with donor as random effect followed by Tukey post-hoc ( $F_{\text{low dose}}(4, 52)=5.58$ ;  $F_{\text{medium dose}}(4, 52)=4.44$ ;  $F_{\text{high dose}}(4, 53)=4.23$ ).

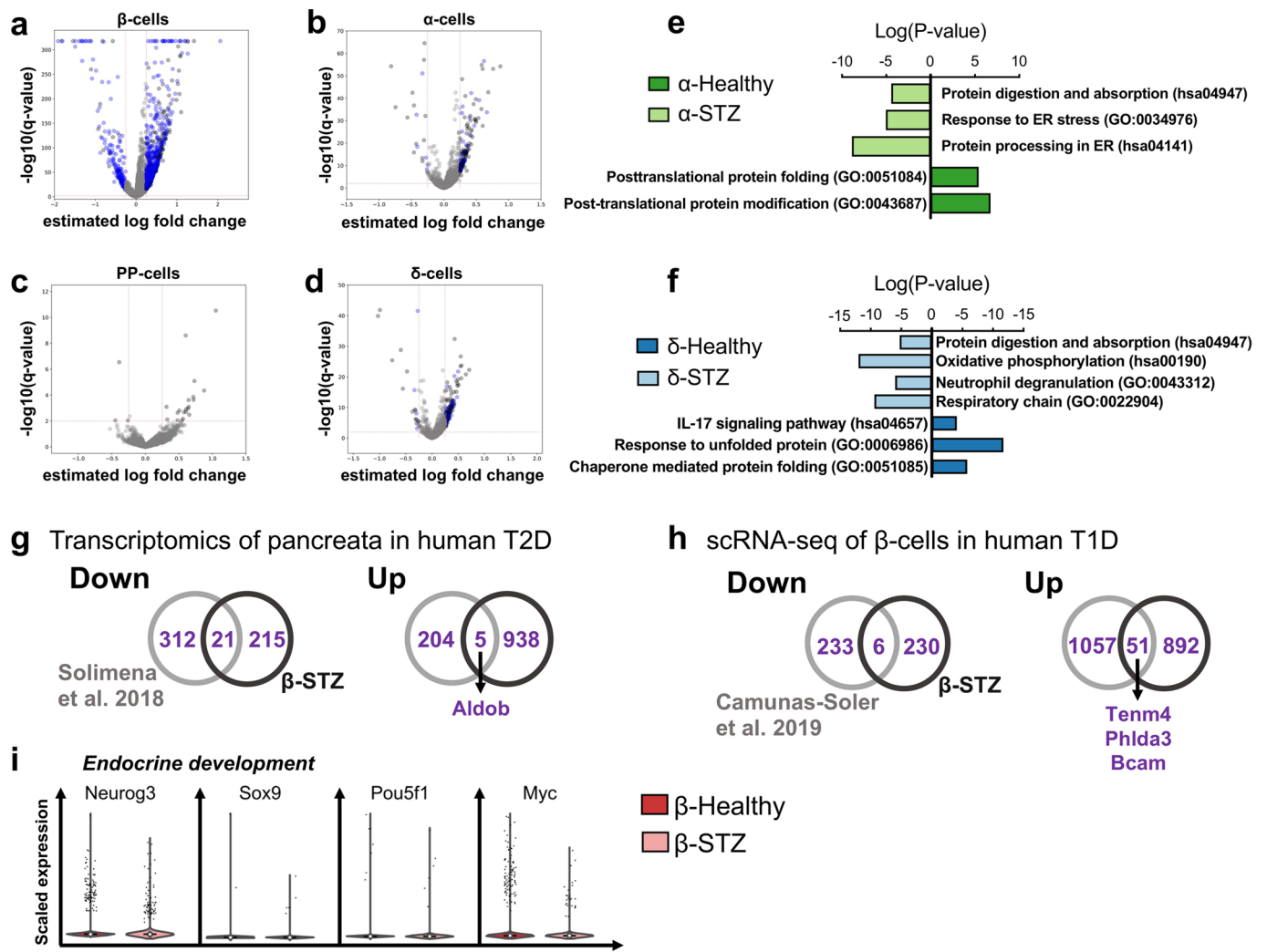
- Vehicle
- Estrogen (100 nmol/kg)
- PEG-insulin (25nmol/kg)
- Vehicle - STZ
- GLP1 (100 nmol/kg)
- GLP1-estrogen (100 nmol/kg)
- GLP1-estrogen (100 nmol/kg) + PEG-insulin (10 nmol/kg)



**Extended Data Fig. 5 | Physiological characteristics of mice used for scRNA-seq.** Representative mice ( $n=3$ ) of each treatment were used for scRNA-seq. **a**, Fasting glucose levels. One-way ANOVA with Tukey post-hoc test among mSTZ, oestrogen, GLP-1 and GLP-1/oestrogen treated mice ( $F(3, 8)=21.23$ ). One-way ANOVA with Tukey post-hoc among no STZ, GLP-1/oestrogen, PEG-insulin, and co-treated mice ( $F(3, 8)=94.06$ ). **b**, Fasting C-peptide levels. One-way ANOVA with Tukey post-hoc test among STZ, oestrogen, GLP-1 and GLP-1/oestrogen treated mice ( $F(3, 8)=9.073$ ). All data are mean  $\pm$  SEM.



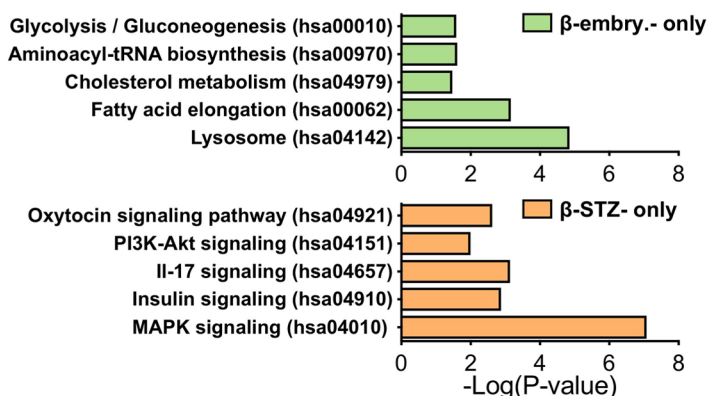
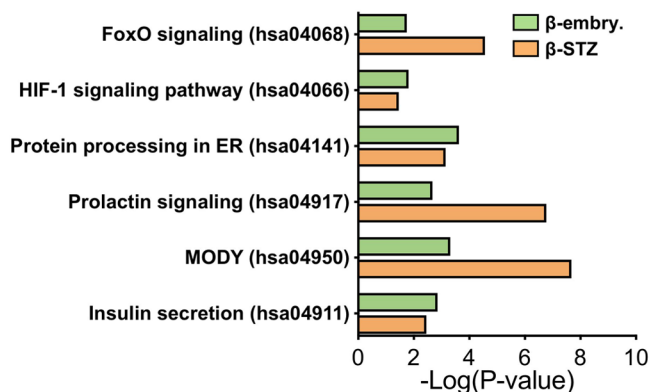
**Extended Data Fig. 6 |  $\beta$ -cell heterogeneity in healthy mice.** **a**, Endocrine cell annotation is based on the hormone expression of insulin (Ins), glucagon (Gcg), somatostatin (Sst), and pancreatic polypeptide (PP). **b**, UMAP plot showing all endocrine cells (7578 cells in total) from healthy mice. The cell number and proportion of each endocrine cluster is indicated. **c**, Redefined clustering of the  $\text{Ins}^+$   $\beta$  cells revealed two main  $\beta$ -cell subpopulations. **d**, Expression changes of genes from selected pathways along a pseudotime trajectory from  $\beta 2$ - to  $\beta 1$  cells.  $\beta 1$ -cells were downsampled to 1000 cells for better visualisation. **e**, GO term and KEGG pathway enrichment analysis of up- ( $\log(\text{fold change}) > 0.25$ ) and downregulated ( $\log(\text{fold change}) < -0.25$ ) genes in  $\beta 2$ - (278 cells) compared to  $\beta 1$  cells (5380 cells). Cells were pooled from  $n = 3$  mice. Representative terms from Supplementary table 2 are depicted. We used limma-trend to find differentially expressed genes (M&M). Gene enrichment was done with EnrichR using Fisher's exact test to identify regulated ontologies/pathways (M&M). **f**, Violin plots showing the distribution of the expression of proliferation and  $\beta$ -cell maturation genes suggesting an immature phenotype of cycling  $\beta$ -cells. Accordingly, 16/403 of the  $\beta 2$  cells, whereas only 2/5319 of the mature  $\beta 1$  cells were classified as cycling (M&M). Cells were pooled from  $n = 3$  mice. Violin shows the distribution as a kernel density estimate fit. Points in violin interior show individual data points. Boxplot in violin interior shows median, quartile and whisker values. **g**, Measured proportion and expected doublet frequency of polyhormonal cell clusters. Expected doublet frequency is calculated given a doublet rate of 10% (M&M). **h**, Boxplot displaying the doublet score distribution of mono- and polyhormonal cell clusters. A high score indicates a high doublet probability. Cells were pooled from  $n = 3$  mice. Boxplot shows the quartile values and extreme values. Whiskers extend to 1.5 IQRs of the lower and upper quartile. Outliers are displayed individually.



**Extended Data Fig. 7 |  $\beta$ -cell dedifferentiation in mSTZ-diabetic mice.** (a–d), Volcano plots showing differential expression and its significance ( $-\log_{10}(\text{adjusted p-Value})$ , limma-trend) for each gene in (a)  $\beta$ -, (b)  $\alpha$ -, (c) PP-, and (d)  $\delta$ -cells from mSTZ treated versus healthy mice. Red line indicates thresholds used on significance level and gene expression change. Significantly regulated genes are highlighted in black. Genes significantly regulated in only one cell type but not the others are highlighted in blue. p-values were correct for multiple testing using BH. Cells were pooled from no STZ ( $n=3$ ) and mSTZ-vehicle ( $n=3$ ) treated mice. (e, f) GO term and KEGG pathway enrichment analysis of up- ( $\log(\text{fold change}) > 0.25$ ) and downregulated ( $\log(\text{fold change}) < -0.25$ ) genes in (e)  $\alpha$ - and (f)  $\delta$  cells in mSTZ treated versus healthy mice. We used limma-trend to find differentially expressed genes (M&M). Gene enrichment was done with EnrichR using Fisher's exact test to identify regulated ontologies/pathways (M&M). Cells were pooled from no STZ ( $n=3$ ) and mSTZ-vehicle ( $n=3$ ) treated mice. (g, h) Comparison between dysregulated genes in mSTZ- $\beta$  cells in mice with (g) data from RNA-seq of human T2D pancreata and (h) from scRNA-seq of human T1D  $\beta$  cells. Gene names of overlapping genes and identified dedifferentiation markers in Fig. 4e) are listed. i, Violin plots showing the distribution of the expression of endocrine developmental genes in beta cells of mSTZ treated and healthy mice. Violin shows the distribution as a kernel density estimate fit. Points in violin interior show individual data points. Boxplot in violin interior shows median, quartile and whisker values. Cells were pooled from no STZ ( $n=3$ ) and mSTZ-vehicle ( $n=3$ ) treated mice.

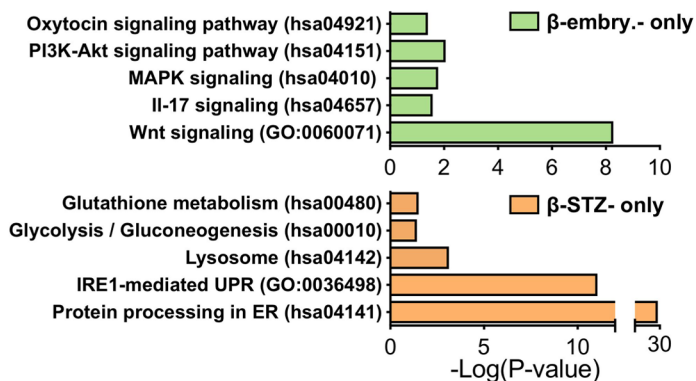
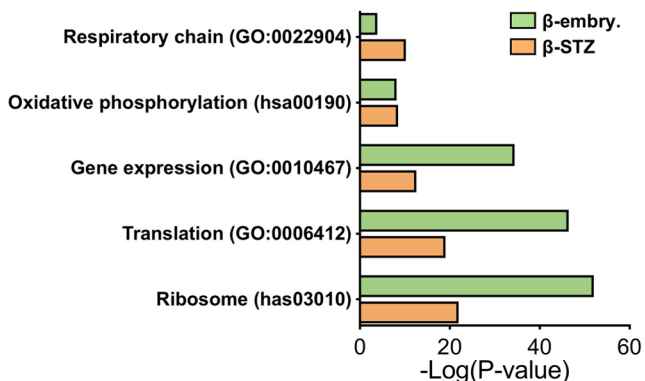
## a Regulated pathways of downregulated genes

### Common in $\beta$ -embryonic (P17.5/P0) and $\beta$ -STZ

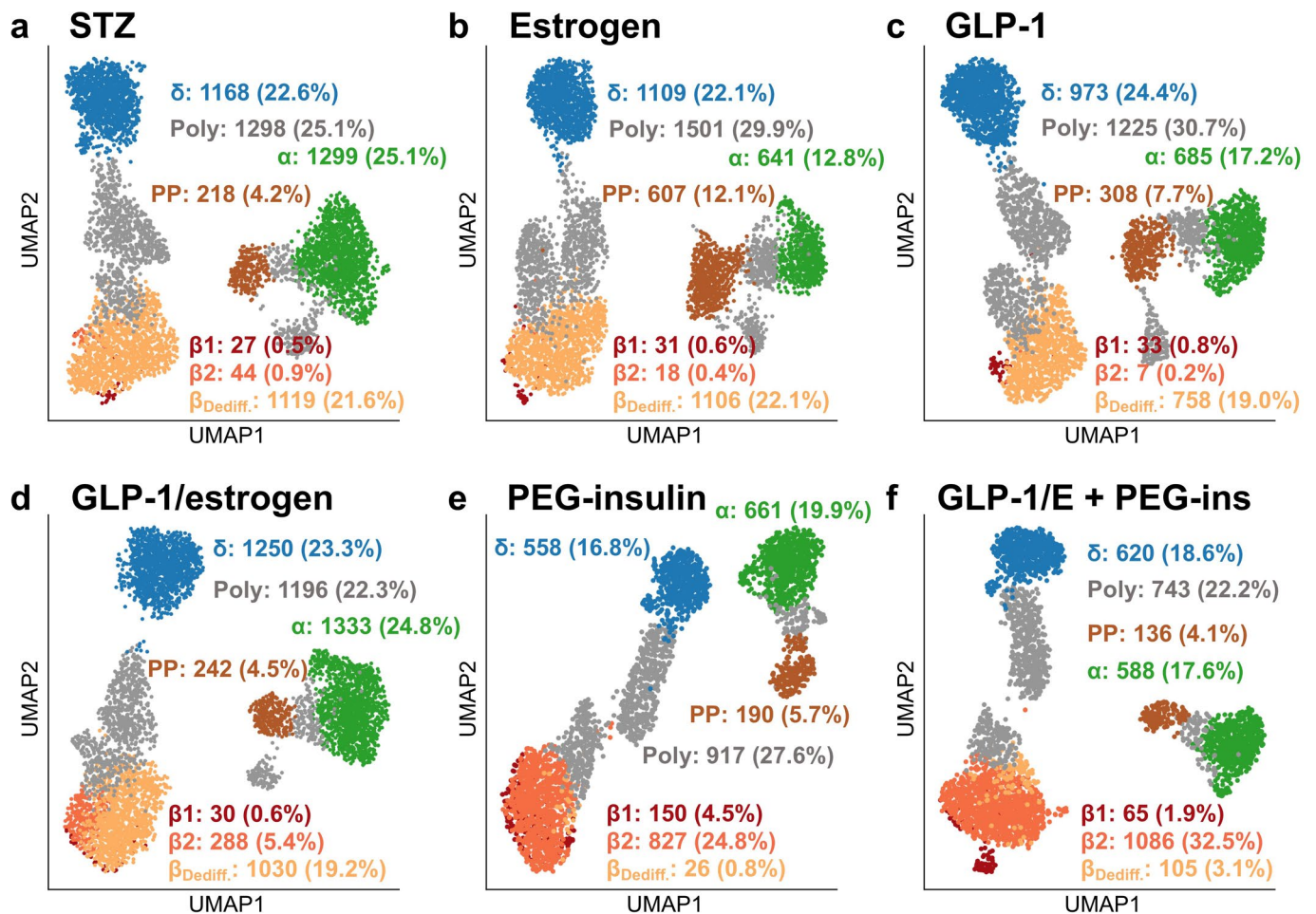


## b Regulated pathways of upregulated genes

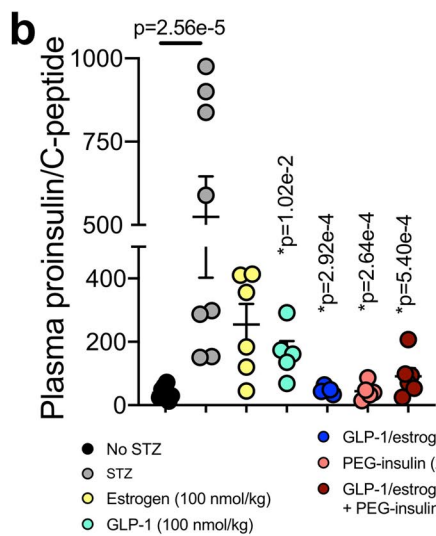
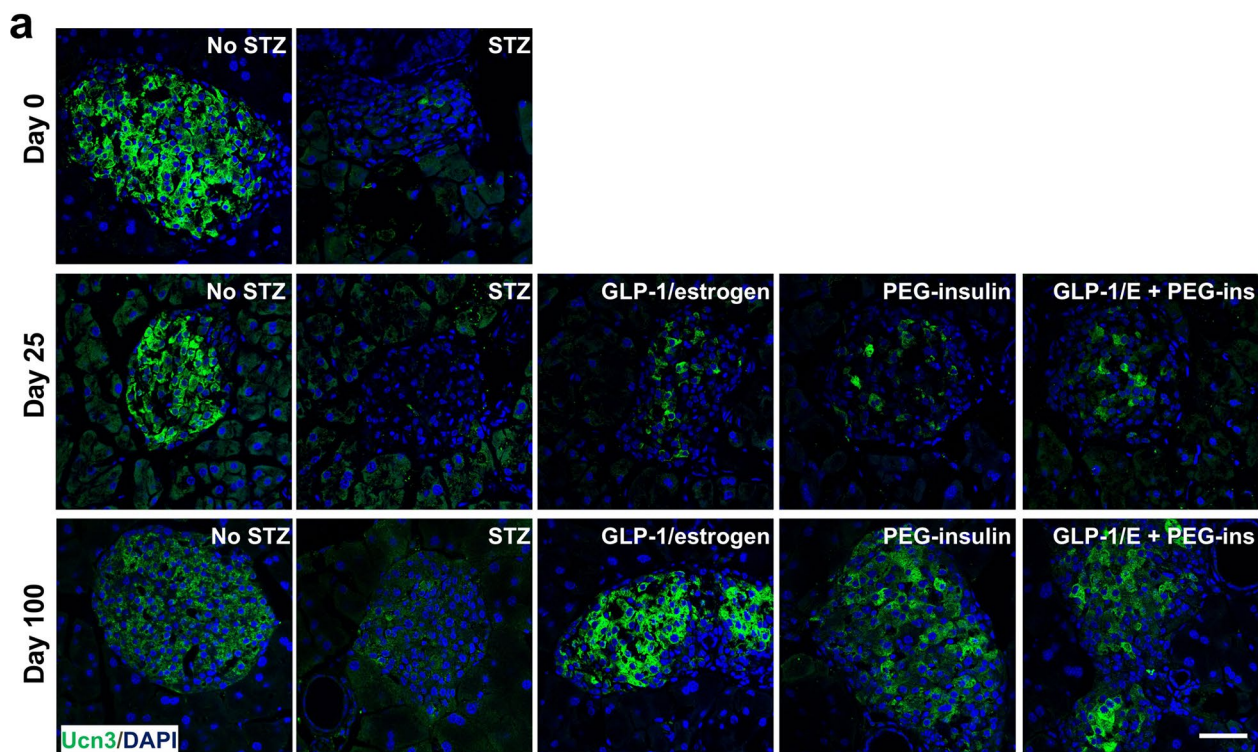
### Common in $\beta$ -embryonic (P17.5/P0) and $\beta$ -STZ



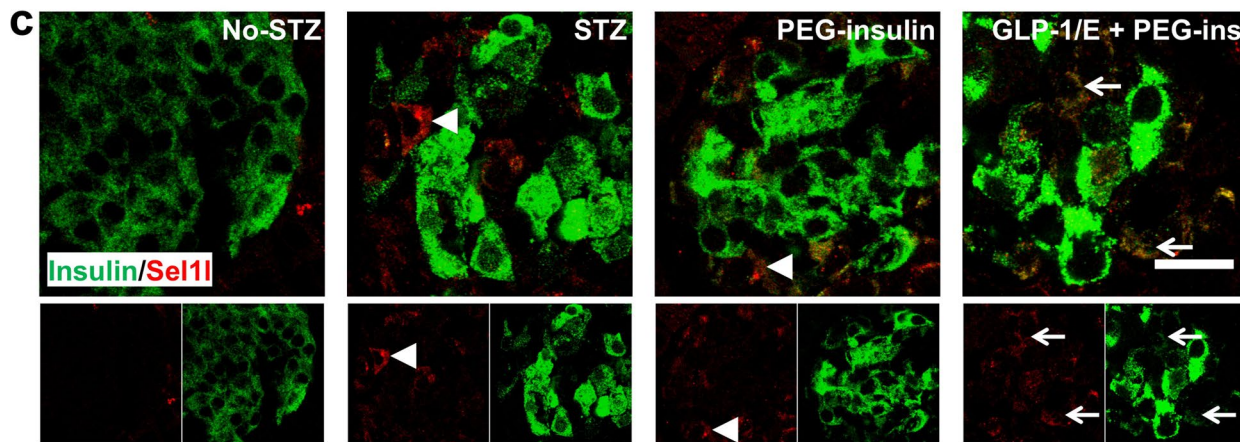
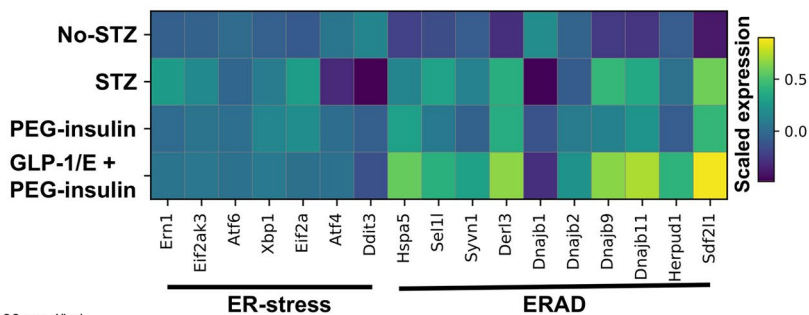
**Extended Data Fig. 8 | Common and distinct pathways of embryonic and dedifferentiated  $\beta$  cells.** Gene ontologies (Pvalue < 0.0001) and KEGG pathways (Pvalue < 0.05) that are commonly and specifically (a) down- and (b) upregulated in embryonic and mSTZ-derived  $\beta$ -cells. Representative terms from Supplementary Table 4 are depicted. Gene enrichment was done with EnrichR using Fisher's exact test to identify regulated ontologies/pathways (M&M). Cells were pooled from mSTZ-vehicle (n=3) treated mice.



**Extended Data Fig. 9 | Effects on endocrine cells of different treatments.** UMAP plot of all endocrine cells after 100 days of treatment showing endocrine cell distribution in each individual treatment. Total cell number for (a) mSTZ diabetic mice 5001, for (b) oestrogen treated mice 4889, for (c) GLP-1 treated mice 3874, for (d) GLP-1/oestrogen treated mice 5201, for (e) PEG-insulin treated mice 3217, and for (f) GLP-1/oestrogen (GLP-1/E) and PEG-insulin (PEG-ins) co-treated mice 3276. Values indicate the proportions of each cell cluster. Cells of  $n=3$  mice for each treatment were pooled.



**d** *scRNA-seq of  $\beta$ -cells at study end*



Extended Data Fig. 10 | See next page for caption.

**Extended Data Fig. 10 |  $\beta$ -cell maturation after compound treatment.** **a**, Immunohistochemical analysis of Ucn3 expression during the course of the study. Scale bar, 50  $\mu$ m. Day 0: Images are representative of no STZ-vehicle (n = 3) and mSTZ-vehicle (n = 3) treated mice. Day 25: Images are representative of no STZ-vehicle (n = 3), mSTZ-vehicle (n = 3), GLP-1/oestrogen (n = 3), PEG-insulin (n = 3), and GLP-1/oestrogen and PEG-insulin (n = 3) co-treated mice. Day 100: Images are representative of no STZ-vehicle (n = 3), mSTZ-vehicle (n = 3), GLP-1/oestrogen (n = 2), PEG-insulin (n = 2), and GLP-1/oestrogen and PEG-insulin (n = 3) co-treated mice. **b**, Plasma proinsulin/C-peptide ration in the end of the study (no STZ-vehicle, n = 8; mSTZ-vehicle, n = 8; oestrogen, n = 6; GLP-1, n = 5; GLP-1/oestrogen, n = 6; PEG-insulin, n = 6, GLP-1/oestrogen and PEG-insulin, n = 6; one-way ANOVA with Tukey post-hoc:  $F(6, 36) = 8.12$ ). Data are mean  $\pm$  SEM. **c**, Representative staining for insulin and Sel1l after 25 days of treatment. Arrow indicates Sel1l + insulin + -cells, which were especially found in GLP-1/oestrogen and PEG-insulin co-treated mice. Sel1l + insulin-cells (arrow head) were more common in mSTZ-diabetic and PEG-insulin treated mice. Images are representative of no STZ-vehicle (n = 3), mSTZ-vehicle (n = 3), PEG-insulin (n = 3), and GLP-1/oestrogen + PEG-insulin (n = 3) co-treated mice. Scale bar, 20  $\mu$ m. **d**, Expression of selected ER stress and ERAD-associated genes by scRNA-seq at study end.



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

Microsoft office excel 2018  
qPCR was carried out using the Viia7 Real Time PCR System (Thermo Fisher Scientific).  
All histological images were obtained with a Leica microscope of the type DMI 6000 using the LAS AF software. Images were analyzed using the LAS AF and/or ImageJ software program.  
Automatic image analyses were done using the commercially available image analysis software Definiens Developer XD 2 (Definiens AG, Munich, Germany).

#### Data analysis

Prism software 8 (Graphpad) <http://www.graphpad.com/scientific-software/> N/A prism/.

All histological images were obtained with a Leica microscope of the type DMI 6000 using the LAS AF software. Images were analyzed using the LAS AF and/or ImageJ software program.

Automatic image analyses were done using the commercially available image analysis software Definiens Developer XD 2 (Definiens AG, Munich, Germany).

R version 3.5.2 (2018-12-20)

scRNAseq data was analyzed using python 3.5 and 3.6. The raw reads were processed using the CellRanger analysis pipeline (v2.0.0) provided by 10X Genomics. Subsequent analyzes were carried out using custom scripts, Scanpy (<https://github.com/theislab/scanpy>, v1.0.4+92.g9a754bb), scVelo (v0.1.16.dev13+c1a6dad, <https://github.com/theislab/scvelo> with scanpy v1.3.2), velocity (v0.17.7, <http://velocityto.org>), Scrublet (v0.1, <https://github.com/AllonKleinLab/scrublet>), limma (<http://bioinf.wehi.edu.au/limma/>, v3.28.10) via an rpy2 interface (v2.9.1) and EnrichR ([amp.pharm.mssm.edu/Enrichr/](http://amp.pharm.mssm.edu/Enrichr/)).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and processed single-cell RNA sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE128565 (password on request). Otherwise, all data generated or analysed during this study are included in this published article (and its supplementary information files).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for in vivo analysis (mSTZ study) was calculated based on $\alpha$ -error=0.05, power=0.8, and effect size=1.58. We obtained a sample size of $n \geq 8$ /treatment group for physiological end points in the end of the treatment (e.g. blood glucose, C-peptide). For single-cell RNA-seq a minimum expected requirement of ~3000 cells per experiment was used and by far exceeded in most cases for confident identification of rare cell subpopulations (~0.05% of cells). No statistical test or power analyzes were used to pre-determine sample size. Otherwise, we chose the sample size of individual experiments based on past experience on detecting differences with a given method.
Data exclusions	For the single-cell RNAseq data cells were filtered using previously described standards for quality control. For each experiment the count matrix was filtered as follows: genes with expression in less than 10 cells were removed; cells with a fraction of UMI counts from mitochondrially encoded genes of 40% or more, with more than 100'000 UMI counts or with more than 7000 genes expressed were excluded. These criteria were based on the cell quality within this study. Otherwise, no data were excluded.
Replication	All attempts of in vivo replication were successful. For single-cell RNAseq experiments cells from three independently treated mice that showed similar responses in assessed physiological parameters were pooled to account for inter-individual variation.
Randomization	Diabetic mice were randomized to different treatment groups by 4h-fasting blood glucose levels. Rats were randomized by body weight.
Blinding	The investigators were not blinded to group allocation during in vivo experiments to be able to assess the health status of diabetic mice.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used

Guinea pig polyclonal anti-insulin, Thermo Scientific PA1-26938; 1:300  
Goat polyclonal anti-Glut2 Abcam ab111117; 1:500  
Goat polyclonal anti-Nkx6.1 R&D systems AF5857; 1:200  
Goat polyclonal anti-somatostatin Santa Cruz sc-7819; 1:500

Rat monoclonal anti-somatostatin Invitrogen MA5-16987; 1:300  
 Rabbit anti-Ngn3 Donated by H. Edlund N/A; 1:800  
 Rabbit polyclonal anti-urocortin 3 Phoenix Pharmaceuticals H-019-29; 1:300  
 Rabbit monoclonal anti-insulin Cell Signaling 3014; 1:300  
 Guinea pig polyclonal anti-glucacon Takara M182; 1:500  
 Guinea pig polyclonal anti-insulin ABD Serotec 5330-0104G; 1:300  
 Rabbit polyclonal cleaved caspase-3 (Asp 175) Cell Signaling 9661; 1:300  
 Rabbit polyclonal anti-Aldh1a3 Novus NBP2-15339; 1:300  
 Rabbit monoclonal anti-GLP-1R Novo Nordisk N/A; 1µg/ml  
 Rabbit polyclonal anti-gastrin Abcam Ab16035; 1:100  
 Rabbit polyclonal anti-cholecystokinin ENZO Life Sciences BML-CA1124; 1:100  
 Goat polyclonal anti-Sel1l Novus Biologicals NB100-93463; 1:300  
 Donkey anti-Goat IgG (H+L) secondary antibody, Alexa Fluor 633 Invitrogen A-21082; 1:800  
 Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 555 Invitrogen A-31572; 1:800  
 Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 488 Invitrogen A-21206; 1:800  
 Donkey anti-Guinea pig IgG (H+L) secondary antibody, DyLight 649 Dianova 706-495-148; 1:800  
 Donkey anti-Rat IgG (H+L) secondary antibody, DyLight 647 Dianova 711-605-152; 1:800  
 Donkey anti-Rat IgG (H+L) secondary antibody, Cy3 Dianova 712-165-153; 1:800  
 Donkey anti-Guinea pig (H+L) secondary antibody, Alexa Fluor 488 Dianova 706-545-148; 1:800

## Validation

Guinea pig polyclonal anti-insulin, <https://www.thermofisher.com/antibody/product/Insulin-Antibody-Polyclonal/PA1-26938>  
 Goat polyclonal anti-Glut2 <https://www.abcam.com/glucose-transporter-glut2-antibody-ab111117.html>  
 Goat polyclonal anti-Nkx6.1 [https://www.rndsystems.com/products/human-mouse-nkx61-antibody\\_af5857](https://www.rndsystems.com/products/human-mouse-nkx61-antibody_af5857)  
 Goat polyclonal anti-somatostatin <https://www.scbt.com/scbt/de/product/somatostatin-antibody-d-20>  
 Rat monoclonal anti-somatostatin <https://www.thermofisher.com/antibody/product/Somatostatin-Antibody-clone-YC7-Monoclonal/MA5-16987>  
 Rabbit anti-Ngn3 We used embryonic mouse pancreata to ensure binding of Ngn3 included in this study. We used no 2nd as control stainings to show specificity.  
 Rabbit polyclonal anti-urocortin 3 <https://www.phoenixpeptide.com/products/view/Antibodies/H-019-29>  
 Rabbit monoclonal anti-insulin <https://www.cellsignal.de/products/primary-antibodies/insulin-c27c9-rabbit-mab/3014>  
 Guinea pig polyclonal anti-glucacon <https://www.takarabio.com/products/antibodies-and-elisa/primary-antibodies-and-elisas-by-research-area/metabolic-diseases/glucacon?catalog=M182>  
 Guinea pig polyclonal anti-insulin <https://www.bio-rad-antibodies.com/polyclonal/pig-porcine-insulin-antibody-5330-0054.html?f=purified>  
 Rabbit polyclonal cleaved caspase-3 (Asp 175) <https://www.cellsignal.de/products/primary-antibodies/cleaved-caspase-3-asp175-antibody/9661>  
 Goat polyclonal anti-Sel1l [https://www.novusbio.com/products/sel1l-antibody\\_nb100-93463](https://www.novusbio.com/products/sel1l-antibody_nb100-93463)  
 Rabbit polyclonal anti-Aldh1a3 [https://www.novusbio.com/products/aldh1a3-antibody\\_nbp2-15339](https://www.novusbio.com/products/aldh1a3-antibody_nbp2-15339)  
 Rabbit monoclonal anti-GLP-1R This is a antibody produced and checked for specificity by Novo Nordisk. We demonstrate binding of this antibody exclusively to beta cells in the islet of Langerhans by using a GLP-1R knock out mouse.  
 Rabbit polyclonal anti-gastrin <https://www.abcam.com/gastrin-antibody-ab16035.html>  
 Rabbit polyclonal anti-cholecystokinin <http://www.enzolifesciences.com/BML-CA1124/cholecystokinin-39-polyclonal-antibody/>  
 Donkey anti-Goat IgG (H+L) secondary antibody, Alexa Fluor 633 <https://www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21082>  
 Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 555 <https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-31572>  
 Donkey anti-Rabbit IgG (H+L) secondary antibody, Alexa Fluor 488 <https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21206>  
 Donkey anti-Guinea pig IgG (H+L) secondary antibody, DyLight 649 <https://www.dianova.com/produkte/706-605-148-esel-igg-anti-meerschweinchen-igg-hl-alexa-fluor-647-minx-bockgohshohumsrbrtsh-500-µg/>  
 Donkey anti-Rat IgG (H+L) secondary antibody, DyLight 647 <https://www.dianova.com/produkte/711-605-152-esel-igg-anti-kaninchen-igg-hl-alexa-fluor-647-minx-bockgoghphshohumsrtsh-500-µg/>  
 Donkey anti-Rat IgG (H+L) secondary antibody, Cy3 <https://www.dianova.com/produkte/712-165-153-esel-igg-anti-ratte-igg-hl-cy3-minx-bockgoghphshohumsrbsh-500-µg/>  
 Donkey anti-Guinea pig (H+L) secondary antibody, Alexa Fluor 488 <https://www.dianova.com/produkte/706-545-148-esel-igg-anti-meerschweinchen-igg-hl-alexa-fluor-488-minx-bockgohshohumsrbrtsh-500-µg/>

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

### Laboratory animals

8-week old male C57BLJ/6 mice (Janvier Labs);  
 male GLP-1R KO mouse, own breeding, C57BLJ/6 background, age=4 months;  
 8-week old male FVFPBFDHOM mice (Bastidas-Ponce, A. et al. Foxa2 and Pdx1 cooperatively regulate postnatal maturation of pancreatic  $\beta$ -cells. *Mol. Metab.* (2017). doi:10.1016/j.molmet.2017.03.007);  
 40 ovariectomized female Sprague-Dawley rats (Charles River UK, Ltd) were used, rats were 9-12 weeks of age.

### Wild animals

The study did not involve wild animals

### Field-collected samples

The study did not involve samples collected from the wild

### Ethics oversight

Government of Upper Bavaria, Germany.

## Ethics oversight

Rat study was further designed according to accepted pharmacological principles and followed Good Laboratory Practice conducted by Envigo CRS Limited, UK.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

We used isolated islets from FVFPBF mice for FACS analysis.

Islet isolation was performed by collagenase P (Roche) digestion of the adult pancreas. Briefly, 3 mL of collagenase P (1 mg/mL) was injected into the bile duct and the perfused pancreas was consequently dissected and placed into another 3 mL collagenase P for 15 min at 37 °C. 10 mL of G-solution (HBSS (Lonza) + 1% BSA (Sigma)) was added to the samples followed by centrifugation at 1600 rpm at 4 °C. After another washing step with G-solution, the pellets were re-suspended in 5.5 mL of gradient preparation (5 mL 10% RPMI (Lonza) + 3 mL 40% Optiprep (Sigma)/ per sample), and placed on top of 2.5 mL of the same solution. To form a 3-layers gradient, 6 mL of G-solution was added on the top. Samples were then incubated for 10 min at RT before subjecting to centrifugation at 1700 rpm. Finally, the interphase between the upper and the middle layers of the gradient was harvested and filtered through a 70 µm Nylon filter and washed with G-solution. Islets were handpicked under the microscope.

In order to achieve a single cell suspension of islets, islets were handpicked in an 1.5 mL Eppendorf tube, pelleted (800 rpm, 1 min) washed with PBS (-Mg/Ca, Gibco) and digested with 0.25% Trypsin with EDTA (Gibco) at 37°C for 8 min. Mechanical disaggregation every 2-3 min was required. After, the digestive reaction was stopped and cells were pelleted (1200 rpm, 5 min).

#### Instrument

Single cells were analyzed by FACS-Aria III (BD).

#### Software

Data were analyzed by FlowJo.

#### Cell population abundance

Endocrine cell populations are reported in this manuscript.

#### Gating strategy

Single cells were gated according to their FSC-A (front scatter area) and SSC-A (side scatter area). Singlets were gated dependent on the FSC-W (front scatter width) and FSC-H (front scatter height) and dead cells were excluded using the marker 7AAD (eBioscience). The FVF endocrine populations were discriminated upon their Venus fluorescence emission at 488 nm and the  $\beta$ - and  $\alpha$ -lineages according to their BFP emission at 405 nm (positive and negative respectively). To enrich for  $\beta$ -cells the distinct SSC-A high populations were gated.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

## Appendix C

This is the published version of the article in Nature Metabolism following peer review. The article is open access thus the published version is inserted here. This article is published in Nature Metabolism under a Creative Commons Attribution license (CC-BY 4.0).

Alexandra Aliluev\*, **Sophie Tritschler\***, Michael Sterr, Lena Oppenländer, Julia Hinterdobler, Tobias Greisle, Martin Irmeler, Johannes Beckers, Na Sun, Axel Walch, Kerstin Stemmer, Alida Kindt, Jan Krumsiek, Matthias H Tschöp, Malte D Luecken, Fabian J Theis, Heiko Lickert, Anika Böttcher, **Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice.** *Nature Metabolism*, 3, 1202–1216 (2021)



OPEN

# Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice

Alexandra Aliluev<sup>1,2,12</sup>, Sophie Triteschler<sup>1,2,3,4,12</sup>, Michael Sterr<sup>1,2</sup>, Lena Oppenländer<sup>1,2</sup>, Julia Hinterdobler<sup>1</sup>, Tobias Greisle<sup>1,2</sup>, Martin Irmeler<sup>5</sup>, Johannes Beckers<sup>2,5,6</sup>, Na Sun<sup>7</sup>, Axel Walch<sup>7</sup>, Kerstin Stemmer<sup>8,9</sup>, Alida Kindt<sup>3</sup>, Jan Krumsiek<sup>3</sup>, Matthias H. Tschöp<sup>2,8,10</sup>, Malte D. Luecken<sup>3</sup>, Fabian J. Theis<sup>3,11,13</sup>✉, Heiko Lickert<sup>1,2,11,13</sup>✉ and Anika Böttcher<sup>1,2,13</sup>✉

**Excess nutrient uptake and altered hormone secretion in the gut contribute to a systemic energy imbalance, which causes obesity and an increased risk of type 2 diabetes and colorectal cancer. This functional maladaptation is thought to emerge at the level of the intestinal stem cells (ISCs). However, it is not clear how an obesogenic diet affects ISC identity and fate. Here we show that an obesogenic diet induces ISC and progenitor hyperproliferation, enhances ISC differentiation and cell turnover and changes the regional identities of ISCs and enterocytes in mice. Single-cell resolution of the enteroendocrine lineage reveals an increase in progenitors and peptidergic enteroendocrine cell types and a decrease in serotonergic enteroendocrine cell types. Mechanistically, we link increased fatty acid synthesis, Ppar signaling and the *Insr-Igf1r-Akt* pathway to mucosal changes. This study describes molecular mechanisms of diet-induced intestinal maladaptation that promote obesity and therefore underlie the pathogenesis of the metabolic syndrome and associated complications.**

Diet-induced obesity is a serious public health and economic problem. Obese people are at higher risk of developing type 2 diabetes (T2D), cardiovascular diseases and cancer, all leading causes of death worldwide (<https://www.who.int>). Today, bariatric surgery is the most effective treatment to achieve long-term weight loss and notably leads to diabetes remission<sup>1</sup>. Surgical procedures cause profound changes in secretion of gut hormones with beneficial effects on whole body metabolism, appetite and food intake<sup>2</sup>. These findings suggest gut hormones as candidates for new therapies against obesity and diabetes<sup>3</sup>.

The gut, as the body's digestive and largest endocrine system, serves as a central regulator of energy and glucose homeostasis and quickly responds to dietary and nutritional changes<sup>4–8</sup>. Constant overnutrition is thought to lead to intestinal maladaptation and dysfunction and to contribute to the development of obesity and prediabetes<sup>8</sup>. This is evident as two hallmarks of obesity, excessive food intake and a reduced stimulation of postprandial insulin secretion by gut hormones, are linked to impaired gut function<sup>9</sup>. Moreover, differences in gut morphology and physiology have been observed between lean and obese individuals<sup>8,10,11</sup>. Tackling gut dysfunction at an early stage of disease might therefore be a promising treatment option to fight obesity and the resulting risks and complications.

Intestinal functions are carried out by specialized epithelial cells lining the gut: absorptive enterocytes, antimicrobial-peptide-secreting Paneth cells, hormone-secreting enteroendocrine cells (EECs), mucus-secreting goblet and chemosensory tuft cells. The cells of the intestinal epithelium are constantly generated from ISCs<sup>12</sup>.

ISC identity is defined by multi-lineage potential and self-renewal capacity, but also by properties that are not hard-wired, such as the proliferative, epigenetic and metabolic state<sup>12,13</sup>. High cell turnover and cellular plasticity contribute to the natural adaptive capacity of the gut, but the mechanisms underlying maladaptation in response to an obesogenic diet are still unclear. Specifically, we do not know whether the hormonal imbalance and increased absorptive capacity emerge at the level of early lineage commitment from ISCs<sup>9,14</sup>. Moreover, the functions of enterocytes and EECs differ between gut regions<sup>15–17</sup>. Proximal enterocytes, for instance, are specialized to absorb iron and nutrients (carbohydrates, fat and protein), whereas distal enterocytes absorb bile acids and vitamin B12. Proximal EEC types secrete serotonin and ghrelin, whereas distally located EECs preferentially secrete Glp-1 (ref. 17). Gut functions are also spatially compartmentalized along the crypt–villus axis. Enterocytes shift their expression profile from an antimicrobial to a nutrient absorption to an immunomodulatory program while migrating from the bottom of the villus to its tip. The spatial–functional compartmentalization of nutrient absorption is achieved through zoned expression of nutrient transporters, with the highest expression of carbohydrate and amino acid transporters in the mid-villus region and of apolipoproteins and fatty acid transporters, such as *Apoa4* and *Fabp1*, at the villus tip<sup>18</sup>. Similarly, enteroendocrine cell types switch their hormone expression pattern along the crypt–villus axis<sup>19</sup>. Regional identities are thought to be determined at the level of the ISCs by epigenetic mechanisms, and the crypt–villus compartmentalization is at least partly established by growth-factor

<sup>1</sup>Institute of Diabetes and Regeneration Research, Helmholtz Diabetes Center, Helmholtz Center Munich, Neuherberg, Germany. <sup>2</sup>German Center for Diabetes Research (DZD), Neuherberg, Germany. <sup>3</sup>Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany. <sup>4</sup>School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. <sup>5</sup>Institute of Experimental Genetics, Helmholtz Center Munich, Neuherberg, Germany. <sup>6</sup>Technical University of Munich, Freising, Germany. <sup>7</sup>Research Unit of Analytical Pathology, Helmholtz Center Munich, Neuherberg, Germany. <sup>8</sup>Institute of Diabetes and Obesity, Helmholtz Diabetes Center, Helmholtz Center Munich, Neuherberg, Germany. <sup>9</sup>Rudolf-Buchheim-Institute of Pharmacology, Justus Liebig University, Giessen, Germany. <sup>10</sup>Division of Metabolic Diseases, Department of Medicine, Technical University of Munich, Munich, Germany. <sup>11</sup>Technical University of Munich, Munich, Germany. <sup>12</sup>These authors contributed equally: Alexandra Aliluev, Sophie Triteschler. <sup>13</sup>These authors jointly supervised this work: Fabian J. Theis, Heiko Lickert, Anika Böttcher. ✉e-mail: [fabian.theis@helmholtz-muenchen.de](mailto:fabian.theis@helmholtz-muenchen.de); [heiko.lickert@helmholtz-muenchen.de](mailto:heiko.lickert@helmholtz-muenchen.de); [anika.boettcher@helmholtz-muenchen.de](mailto:anika.boettcher@helmholtz-muenchen.de)

gradients, but it is unknown whether and how the compartmentalization of gut functions is affected by an obesogenic diet<sup>20,21</sup>.

Developing pharmacological approaches to counteract obesity and diabetes require in-depth understanding of the mechanisms that underlie maladaptation and endocrine dysfunction in the gut, specifically during the transition from a healthy to an obese and to a prediabetic state. In this study we combined single-cell profiling with genetic lineage labelling and tracing of ISC fate decisions and in situ metabolomics to elucidate the cellular and molecular mechanisms that underlie intestinal maladaptation to an obesogenic western-style high-fat/high-sugar diet (HFHSD) in mice.

## Results

### HFHSD remodels the intestinal mucosa and leads to obesity.

To study mechanisms of intestinal maladaptation to an HFHSD, we maintained male *Foxa2*-Venus fusion (FVF) reporter mice on a diet regimen for 12 weeks<sup>22</sup>. Over this time, mice on an HFHSD gained significantly more body weight (fat and lean mass) than did control diet (CD)-fed mice, and this was accompanied by an increase in length and weight of the small intestine (SI), increased villus length and decreased crypt density (Extended Data Fig. 1a–l). Crypt depth, number of cells per crypt and cell sizes in crypts and villi were not changed (Extended Data Fig. 1k,m–p). Histological assessment of the SI mucosa showed cellular fat inclusions in the villi of HFHSD-fed mice, suggesting disturbances in fatty acid metabolism in the gut epithelium (Extended Data Fig. 1q). Metabolic assessment showed that our diet-induced obese mice developed prediabetes, which was characterized by fasting hyperglycaemia, impaired glucose tolerance and pronounced hyperinsulinaemia as well as insulin resistance (Extended Data Fig. 1r–u). Thus, an HFHSD changes mucosal morphology, which is indicative of altered ISC homeostasis and lineage recruitment.

### HFHSD alters ISC lineage allocation and regional identity.

To dissect the metabolic impact of an HFHSD on ISC lineage recruitment, we employed single-cell RNA-sequencing (scRNA-seq) of crypt cells from the SI of CD- and HFHSD-fed FVF reporter mice (Fig. 1a). FVF-lineage labelling enabled us to flow sort and enrich for ISCs and EECs, cell types that together usually make up less than 6% of the intestinal epithelial cells (Fig. 1a and Extended Data Fig. 2a)<sup>17</sup>. This enrichment strategy not only allowed us to assess compositional changes on an HFHSD, but also to molecularly and functionally characterize rare crypt-cell types. We profiled 27,687 cells obtained from three biological replicates of CD- and HFHSD-fed mice and detected on average 3,500 genes per cell (Extended Data Fig. 2b). Unsupervised graph-based clustering and annotation based on known marker genes revealed all mature intestinal cell lineages as well as ISCs and distinct progenitor states for each intestinal lineage (Fig. 1b,c and Extended Data Fig. 2c,d). Lineage marker gene expression as well as the expression of known

regulators that drive intestinal lineage decisions toward the absorptive (for example *Notch1*) and secretory lineages (*Dll1*, *Dll4* and *Atoh1*) and PAGA topology<sup>23</sup> were unchanged between diet conditions (Fig. 1c, Extended Data Fig. 2c–f and Supplementary Table 1). Corresponding subtypes showed high correlation in their transcriptomes, indicating that an HFHSD did not alter lineage identities (Extended Data Fig. 2g). Instead, we observed a shift of cell densities indicating that an HFHSD impacts the composition of the mucosal epithelium (Fig. 1d,e and Extended Data Fig. 2d,h). Note that we excluded the Paneth cell lineage from the scRNA-seq analysis because of reported sampling issues, which might impact cell-type ratios<sup>17</sup>. To assess whether the HFHSD affects regional subtypes, we separated cell types into proximal and distal cells using recently described regional signature genes (Extended Data Fig. 3a)<sup>17</sup>. We found that the fractions of ISCs, enterocytes and goblet cells with proximal identity increased in mice fed an HFHSD (Fig. 1f,g and Extended Data Fig. 3b). Together, these data suggest that an HFHSD boosts formation of the absorptive and goblet cell lineage and promotes proximal cell identities, indicating that the intestine adapts cell-type composition and specific cellular functions to nutrient availability.

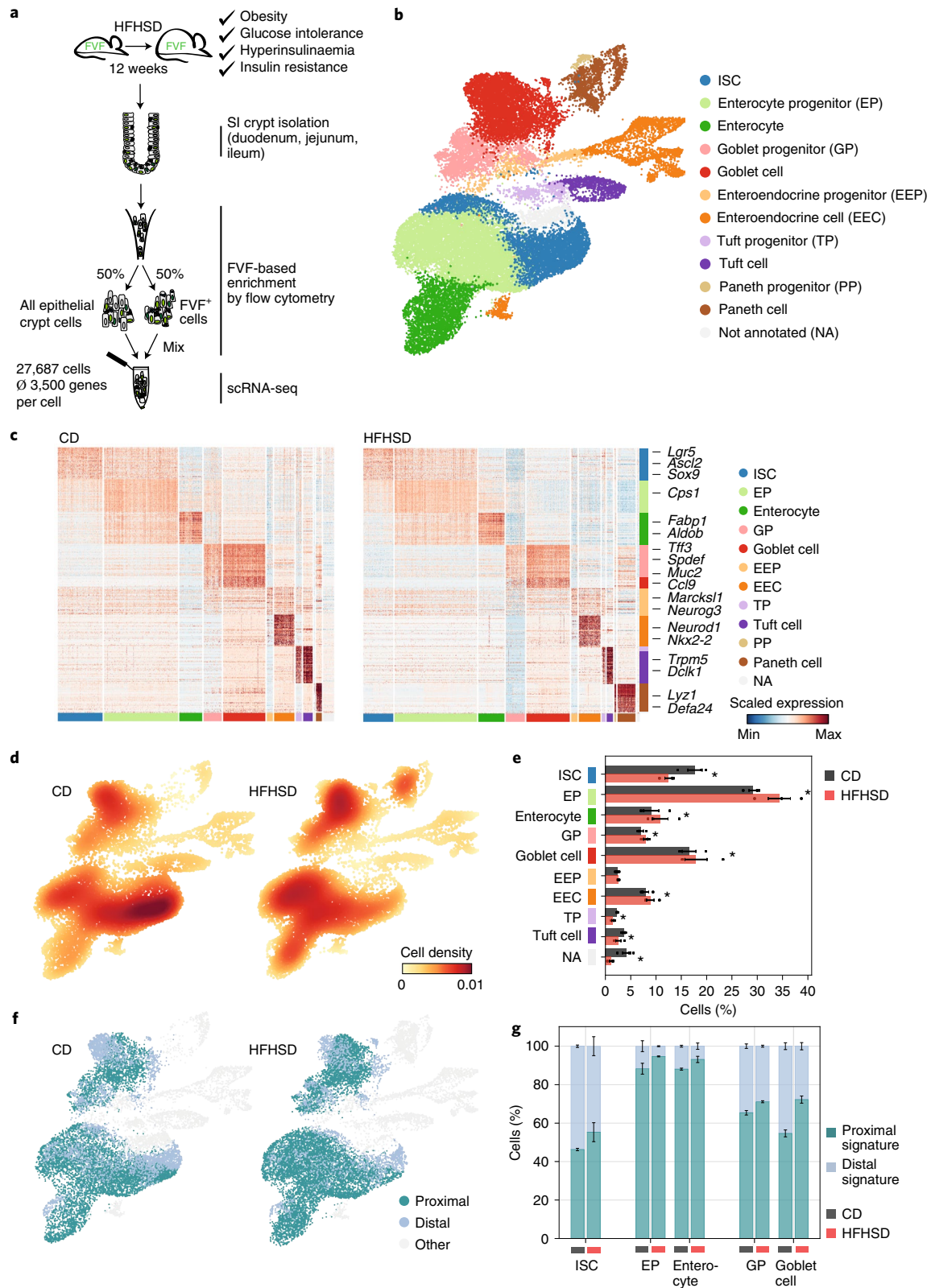
### Altered lineage relations in the crypt translate into the villus.

As our scRNA-seq experiment indicated alterations of lineage abundances in the crypt on HFHSD, we next performed short-term genetic lineage tracing of ISC fate decisions using a dual-fluorescent inducible Cre-reporter *Foxa2*<sup>2nEGFP-CreERT2/+;Gt(ROSA)26<sup>mTmG</sup>/+</sup> mouse (Fig. 2a)<sup>24,25</sup> to determine how these changes impact the composition of the mature villus compartment. *Foxa2* is expressed in quiescent and rapidly dividing ISCs (Supplementary Fig. 1). In this mouse model, *Foxa2* expression-driven CreERT2 induces a switch from membrane-Tomato (mT) to membrane-GFP (mG) in ISCs after tamoxifen administration. At 48 h after Cre-ERT2 activation by tamoxifen, we observed labelled mG-positive ISCs next to Paneth cells at the crypt bottom and labelled single cells and small cell clusters in duodenal crypts at similar frequencies in CD- and HFHSD-fed mice (Extended Data Fig. 4a). To determine the effect of an obesogenic diet on ISC lineage recruitment, we analysed the abundance of mature enterocytes, goblet cells and EECs in Cre-reporter labelled mG-positive lineage ribbons in the villi of CD- and HFHSD-fed mice 70 h after tamoxifen induction (Fig. 2b). In mice fed an HFHSD, longer ribbons (>6 cells) appeared more frequently, suggesting higher ISC or progenitor proliferation or increased cell turnover (Fig. 2c,d and Extended Data Fig. 4b). Within the lineage ribbons, total numbers of villin<sup>+</sup> enterocytes and Muc2<sup>+</sup> goblet cells were increased, whereas the number of ChgA<sup>+</sup> EECs was reduced (Fig. 2c,d). The relative abundances of enterocytes and ChgA<sup>+</sup> EECs within lineage ribbons were diminished owing to the large increase in goblet cell numbers in mice fed an HFHSD (Fig. 2c and Extended Data Fig. 4c). In accordance,

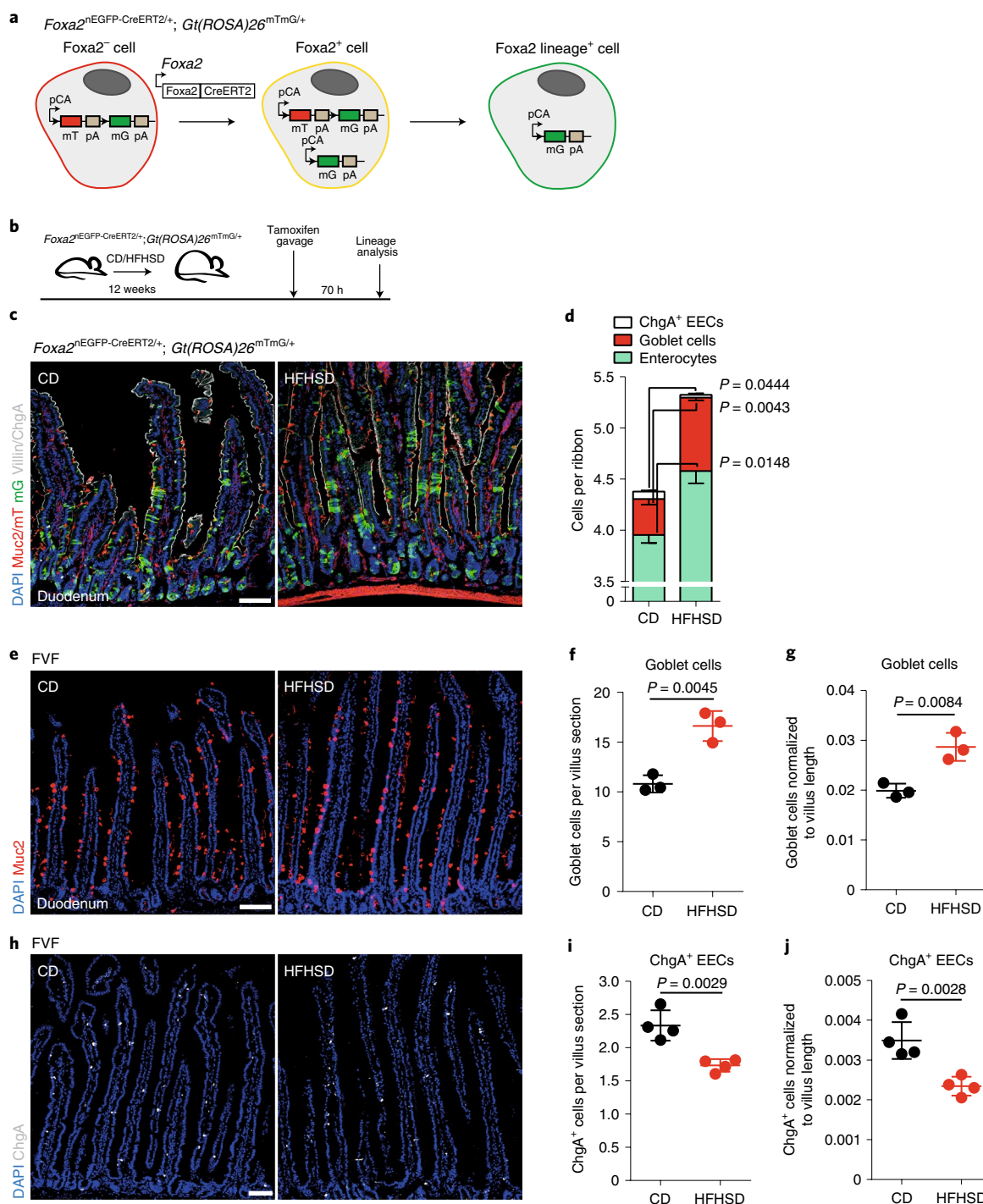
**Fig. 1 | HFHSD alters lineage allocation from ISCs and shifts the regional identity of cells.** **a**, Experimental design for FVF-based SI crypt-cell enrichment by flow cytometry and scRNA-seq. Single FVF<sup>+</sup> and whole crypt cells from SI crypts of CD- and HFHSD-fed FVF mice were isolated in equal numbers by flow cytometry and combined for each sample. FVF-enriched single-cell samples were then transcriptionally profiled by scRNA-seq. **b**, Uniform manifold approximation and projection (UMAP) plot of 27,687 profiled single SI crypt cells. Colours highlight clustering into major intestinal cell types based on the expression of previously published marker genes. One cluster of cells could not be assigned owing to a missing marker gene signature (NA). **c**, Heatmap depicting scaled expression of cell-type-specific gene signatures in CD- and HFHSD-derived cells. Cells are represented in columns and genes are represented in rows. Colour bars indicate cell types assigned to both cells and genes. Selected known marker genes for every lineage are indicated. **d,e**, Cell-type composition differences in CD- and HFHSD-derived single-cell samples visualized by cell density projected onto the two-dimensional UMAP embedding (**d**) and quantified as proportions over cell types (mean ± s.e.m. of biologically independent samples, *n* = 3 mice per group, Dirichlet multinomial model) (**e**). Densities were estimated using Gaussian kernels. \* Indicates a credible shift (95% highest posterior density interval of the frequency shift modelled by a Dirichlet multinomial model does not overlap 0). *n* = 3 CD mice, *n* = 3 HFHSD mice. **f**, UMAP coloured by the regional identities of cells in ISCs and goblet cell and enterocyte cell lineage. Subclusters were classified on the basis of expression of regional marker genes. **g**, Proportions of cells with proximal or distal identity in ISCs and enterocyte and goblet cell lineage (mean ± s.e.m. of biologically independent samples). *n* = 3 CD mice, *n* = 3 HFHSD mice.

goblet cell numbers in the duodenal villi were increased and numbers of ChgA<sup>+</sup> EECs were reduced in the duodenum and ileum of HFHSD-fed FVF mice (Fig. 2e–j and Extended Data Fig. 4d). Note that the number of ileal goblet cells was not altered, suggesting that gut regions are differently affected by an HFHSD (Extended Data

Fig. 4e). As we excluded Paneth cells from our scRNA-seq analysis, we determined the Paneth cell numbers in situ and found no difference between CD- and HFHSD-fed mice (Extended Data Fig. 4f,g). To understand whether the shift in regional identity in crypt cells translates to the villi and results in altered function, we performed







**Fig. 2 | Altered lineage relations in the crypt translate into the mature villus compartment.** **a**, *Foxa2*<sup>nEGFP-CreERT2/+</sup>; *Gt(Rosa)26*<sup>mTmG/+</sup> lineage-tracing model. mT *Foxa2*-negative cells (red) convert into mG *Foxa2*-lineage-positive cells (green) upon *Foxa2*-promoter-driven Cre expression via an intermediate (mTmG<sup>+</sup>, yellow) state. pCA, chicken  $\beta$ -actin core promoter with a CMV enhancer. **b**, Experimental scheme of short-term lineage tracing of *Foxa2*-positive cells using the *Foxa2*<sup>nEGFP-CreERT2/+</sup>; *Gt(Rosa)26*<sup>mTmG/+</sup> mouse model. **c,d**, Representative laser scanning microscopy (LSM) images of Cre-driven recombination in the duodenum of CD- and HFHSD-fed *Foxa2*<sup>nEGFP-CreERT2/+</sup>; *Gt(Rosa)26*<sup>mTmG/+</sup> (**c**) and analysis of lineage-positive cells (**d**) 70 h after tamoxifen induction. Single converted mG<sup>+</sup> cells and lineage ribbons (green) containing Muc2<sup>+</sup> goblet cells (red), ChgA<sup>+</sup> EECs (white) and villin<sup>+</sup> enterocytes (white) are observed in crypts and villi. Scale bar, 100  $\mu$ m. For *Foxa2*-lineage analysis, only confluent mG<sup>+</sup> cell patches located in the villi were considered ( $n = 3$  mice per group). Data are mean  $\pm$  s.e.m. Statistical significance was determined by two-tailed Student's *t*-test. DAPI, 4,6-diamidino-2-phenylindole. **e-j**, Abundances of mature intestinal cell types are altered in HFHSD-fed FVF mice. **e,g**, Representative LSM images (**e**) of goblet cells (Muc2<sup>+</sup>) and quantification thereof (**f,g**). Scale bar, 75  $\mu$ m,  $n = 3$  mice per group. **h-j**, Representative LSM images (**h**) of ChgA<sup>+</sup> EECs and quantification thereof (**i,j**). Scale bar, 75  $\mu$ m,  $n = 4$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's *t*-test.

scRNA-seq of villus cells from the SI (Extended Data Fig. 5a–c). We focused on enterocytes because regional and spatial compartmentalization of enterocyte function is well described and classified them into proximal and distal enterocytes using reported regional transcription factors (Extended Data Fig. 5d,e)<sup>17,18</sup>. Consistent with crypt scRNA-seq data, we also found an increased fraction of proximal-type villus enterocytes in mice fed an HFHSD (Extended Data Fig. 5f). Furthermore, proximal and distal-type enterocytes showed upregulation of functional proximal genes such as *Fabp1* and *Apoa4* (Extended Data Fig. 5g).

Next, we assessed whether enterocyte zonation is affected by an HFHSD. We inferred a pseudospacial ordering of cells using previously reported zonation markers and partitioned cells into five zones along the axis from villus bottom to tip (Extended Data Fig. 6a)<sup>18</sup>. We found that in proximal and distal enterocytes, the spatial expression pattern of several genes associated with carbohydrate and fatty acid absorption was altered in HFHSD-fed mice (Extended Data Fig. 6b)<sup>18</sup>. Analysis of *Fabp1* and *Apoa4* immunolocalization in the ileum and duodenum confirmed the proximalization of enterocytes and altered enterocyte zonation. The expression zones of *Fabp1* and *Apoa4* were enlarged on an HFHSD and reached from villus tip to villus bottom (Extended Data Fig. 6c–k).

Thus, more goblet cells and proximal-type enterocytes are generated under HFHSD conditions, which leads to morphological changes of the gut mucosa and an increase in fatty acid transport and absorption. Reduced ChgA<sup>+</sup> EEC numbers imply changes in specific EEC subsets.

### HFHSD changes the allocation of the enteroendocrine lineage.

Secreted gut hormones are critical regulators of food intake and systemic metabolism along the gut–brain–pancreas axis and hormonal imbalance is linked to the metabolic syndrome<sup>26,27</sup>. To elucidate the mechanisms underlying EEC dysfunction in response to an obesogenic diet, we first refined the clustering of our 2,865 EEC lineage cells into distinct subpopulations, which revealed five enteroendocrine progenitor (EEP) clusters characterized by expression of *Sox4*, *Ngn3*, *Arx/Isl1*, *Ghrl* and *Pax4*, respectively, and six polyhormonal EEC clusters: SILA cells (coexpress *Sct*, *Cck*, *Gcg*, *Ghrl* and *Gal*), SILP cells (coexpress *Sct*, *Cck*, *Gcg* and *Pyy*), SIK cells (coexpress *Sct*, *Cck* and *Gip*), SAKD cells (coexpress *Sct*, *Cck*, *Ghrl*, *Gip* and *Sst*), SIN cells (coexpress *Sct*, *Cck*, *Gcg* and *Nts*), enterchromaffin (EC) cells (coexpress *Sct*, *Tac1*, *Tph1* and *Reg4*<sup>+</sup> EC cells (coexpress *Sct*, *Tac1*, *Tph1*, *Ucn3* and *Reg4*) (Fig. 3a,b and Extended Data Fig. 7a,b)<sup>17,19,28</sup>. We also identified a population of heterogeneous EECs coexpressing endocrine and ISC markers that we termed *Lgr5*<sup>+</sup> EECs and which are reminiscent of *Lgr5*<sup>+</sup> label-retaining cells (Fig. 3a,b and Extended Data Fig. 7a)<sup>29</sup>. Notably, we found that *Lgr5*<sup>+</sup> EECs are characterized by active Bmp signalling, which has been shown to regulate the hormonal plasticity of EECs (Extended Data Fig. 7c)<sup>19</sup>. To determine EEC lineage hierarchy and to understand the relationship of EEP and mature clusters,

we used PAGA in combination with RNA velocity (Methods)<sup>30</sup>. The abstracted graph represents possible differentiation paths that cells follow and RNA velocity predicts the future state of a cell based on gene expression state (gene induction or repression), thus indicating the direction of differentiation. We observed two main differentiation trajectories from the *Sox4*<sup>+</sup> progenitors to mature subtypes: (1) a path via *Arx/Isl1*<sup>+</sup> and *Ghrl*<sup>+</sup> progenitors to peptidergic EECs (SILA, SILP, SIK and SAKD) and (2) a route via *Pax4*<sup>+</sup> progenitors to serotonergic EC and mature *Reg4*<sup>+</sup> EC cells (Fig. 3c). To identify transcriptional signatures and regulators of EEC lineage allocation, we extracted genes that were either only transiently expressed in a specific EEP stage or mature cell subtype (state-specific genes) or turned on with sustained expression in subsequent states (global or lineage-specific genes) (Fig. 3d, Extended Data Fig. 7a,d and Supplementary Tables 2 and 3). Transiently expressed markers in progenitor stages (for example *Sfrp5* in *Ngn3*<sup>+</sup> progenitor) are potentially important for specification. Global or lineage-specific markers (for example *Nefm* in peptidergic lineage) might regulate endocrine cell identity. Together, our data confirm EEC lineage differentiation from an early common progenitor into EC-biased (*Pax4*<sup>+</sup>) and non-EC-biased (*Arx/Isl1*<sup>+</sup>) progenitors and their respective molecular programs<sup>17,28,31</sup>. We then compared EEC lineage allocation in CD- and HFHSD-fed animals. We found that an HFHSD reduces the number of *Sox4*<sup>+</sup> early EEPs, increases the fraction of *Ngn3*<sup>+</sup> EEPs, reduces the number of *Lyz1*<sup>+</sup>*BrdU*<sup>+</sup> cells that correspond to label-retaining *Lgr5*<sup>+</sup> EECs and affects mature SILA and *Reg4*<sup>+</sup> EC cells, which are most abundant in the duodenum (Fig. 3e–o and Extended Data Fig. 7e–i)<sup>17</sup>. Serotonin (5-HT)<sup>+</sup>*Reg4*<sup>+</sup> EC numbers were decreased in HFHSD crypts, whereas *Ghrl*<sup>+</sup> SILA cells were more abundant in the duodenal crypts and villi (Fig. 3h–m). Numbers of ileal located SILA/SILP cells expressing *Gcg*, which encodes the incretin *Glp-1*, were also increased with an HFHSD (Fig. 3n,o). We next assessed whether an HFHSD induces transcriptome changes that might affect EEC lineage allocation or function. Single-cell messenger RNA expression levels of hormones did not differ between CD and HFHSD EEC subsets (Extended Data Fig. 7j). However, genes associated with metabolism (for example *Slc5a1*, *Acadl* in SIK cells), vesicular trafficking machinery and the secretome (for example *Cplx2* in EC and *Sct* in SIK cells) as well as signalling (transduction) (for example *ID1/ID3* in *Sox4*<sup>+</sup> progenitors, *Gnas* in EC cells) and transcription factors (for example *Cdx2* in *Sox4*<sup>+</sup> progenitor, *Hmgn3* in EC cells) were differentially expressed between CD and HFHSD conditions (Extended Data Fig. 7k–m and Supplementary Table 4).

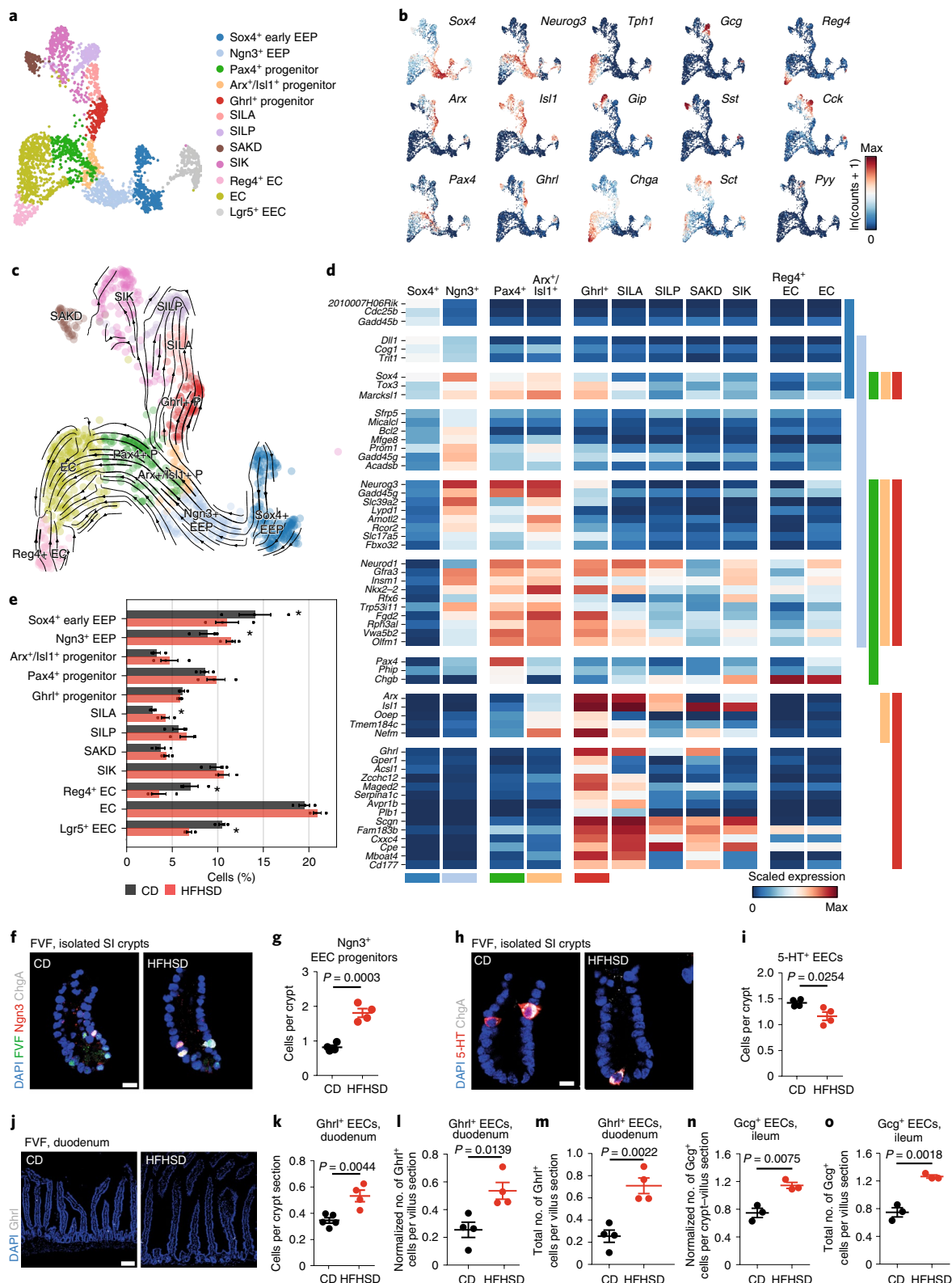
Next, we examined EEC functionality by assessing hormone secretion (Extended Data Fig. 7n). Intestinal EC cells produce over 90% of the body's serotonin<sup>32</sup>. We found that basal plasma levels of serotonin were lower in mice fed an HFHSD, which correlates with reduced numbers of 5-HT<sup>+</sup> EECs (*Reg4*<sup>+</sup> EC cells) (Extended Data Fig. 7o). By contrast, despite a higher number of *Ghrl*<sup>+</sup> SILA EECs, plasma levels of ghrelin were reduced in HFHSD-fed mice

**Fig. 3 | EEC lineage formation and composition in homeostasis and upon HFHSD.** **a**, Colour-coded UMAP plot of 2,865 EEC lineage cells from CD- and HFHSD-derived samples. Cluster annotation was based on known marker genes and labelling of mature EEC subtypes as previously described<sup>17</sup>. **b**, Expression levels of selected EEC markers (hormones and transcription factors) across EEC clusters plotted in UMAP space. **c**, Streamline plot of RNA velocity projected into UMAP space showing the direction of cell differentiation along trajectories for the endocrine lineage in CD-derived samples. Arrows indicate estimated future states of cells. **d**, Heatmap showing mean expression values per cluster of genes upregulated in EEP clusters during endocrine lineage formation. Genes were selected from (pan)endocrine, lineage- and stage-specific markers (Methods). Colour bars on the side indicate expression in the progenitor populations (Fig. 3a). **e**, Cell proportions in EEC lineage subsets from CD- and HFHSD-derived samples (mean ± s.e.m. of biologically independent samples, *n* = 3 mice per group, Dirichlet multinomial model). \*Indicates a credible shift (95% highest posterior density interval of the frequency shift modeled by a Dirichlet multinomial model does not overlap 0). **f–o**, Validation of EEC frequencies in CD- and HFHSD-fed FVF mice by immunofluorescence staining. Representative LSM images (**f**) and quantification of *Ngn3*<sup>+</sup> cells in isolated small intestinal crypts (**g**). Scale bar, 10 μm, *n* = 4 mice per group. Representative LSM images (**h**) and quantification of 5-HT<sup>+</sup> cells in isolated small intestinal crypts (**i**). Scale bar, 10 μm, *n* = 4 mice per group. Representative LSM images (**j**) and quantification of *Ghrl*<sup>+</sup> cells in duodenal sections (**k–m**). Scale bar, 100 μm; crypt, *n* = 5 for CD mice and *n* = 4 HFHSD mice; villus, *n* = 4 CD and HFHSD mice. Quantification of *Gcg*<sup>+</sup> cells in ileal sections (**n,o**), *n* = 3 CD and HFHSD mice. Data are shown as mean ± s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's *t*-test.

(Extended Data Fig. 7p), suggesting that *Ghrl*<sup>+</sup> EECs are functionally impaired upon HFHSD. *Glp-1* plasma levels were also increased, which corresponds with increased numbers of ileal *Gcg*<sup>+</sup> cells (Extended Data Fig. 7q). Taken together, these results show that an HFHSD impacts EEPs, alters the number of specific mature EEC subtypes and circulating gut hormone levels and affects expression

of genes that are potentially important for endocrine cell differentiation and/or function.

**HFHSD induces hyperproliferation of ISC and progenitors.** We next asked whether the increased number of *Ngn3*<sup>+</sup> EEPs, as well as the increase in villus length and higher abundance of



the enterocyte and goblet cell lineage under HFHSD conditions, resulted from enhanced proliferation. To test this, we identified proliferating cells using a cell-cycle signature score that assigns each cell to a cell-cycle state (G1, S and G2M; Methods) (Extended Data Fig. 8a). HFHSD mainly increased the proportion of cells in G2M phase, indicating that HFHSD alters cell-cycle dynamics (Fig. 4a,b). In particular, in ISCs, enterocytes, goblet cells and the early *Sox4*<sup>+</sup> and *Ngn3*<sup>+</sup> EEP clusters, the proportion of cycling cells was clearly increased (Fig. 4c and Extended Data Fig. 8b). Moreover, several genes, including the known cell-cycle regulators *Ccnb1*, *Cenpa*, *Dut*, *Pbk* and *Smc2*, which correlated with the S/G2M cell-cycle score, were differentially expressed in proliferating cells upon HFHSD, further suggesting that cell-cycle dynamics of ISCs and progenitors is accelerated (Fig. 4d and Supplementary Table 5).

To confirm the increase in proliferative activity, we first identified 5-ethynyl-2'-deoxyuridine (EdU)-positive progenitor cells 2 h after an EdU pulse. Consistent with the crypt scRNA-seq data, we found a higher number of proliferative progenitors in the transit-amplifying (TA) zone in SI crypts of HFHSD-fed FVF mice (Fig. 4e,f). To directly assess the number of proliferating ISCs we put *Lgr5*-ki mice<sup>33</sup> on an HFHSD (Extended Data Fig. 8c). In contrast to our scRNA-seq data, we found no difference in the numbers of ISCs between CD and HFHSD by flow cytometry on the basis of high *Lgr5*-EGFP fluorescence intensity (Extended Data Fig. 8d and Fig. 4g). However, the numbers of EdU-positive ISCs were increased on HFHSD (Fig. 4h,i).

Further, we determined antigen immunoreactivity to Ki67 and the migration rate of crypt cells. We observed that Ki67-labeled domains were significantly larger in mice fed an HFHSD and, in contrast to controls, extended into the villi. Also, the crypt-to-villus migration rate of pulse-labeled 5-bromo-2'-deoxyuridine (BrdU)-positive cells was increased on HFHSD (Extended Data Fig. 8e-i) which altogether suggests that an HFHSD enhances cell turnover.

In summary, these data demonstrate that an obesogenic diet increases the proliferation rate of ISCs and progenitors leading to increased villus length. Enhanced cell-cycle activity of ISCs did not result in physical expansion of the ISC pool due to accelerated differentiation and cell turnover. This is reflected by the decreased fraction of ISCs and increased fraction of progenitors in our crypt scRNA-seq data (Fig. 1e). These results further indicate that ISC homeostasis and identity is disturbed in HFHSD-fed animals.

#### HFHSD upregulates fatty acid synthesis and Ppar signalling.

A central role of metabolic pathways in the regulation of stem cell maintenance and fate control has been described in several adult stem cell systems<sup>34,35</sup>. Both Wnt/ $\beta$ -catenin and *Igf1*/insulin signalling integrate metabolic and proliferative cues, and increased

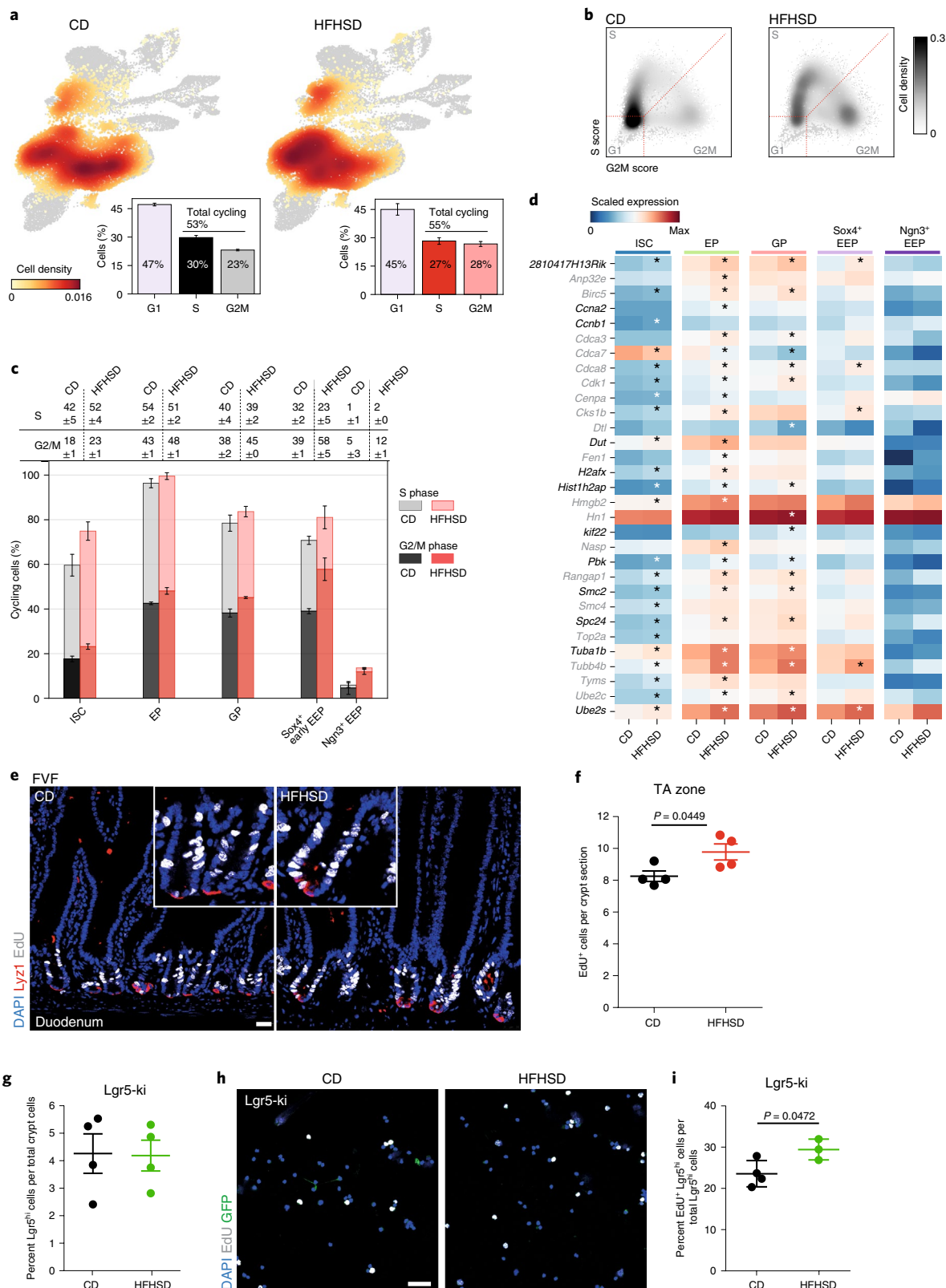
activity of these pathways has been shown to induce ISC hyperproliferation upon high-fat diet (HFD) feeding<sup>6,36–38</sup>. We assessed Wnt/ $\beta$ -catenin pathway activation in bulk ISC-enriched FVF<sup>low</sup> cells and FVF<sup>neg</sup> enterocyte progenitors, which we isolated by flow cytometry (Supplementary Fig. 2a,b). Unexpectedly, the level of nuclear  $\beta$ -catenin was decreased or unchanged in ISC-enriched FVF<sup>low</sup> cells or enterocyte progenitor-enriched FVF<sup>neg</sup> cells, respectively (Supplementary Fig. 2c,d). Consistently, expression of Wnt/ $\beta$ -catenin target genes was downregulated in ISC-enriched FVF<sup>low</sup> cells and HFHSD-derived single-cell ISCs (Supplementary Fig. 2e–g and Supplementary Table 1). As glycogen synthase kinase 3 $\beta$  (*Gsk3 $\beta$* ) is a negative regulator of Wnt/ $\beta$ -catenin signalling and an important modulator of cellular metabolism, we checked *Gsk3 $\beta$*  activity<sup>39</sup>. Decreased levels of phosphorylated *Gsk3 $\beta$*  in HFHSD-derived SI crypts indicated increased *Gsk3 $\beta$*  activity and enhanced *Gsk3 $\beta$* -mediated destruction of  $\beta$ -catenin (Supplementary Fig. 2h,i). Together, these results demonstrate that Wnt/ $\beta$ -catenin signalling does not drive hyperproliferation of ISCs and progenitors and enhanced progenitor recruitment in our HFHSD mouse model. Instead, decreased Wnt/ $\beta$ -catenin signalling indicates that differentiation of ISCs is accelerated and thus provides further evidence of disturbed ISC homeostasis<sup>40,41</sup>. HFHSD caused obesity and hyperinsulinaemia in our mouse model. Analysis of *InsrIgf1rAkt* pathway activity in SI crypt lysates showed increased phosphorylation of *InsrIgf1r* and *Akt* as well as upregulation of several genes associated with PI3K and *Akt* signalling, which is in line with pronounced hyperinsulinaemia and confirms previous findings that *Igf1* insulin signalling induces hyperproliferation in diet-induced obesity (Extended Data Figs. 1r–u and 9a–c).

To uncover additional metabolic pathways that link HFHSD to hyperproliferation and cell fate changes, we compared metabolite profiles from SI crypt regions of CD- and HFHSD-fed FVF mice using matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI)<sup>42</sup>. MALDI-MSI allows analysis of metabolites directly in tissue sections without isolation bias. We identified 297 discriminative masses; of these, 257 were enriched and 40 were less abundant in SI crypts of HFHSD-fed animals (Fig. 5a). Pathway enrichment analysis revealed that HFHSD upregulated metabolite signatures related to fatty acid biosynthesis (for example, octadecanoic acid) and linoleic acid metabolism (for example, phosphatidylcholine) and downregulated metabolites linked to pathways of glucose metabolism, such as the pentose phosphate pathway and pentose glucuronate interconversions (for example, D-glyceraldehyde 3-phosphate) (Fig. 5b and Supplementary Table 6). To map the metabolic changes from tissue to cell-type level, we integrated metabolomics data with bulk transcriptomes of ISC-enriched FVF<sup>low</sup> cells, secretory progenitor-enriched FVF<sup>high</sup> cells and enterocyte progenitor-enriched FVF<sup>neg</sup> cells (Extended

**Fig. 4 | HFHSD induces hyperproliferation of ISCs and progenitors.** **a**, Distribution of cycling cells across CD- and HFHSD-derived cell clusters depicted as cell densities projected onto the UMAP plot and quantified as proportions of cells in each cell-cycle stage. Cells were classified using a cell-cycle score, calculated using the expression of genes related to cell cycle. Data are mean  $\pm$  s.e.m. of biologically independent samples,  $n = 3$  mice per group. **b**, Distribution of ISCs and progenitors over the three cell-cycle stages visualized as cell densities in a scatter-plot of S- versus G2/M-phase score levels. Higher score levels indicate higher expression of involved genes. Dotted lines depict classification borders. Densities are Gaussian kernel estimates. **c**, Proportions of cycling cells in CD and HFHSD-derived ISCs and progenitor clusters. Table indicates percentages. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples,  $n = 3$  mice per group. EP, enterocyte progenitor; GP, goblet progenitor. **d**, Heatmap of mean expression values per cluster of selected genes used for cell-cycle scoring (black) or highly correlating with S and G2M scores (grey, Pearson correlation  $>0.7$ ). Only cells classified as cycling (S or G2M phase) are shown. \* Indicates differentially expressed genes between CD and HFHSD conditions (two-sided, limma, adjusted  $P < 0.01$ , logFC  $> 0.1$ ),  $n = 3$  mice per group. FC, fold change. *P* values are provided in Supplementary Table 5. **e, f**, Representative LSM images (**e**) and quantification (**f**) of EdU incorporation after a 2-h EdU (white) pulse in the TA zone in duodenal sections of CD- and HFHSD FVF mice. *Lyz1*<sup>+</sup> Paneth cells are shown in red. Scale bar, 25  $\mu$ m,  $n = 4$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's *t*-test. FDR, false discovery rate. **g**, Determination of *Lgr5*-EGFP<sup>hi</sup> cells from CD- and HFHSD-fed *Lgr5*-ki mice by flow cytometry,  $n = 4$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. **h, i**, Representative LSM images from cytospin (**h**) and quantification (**i**) of EdU<sup>+</sup>*Lgr5*-EGFP<sup>hi</sup> cells from CD- and HFHSD-fed *Lgr5*-ki mice after a 2-h EdU (white) pulse. Scale bar, 40  $\mu$ m,  $n = 4$  CD mice,  $n = 3$  HFHSD mice. Data are mean  $\pm$  s.d. of biologically independent samples. Statistical significance was determined by two-tailed Student's *t*-test.

Data Fig. 9d, Fig. 5c and Methods). The HFHSD metabolite signature overlapped with regulated genes involved in carbohydrate metabolism (for example, sucrose degradation, maturity onset diabetes of young (MODY) signalling) and pro-proliferative fatty acid biosynthesis pathways (for example, stearate synthesis

and acyl-CoA-hydrolysis)<sup>43</sup>. Regulation of these pathways was most pronounced in enterocyte progenitor-enriched FVF<sup>neg</sup> and ISC-enriched FVF<sup>low</sup> cells (Fig. 5d and Supplementary Table 7). Next, to determine the cell subtypes in which an HFHSD altered metabolism and to identify the molecular pathways associated with



deregulated HFHSD metabolites, we compared the transcriptional profiles of CD- and HFHSD-derived cells. We found that gene signatures associated with Ppar signalling (for example *Hmgcs2*, *Acdvl*, *Acaa2*, *Fabp1* and *Fabp2*) and fatty acid biosynthesis (for example *Acot1*, *Acox1*, *Scd2*, *Srebf1*, *Fads1* and *Me1*) were upregulated in HFHSD-derived ISCs and progenitor clusters (Fig. 5e–g and Supplementary Tables 1, 8 and 9). Upregulation of these gene sets was stronger in subpopulations with proximal identity; in particular, in proximal ISCs, enterocyte and goblet progenitors (Fig. 5e, f and Supplementary Table 8). Moreover, genes associated with carbohydrate metabolism (for example glycolysis, gluconeogenesis and the pentose phosphate pathway) were downregulated in these subtypes (Fig. 5e–g). We confirmed upregulation of active *Srebp1* (mature, m-*Srebp1*), *Acc*, *Pparγ* and *Scd1*, the master transcriptional regulators and key enzymes of Ppar signalling and fatty acid synthesis in HFHSD crypt protein lysates and in ISCs and progenitors, by a sensitive, targeted single-cell qPCR approach (Fig. 5h, i, Extended Data Fig. 10 and Supplementary Table 10). Finally, consistent with crypt enterocytes, we also found that mature villus enterocytes increase the expression of genes associated with intracellular fat accumulation and lipid uptake, de novo lipogenesis and peroxisomal fatty acid oxidation (Supplementary Fig. 3).

Thus, by integration of in situ metabolomics, bulk and single-cell transcriptomics and targeted protein expression analysis, we have revealed diet-induced metabolic rewiring and cell- and subtype-specific transcriptional changes that correlate with an increase in proliferation and endocrine dysfunction.

## Discussion

In this study we provide a basic mechanistic explanation of diet-induced ISC dysfunction and intestinal maladaptation that underlie the development of obesity and prediabetes and increase the risk for gastrointestinal cancer.

With our study we aimed to determine the immediate effects of a western-style HFHSD on intestinal function. We observed enlargement of the SI, longer villi and decreased crypt density and an altered cellular composition in the crypts, which was confirmed in the villi by lineage-tracing studies. We are aware that our lineage-tracing approach using the dual-fluorescent, inducible Cre-reporter *Foxa2*<sup>EGFP-CreERT2/+</sup>; *Gt(ROSA)26<sup>mTmG</sup>/+* mouse has its limitations; for example, inefficient labelling of stem cells, which generates fewer clonal ribbons in comparison to the *Lgr5*-ki reporter<sup>33</sup> and analysis of only one time point. Therefore, we validated our findings from the crypt scRNA-seq experiment not only by a lineage-tracing approach but also independent of a reporter in tissue sections.

We show that ISCs and progenitors are hyperproliferative, and that differentiation and cell turnover are accelerated by an HFHSD.

Accelerated differentiation and cell turnover also explain the discrepancy between the scRNA-seq data (decrease in ISCs) and the data from the *Lgr5*-reporter mice (no change in the number of ISCs). ISCs in HFHSD-fed mice divide faster and newly formed cells are still *Lgr5*-EGFP<sup>hi</sup>, but already express markers of differentiated cells. Whole-transcriptome-based clustering in our scRNA-seq data groups *Lgr5*<sup>hi</sup> ISCs and *Lgr5*<sup>hi</sup> progenitors separately, whereas when using only *Lgr5*-EGFP as a marker in tissue sections and flow cytometry we cannot discriminate between an ISC and a progenitor.

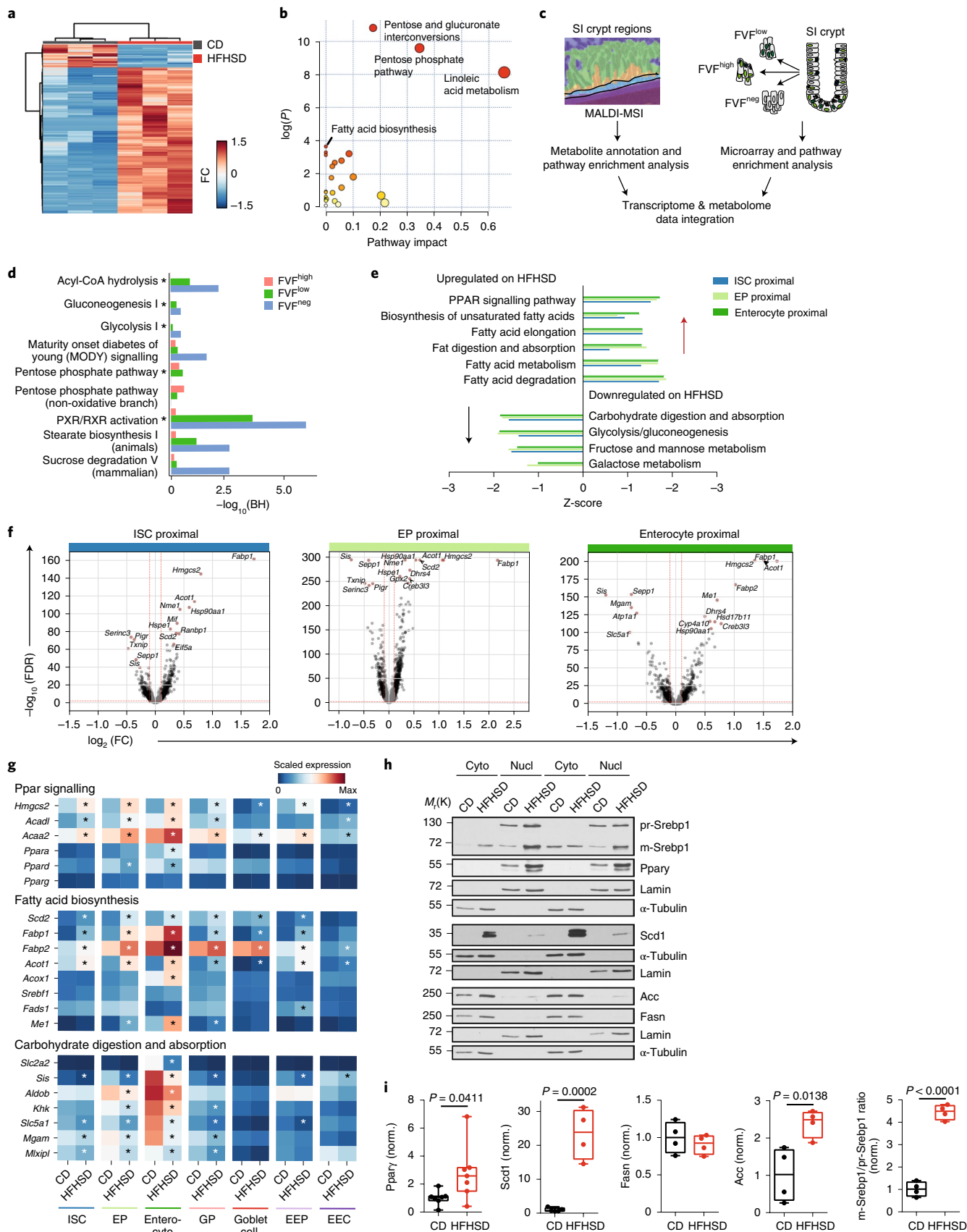
Hyperproliferation in the crypt, which can promote cancer initiation and progression, has previously been associated with an HFD; however, the molecular pathways that couple dietary cues to this cellular response are still debated and probably depend on the type of diet (for example fat and sugar source)<sup>37,38,44</sup>. Also, a role of metabolic pathways in ISC maintenance, number and fate control has been highlighted in several studies<sup>7,44,45</sup>. Hyperproliferation upon HFHSD is not driven by *Pparδ*-mediated activation of the major oncogenic pathway in the gut, Wnt/ $\beta$ -catenin signalling, as previously reported for a lard-based chronic HFD<sup>6,46,47</sup>. Instead, we found that a coconut oil- and sucrose-based high-lipid and -carbohydrate content diet specifically elevates pro-proliferative *Pparγ* signalling, *Srebp1*-mediated lipogenesis and *InsRlgf1r*-Akt signalling. Upregulation of *Pparγ* and *Srebp1*-mediated de novo lipogenesis has been associated with inflammation, increased proliferation and tumour progression in many types of cancer and thus provides a possible link between HFHSD-induced metabolic signalling, crypt-cell hyperproliferation and increased risk of gastrointestinal cancer<sup>43</sup>. As tumour initiation and progression require biomass production and therefore depend on a high nutrient supply, chronic activation of these pathways as observed with an HFHSD might reduce the barrier to oncogenic transformation or tumour growth and proliferation.

Further, we found a profound impact of an HFHSD on the enterocyte lineage. Enterocytes are metabolically rewired, increase the expression of genes linked to carbohydrate and fat uptake and show an intracellular fat accumulation, which is reminiscent of the vesicular accumulation of triglycerides in hepatic steatosis that causes liver fibrosis and cancer<sup>48–50</sup>. In addition, HFHSD induces a regional and spatial repatterning of enterocyte gene expression and function. An HFHSD increased the number of proximal-type enterocytes, which are specialized on carbohydrate and fatty acid absorption and altered enterocyte zonation along the crypt–villus axis (increased expression of the fatty acid transporter *Fabp1*). Thus, our results imply that enterocytes functionally adapt to an HFHSD, which may increase calorie intake and fat accumulation and promote obesity.

**Fig. 5 | Fatty acid synthesis and Ppar signalling are upregulated on HFHSD.** **a**, Heatmap-based clustering analysis of the 297 discriminative metabolite masses ( $FC \geq 2$  and  $P \leq 0.05$ ). Each coloured cell on the map corresponds to an intensity value, with samples in rows and features in columns. Euclidean distance and Ward's method were applied for clustering analysis. Statistical significance was determined by two-tailed Student's *t*-test. **b**, Pathway enrichment analysis of deregulated metabolites was performed with MetaboAnalyst 3.0. The enrichment method was hypergeometric test. Topology analysis was based on relative betweenness centrality. The *P* value was calculated from the enrichment analysis without adjustment ( $FC \geq 2$  and  $P \leq 0.05$ ). Metabolic pathways are represented as circles according to their scores from enrichment (vertical axis) and topology analyses (pathway impact, horizontal axis). **c**, Overview of experimental design for data integration. **d**, Ingenuity pathway analysis showing overlap of significantly deregulated metabolites on HFHSD from MALDI-MSI profiling and genes from microarray analysis. Shown are values from microarray analysis and the red line indicates the significance cutoff. **e**, Enriched KEGG pathways in genes differentially regulated between CD and HFHSD conditions (Enrichr, Fisher's exact test, two-sided). Genes with  $FDR < 0.01$  and  $\log FC > 0.1$  were considered and weighted by  $\log FC$ . **f**, Volcano plots showing differential expression and its significance ( $-\log_{10}(FDR)$ , *limma*) for each gene on HFHSD compared to CD. Red lines indicate thresholds used for significance level and gene expression change and regulated genes are highlighted in black. Annotated genes are the top ten genes ranked by FDR. **g**, Mean expression levels for selected genes. \* Indicates a significant change (*limma*,  $FDR < 0.01$ ,  $\log FC > 0.1$ ). **h, i**, Protein expression analysis by western blot in cytoplasmic (cyto) and nuclear (nucl) extracts from SI crypts of CD- and HFHSD-fed FVF mice. Representative immunoblots (**h**) and relative quantification of band signal intensity (**i**),  $n = 4$  mice per group (*Srebp1*, *Scd1*, *Acc*, *Fasn*) and  $n = 7$  mice per group (*Pparγ*). Data are presented as box-and-whisker plots. The lower and upper boundaries of the boxes represent the 25th and 75th percentiles, respectively. The centre lines indicate the medians and whiskers represent the maximum and minimum values. Statistical significance was determined by two-tailed Student's *t*-test. Circles represent biological independent samples. norm., normalized.

Our in-depth molecular and functional analysis of the EEC lineage reveals mechanisms of EEC dysfunction in obesity that include (1) higher abundance of *Ngn3*<sup>+</sup> EEPs owing to increased

proliferation, (2) lower abundance of serotonergic *Reg4*<sup>+</sup> EC cells and lower plasma serotonin levels, (3) higher abundance of peptide-ghrelin *Ghr1*<sup>+</sup> SILA cells but lower ghrelin plasma levels, (4) increased



numbers of ileal Gcg<sup>+</sup> cells and increased Glp-1 levels and (5) lower abundance of *Lgr5*<sup>+</sup> EECs. The physiological roles of gut-derived serotonin are broad and it regulates various processes both in the gut and systemically<sup>32</sup>. Accumulating evidence indicates a link between peripheral serotonin and systemic glucose and lipid metabolism, as well as metabolic diseases<sup>51</sup>. Whether blood serotonin levels are changed in obesity is controversial, due to difficulties in measuring serotonin<sup>51</sup>. However, intraperitoneal injection of 5-HT to mice inhibits weight gain, hyperglycaemia and insulin resistance on an HFD<sup>52</sup>. Thus, reduced numbers of serotonin-producing cells and blood serotonin levels as observed in our mouse model might promote the development of obesity. The hunger hormone ghrelin is known to increase gastric emptying and decrease insulin secretion<sup>53</sup>. A negative correlation between plasma insulin and ghrelin has been reported in human obesity, which is in line with our data<sup>54</sup>. Notably, we found that lower plasma ghrelin levels are not due to a reduced number of Ghrl<sup>+</sup> SILA cells, suggesting that an impaired secretory machinery might affect ghrelin levels. However, given that ghrelin-secreting cells are also present in the stomach, changes in plasma ghrelin levels may not be solely attributed to duodenal Ghrl<sup>+</sup> EECs. Lower postprandial GLP-1 levels are reported in obesity and incretin hormone secretion and activity are impaired in individuals with prediabetes, although findings are contradictory<sup>55</sup>. In our obesity model, however, the number of Gcg-expressing cells and Glp-1 plasma levels were increased. Glp-1 stimulates insulin secretion from pancreatic  $\beta$ -cells, so the higher levels of circulating Glp-1 in HFHSD-fed mice may be a compensatory response to insulin resistance at the prediabetic state, which leads to hyperinsulinaemia.

Finally, in contrast to other studies, Paneth cell numbers were not affected in our obesity model<sup>6</sup>.

The discrepancies between our results and those of previous studies support the emerging evidence that the intestine fine-tunes its response to environmental stimuli<sup>6,37,38</sup>. Different factors, such as the selected mouse model (for example, diet versus genetically induced obesity), type of diet (for example HFD versus HFHSD), fat source (for example lard versus coconut oil) and/or duration of diet (short-term versus chronic) can influence intestinal remodelling and response. For instance, in contrast to the coconut oil, high-sucrose diet for 12–14 weeks used in this study, a lard-based low-sugar chronic diet for 9–14 months caused shorter villi, increased the number of ISCs, decreased the number of Paneth cells and did not affect goblet cells and EECs<sup>6</sup>. Hence, careful selection and reporting of dietary information in animal studies is crucial to interpret and contextualize results.

In summary, our study reveals that functional maladaptation of the gut in response to an HFHSD is caused by disturbed ISC identity, changes in the regional identity of cells and an altered mature cell-type composition. Further, we describe targetable pathways that are induced by an HFHSD and potentially underlie the pathogenesis of the metabolic syndrome and gastrointestinal cancer. This new understanding of the mechanisms of disease is crucial to develop non-invasive therapeutic options to resolve obesity and insulin-dependent diabetes, for example by counteracting enteroendocrine dysregulation (for example, by elevating peripheral serotonin levels) and increased nutrient absorption (for example, by inducing distal enterocyte phenotypes).

## Methods

**Experimental model. Animals.** Animal experiments were carried out in compliance with the German Animal Protection Act and with the approved guidelines of the Society of Laboratory Animals and of the Federation of Laboratory Animal Science Associations. This study was approved by the institutional Animal Welfare Officer (Helmholtz Center Munich) and by the Government of Upper Bavaria, Germany. Homozygous FVF mice were generated as previously described and backcrossed to C57BL/6 background<sup>22</sup>. *Foxa2*<sup>EGFP-CreERT2</sup> mice<sup>25</sup> (CD1 background) were crossed with *Gt(ROSA)26<sup>mTmG/+</sup>* mice<sup>24</sup> (mixed 129/SvJ, C57BL/6J background) to obtain heterozygous *Fo*

*x2*<sup>EGFP-CreERT2/+;Gt(ROSA)26<sup>mTmG/+</sup> animals and bred in our own facilities.</sup>

Other mouse lines were *Lgr5-EGFP-IRES-creERT2* (ref. <sup>33</sup>) (*Lgr5*-ki, C57BL/6J background) and wild-type C57BL/6N (bred in our own facilities).

Mice were housed in groups of two to four animals and maintained at 23 ± 1 °C and 45–65% humidity on a 12-h dark/light cycle with ad libitum access to diet and water unless otherwise indicated. All experiments were performed using male animals at 3 to 7 months of age.

**Dietary interventions.** For dietary interventions, 10–12-week-old male mice were randomized into test groups matched for body weight, with similar variance, and given ad libitum access to either an obesogenic HFHSD (58% kcal from fat, 25% kcal from carbohydrates, 17% kcal from protein (Research Diets, no. D12331)) or CD (11% kcal from fat, 64% kcal from carbohydrates, 25% kcal from protein (ssniff Spezialdiäten, E15051-04)) for a period of 11–13 weeks. Body weights were measured every second week.

**Body composition analysis.** Lean and fat masses were measured in FVF mice 12 weeks after the start of the diet using quantitative nuclear magnetic resonance technology (EchoMRI).

**Glucose tolerance and insulin secretion tests.** Glucose tolerance was assessed by an oral glucose tolerance test (oGTT) in FVF mice, maintained for 12 weeks on CD or HFHSD. After a 6-h fast, mice received an oral glucose bolus (1.5 mg g<sup>-1</sup> body weight of 20% (*wt/v*) D-(+)-glucose (Sigma-Aldrich) in PBS). Tail blood glucose concentrations were measured with a handheld glucometer (Abbott) before (0 min) and 15, 30, 60 and 120 min after the glucose bolus. To measure the insulin secretion, tail vein blood samples were collected into EDTA-coated microvette tubes (SARSTEDT) at time points 0, 15 and 30 min of the oGTT. Plasma was extracted by centrifugation (3,500 r.p.m., 15 min, 4 °C) and insulin concentration was determined using the Ultra-Sensitive Mouse Insulin ELISA Kit (Crystal Chem, 90080) according to the manufacturer's instructions.

The homeostasis model assessment of insulin resistance (HOMA-IR) and HOMA- $\beta$  were used to assess insulin resistance and beta-cell function, respectively, in FVF mice at 12 weeks after diet start. HOMA indices were calculated from basal blood glucose and plasma insulin levels after a 6-h fast based on the conventional formulas: HOMA-IR = fasting blood glucose (mg per 100 ml) × fasting insulin ( $\mu$ U per ml)/405 and HOMA- $\beta$  = fasting insulin ( $\mu$ U per ml) × 360/fasting glucose (mg per 100 ml) – 63 (ref. <sup>56</sup>).

**Plasma hormone measurements.** Circulating hormones, serotonin, ghrelin and Glp-1 were assessed in FVF mice maintained for 13 weeks on a CD or an HFHSD. For basal levels (ghrelin and serotonin), mice were fasted for 6 h and tail vein blood was sampled into EDTA-coated microvette tubes (SARSTEDT). To compare postprandial plasma hormone levels (Glp-1, ghrelin), fasted mice were gavaged with 250  $\mu$ l of mixed-meal-containing liquid diet (Osmolite HiCal, Abbott) supplemented with dextrose at 20% (*wt/v*) (Sigma-Aldrich). Blood was collected 10 min after the mixed-meal bolus either from the tail vein or, under terminal anaesthesia with isoflurane, from the vena cava<sup>57–59</sup>. For Glp-1 measurement, blood samples were immediately mixed with 0.1 mM Diprotin A (Abcam, 145599) and 500 KIU ml<sup>-1</sup> aprotinin (Sigma, A-1153). Plasma was extracted by centrifugation (13,000 r.p.m., 2 min, 4 °C). Total ghrelin concentrations were determined using an ELISA kit from Millipore-Merck (EZRGRT-91K), total plasma Glp-1 was measured using a mouse Glp-1 ELISA kit (Crystal Chem, 81508) and serotonin concentration was determined using a serotonin ELISA kit from Enzo (ADI-900-175).

**Proliferative cell labelling with EdU and BrdU and tamoxifen administration in mice.** To assess the epithelial replication rate in the SI, EdU (Thermo Fisher Scientific, A10044) or BrdU (Sigma, no. B5002) was administered as an intraperitoneal injection at 100  $\mu$ g g<sup>-1</sup> body weight or at 50  $\mu$ g g<sup>-1</sup> body weight, respectively, each from a 10 mg ml<sup>-1</sup> stock and in sterile PBS. Mice were killed 2 h post-EdU or 24 h post-BrdU administration. For the assessment of BrdU label retention, FVF mice maintained for 10 weeks on a diet were given BrdU in drinking water at 1 mg ml<sup>-1</sup> supplemented with 1% sucrose for 14 d. BrdU was then withdrawn and mice were further maintained on a diet with ad libitum drinking water for a chase period of 21 d. A group of mice was killed after 14 d of continuous BrdU labelling or after a 21 d period of chase. For short-term genetic lineage studies, *Foxa2*<sup>EGFP-CreERT2/+;Gt(ROSA)26<sup>mTmG/+</sup> mice were fasted for 3 h and Cre-recombinase activity was induced by a titrated single dose of tamoxifen administered orally by gavage (Sigma-Aldrich, T5648) at 0.25 mg g<sup>-1</sup> body weight in sunflower oil. Mice were killed 70 h after the tamoxifen gavage.</sup>

**Crypt and villus isolation and single-cell preparation.** Isolation of small intestinal crypts was carried out as previously reported<sup>40</sup>. In brief, SIs were removed and washed with cold PBS. Villi were scraped off with a glass slide. The remaining tissue was cut into 2-cm pieces, washed several times with cold PBS and incubated in 2 mM EDTA/PBS for 35 min at 4 °C on a tube roller. Subsequently, crypts were collected by rigorous shaking and filtered through a 70- $\mu$ m mesh to remove villous fragments. For single-cell preparation, isolated crypts were incubated with TrypLE (Life Technologies, no. 12605) for 5 min on ice and then 5 min at 37 °C and treated



with  $10 \mu\text{g ml}^{-1}$  DNase in crypt complete medium (DMEM/F-12 containing 10% FCS) for 5 min at  $37^\circ\text{C}$ . Single-cell suspension was achieved by gentle repeated pipetting. Cells were washed twice with 2% FCS in PBS and pelleted at 300g for 5 min at  $4^\circ\text{C}$ . For flow cytometry, cells were collected in 1–2 ml FACS buffer (2% FCS, 2 mM EDTA in PBS) (Sigma-Aldrich, no. Y0503) and passed through 40- $\mu\text{m}$  cell strainer caps of FACS tubes.

To obtain a single-cell suspension of villi cells, villi were scraped off and incubated with TrypLE as described for crypt cells.

**Flow cytometry.** For gene expression measurement (microarray, single-cell transcriptomics) and western blotting, small intestinal crypt cells were sorted using FACS-Aria III (FACSDiva software v.6.1.3, BD Bioscience) with a 100- $\mu\text{m}$  nozzle. For all experiments, single cells were gated according to their FSC-A (front scatter area) and SSC-A (side scatter area). Singlets were gated dependent on the FSC-W (front scatter width) and FSC-H (front scatter height) and dead cells were excluded using 7-AAD (eBioscience, no. 00-6993-50). For quantitative PCR with reverse transcription (qRT-PCR), cells were sorted directly into Qiazol lysis reagent (QIAGEN, no. 79306). To obtain FVF-enriched small intestinal crypt-cell samples for scRNA-seq, we sorted 30,000 FVF<sup>+</sup> (FVF<sup>low</sup> and FVF<sup>high</sup>) cells followed by sorting 30,000 live crypt cells per sample. Cells were sorted into modified FACS buffer (2% FCS, 0.02 mM EDTA in PBS).

**RNA isolation for qRT-PCR and microarray.** For bulk gene profiling studies (qPCR, microarray), RNA isolation from crypts or flow-sorted cells was performed using the RNA isolation kit miRNeasy Mini (QIAGEN, no. 217004) or miRNeasy Micro kit (QIAGEN, no. 217084) depending on the amount of the sample. Complementary DNA was synthesized using the SuperScript VILO cDNA synthesis kit (Invitrogen, no. 11754). RNA was reverse transcribed and amplified with the Ovation PicoSL WTA System V2 kit (NuGEN, no. 331248).

**Microarray analysis.** For gene profiling of flow-sorted FVF<sup>low</sup>, FVF<sup>high</sup> and FVF<sup>neg</sup> cells from CD- and HFHSD-fed FVF mice, total RNA was isolated as described above and RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent RNA 6000 Pico Kit). RNA was amplified with the Ovation PicoSL WTA System V2 in combination with the Encore Biotin Module (NuGen). Amplified cDNA was hybridized on Affymetrix Mouse Gene 1.0 ST arrays. Staining (Fluidics script FS450\_0007) and scanning of the microarray were performed according to the Affymetrix expression protocol including minor modifications as suggested in the Encore Biotin kit protocol. Expression Console (v.1.3.0.187, Affymetrix) was used for quality control and annotation of the normalized robust microarray analysis gene-level data, and standard settings, including median polish and sketch-quantile normalization, were used.

**TaqMan qRT-PCR.** For gene expression analysis, real-time qRT-PCR was performed using TaqMan probes (Life Technologies), TaqMan Fast Advanced Master Mix (Applied Biosystems, no. 4444557) or TaqMan Universal Master Mix II (Applied Biosystems, no. 4440040) for amplified cDNA and the ViiA 7 Real-Time PCR System (Thermo Fisher Scientific).

The following probes were used: Mm00782745\_s1 for *Rpl37*, Mm00438890\_m1 for *Lgr5*, Mm01320260\_m1 for *Olfm4*, Mm01268891\_g1 for *Ascl2*, Mm00443610\_m1 for *Axin2*, Mm03928990\_g1 for *RN18S*, Mm02524776\_s1 for *Fzd2*, and Mm00433409\_s1 for *Fzd7*.

**scRNA-seq: RNA preparation, library generation and sequencing.** FVF-enriched single-cell samples of crypts isolated from the small intestinal epithelium (duodenum, jejunum and ileum) and villus samples from the SI of C57BL/6N mice were prepared as described above. Dead cells were excluded by flow cytometry after 7AAD labelling. Dead cell exclusion was controlled by trypan blue staining and sorted cells were counted. Single-cell libraries were generated using the Chromium Single cell 3' library and gel bead kit v2 (10X Genomics, no. 120237) according to the manufacturer's instructions. Libraries were sequenced on a HiSeq4000 (Illumina) with 150-bp paired-end sequencing of read 2.

**MALDI-MSI.** Fresh-frozen samples were cut into 12- $\mu\text{m}$  sections using a cryo-microtome at  $-20^\circ\text{C}$  (Leica CMI1950, Leica Microsystems) and mounted onto precooled conductive indium-tin-oxide-coated MALDI target glass slides (Bruker Daltonics). Sections were coated with 9-aminoacridine hydrochloride monohydrate matrix (Sigma-Aldrich) at  $10 \text{ mg ml}^{-1}$  in water/methanol 30:70 (v/v) by a SunCollect automatic sprayer (Sunchrom). The matrix application was performed at flow rates of 10, 20 and 30, respectively, for the first three layers. The other five layers were performed at  $40 \mu\text{l min}^{-1}$ . MALDI-MSI measurement was performed on a Bruker Solarix 7T FT-ICR-MS (Bruker Daltonics). MALDI-MSI data were acquired over a mass range of  $m/z$  50–1,000 in negative ionization mode with 30- $\mu\text{m}$  spatial resolution using 50 laser shots at a frequency of 500 Hz. The acquired data underwent spectrum processing in FlexImaging v.4.2 (Bruker Daltonics). Following MALDI imaging experiments, the matrix was removed with 70% ethanol. Tissue sections were stained with haematoxylin and eosin. Slides were scanned with a MIRAX DESK digital slide-scanning system (Carl Zeiss MicroImaging).

**Western blot.** For protein expression analyses, whole-cell lysates from isolated crypts or flow-sorted cells were prepared using the RIPA buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Igepal, 0.1% SDS, 0.5% sodium deoxycholate) containing phosphatase (Sigma-Aldrich, P5726, P0044) and protease inhibitors (Sigma-Aldrich, P8340). Nuclear and cytosolic extracts from isolated crypts were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents kit (Thermo Fisher Scientific, no. 78833) according to the manufacturer's instructions. Cell lysates were mixed with Laemmli sample buffer, resolved by SDS-PAGE and blotted onto a PVDF membrane (Bio-Rad). Membranes were blocked with 5% milk in Tris-buffered saline containing 0.2% Tween-20, then incubated overnight with primary antibodies in blocking solution at  $4^\circ\text{C}$ , followed by a 1-h incubation with horseradish peroxidase (HRP)-conjugated IgG secondary antibodies. Protein bands were visualized using a chemiluminescence reagent (Bio-Rad, no. 170-5061) and quantified using ImageJ software. For quantification, expression of proteins was normalized to  $\alpha$ -tubulin or lamin in cytoplasmic or nuclear fractions, respectively.

Primary antibodies used were mouse anti-Srebp1 (1:1,000 dilution, Novus Biologicals, NB600-582S); rabbit anti-Acc (1:1,000 dilution, Cell Signaling Technology, 3676); rabbit anti-Ppary (1:1,000 dilution, Cell Signaling Technology, 2435); goat anti-lamin (1:1,000 dilution, Santa Cruz, sc-6217); mouse anti- $\alpha$ -tubulin (1:1,000 dilution, Sigma-Aldrich, T6199); rabbit anti-Fasn (1:1,000 dilution, Cell Signaling Technology, 3180); rabbit anti-Scd1 (1:1,000 dilution, Cell Signaling Technology, 2794); mouse anti- $\beta$ -catenin (1:1,000 dilution, BD, 610154); rabbit anti-Gsk3 $\beta$  (1:5,000 dilution, Cell Signaling Technology, 12456); and rabbit anti-phospho Gsk3 $\beta$  (1:5,000 dilution, Cell Signaling Technology, 5558). Secondary antibodies used were goat anti-mouse HRP (1:15,000 dilution, Dianova, 115036062); goat anti-rabbit HRP (1:15,000 dilution, Dianova, 111036045) or rabbit anti-goat HRP (1:15,000 dilution, Dianova, 305035045).

**Tissue morphology assessment.** For tissue histology, intestines were flushed and fixed in 4% paraformaldehyde (PFA) overnight, paraffin embedded according to standard procedures and sectioned at 6  $\mu\text{m}$ . Sections were dried, dehydrated through a graded ethanol series and cleared in xylene. Standard haematoxylin and eosin staining was performed and images were acquired using the Zeiss AXIO Scope A1 microscope (Carl Zeiss AG).

**Histochemistry and immunofluorescence.** SIs were isolated, rinsed with ice-cold PBS and fixed with 4% PFA for 3 h at  $4^\circ\text{C}$ . Tissue was cryopreserved through a progressive sucrose gradient (7.5% for 1 h, 15% for 1 h, 30% sucrose overnight), embedded in a tissue-freezing medium (Leica Biosystems, no. 14020108926) and sectioned at 14  $\mu\text{m}$ . For whole-mount staining, isolated small intestinal crypts were fixed in 4% PFA for 30 min at room temperature (RT) and subsequently washed in PBS. For immunofluorescence staining, sections or crypts were permeabilized with 0.5% Triton X-100 in PBS for 30 min at RT, blocked (10% FCS, 0.1% BSA and 3% donkey serum in PBS/0.1% Tween-20) for 1 h and incubated with primary antibodies overnight at  $4^\circ\text{C}$ . Sections or crypts were washed in PBS/0.1% Tween-20 and incubated with secondary antibodies in blocking solution for 1 h at RT, stained with DAPI (ROTH, 6335.1) to visualize the nuclei and mounted with the Elvanol antifade reagent.

To assess proliferation, EdU staining was performed using the Click-iT Staining kit (Invitrogen, no. C10340) according to the manufacturer's instructions.

For BrdU staining, sections were incubated with 3.3 N HCl for 10 min on ice, followed by incubation for 50 min at  $37^\circ\text{C}$  and two wash steps with borate buffer, pH 8.5, to neutralize the reaction (each wash 15 min at RT).

Ki67 immunoreactivity was assessed on paraffin tissue sections after heat-induced antigen retrieval (at  $\sim 90^\circ\text{C}$ , 10 min) using the Antigen Unmasking Solution, Citric Acid-Based (Vector Laboratories, no. H-3300-250).

Fluorescent images were obtained with a Leica SP5 confocal microscope (Leica Microsystems) and analysed using LAS AF software (LAS AF software v.2.6.0-7266).

The primary antibodies used were chicken anti-GFP (1:600 dilution, Aves Labs, GFP-1020); goat anti-ChgA (1:200 dilution, Santa Cruz, sc-1488); rabbit anti-Lyz1 (1:1,000 dilution, DAKO, M0776); rabbit anti-Muc2 (1:1,000 dilution, Santa Cruz, sc-7314); rat anti-BrdU (1:200 dilution, Abcam, ab6326); rabbit anti-5-HT (1:1,000 dilution, Neuromics, RA20080); anti-rabbit Ngn3 (1:100 dilution, a kind gift from H. Edlund); goat anti-villin (1:200 dilution, Santa Cruz, sc-7672); goat anti-ghrelin (1:200 dilution, Santa Cruz, sc-10368); rabbit anti-Ki67 (1:200 dilution, Abcam, ab15580); rabbit anti-E-cadherin (extracellular domain) (1:1,000 dilution, a gift from D. Vestweber); rabbit anti-Fabp1 (1:300 dilution, Abcam, ab222517); and rabbit anti-Apoa4 (1:300 dilution, Abcam, ab231660). The secondary antibodies used were donkey anti-chicken Alexa Fluor 488 (1:800 dilution, Dianova, 703225155); donkey anti-mouse Cy5 (1:800 dilution, Dianova, 715175151); donkey anti-goat Alexa Fluor 555 (1:800 dilution, Invitrogen, A21432); donkey anti-rabbit Alexa Fluor 555 (1:800 dilution, Invitrogen, A31572); and donkey anti-rabbit Alexa Fluor 649 (1:800 dilution, Dianova, 711605152).

**RNAScope in situ hybridization for detection of target RNA.** In situ detection of *FOXA2*, *LGR5* and *OLFM4* mRNA was performed using the RNAScope Intro Pack for Multiplex Fluorescent Reagent Kit v2-Mm (ACD; no. 323136) according to the

manufacturer's protocol. RNAscope 3-Plex Negative Control Probe (dapB; ACD; no. 320871) and RNAscope 3-Plex Positive Control Probe (ACD; no. 320881) were used as internal controls. Probes used were Mm-Foxa2 (ACD; no. 409111), Mm-Olfm4-C2 (ACD; no. 311831-C2) and Mm-Lgr5-C3 (ACD; no. 312171-C3) with respective fluorescent dyes Opal 520 Reagent (AKOYA; no. FP1487001KT) for channel 1, Opal 570 Reagent (AKOYA; no. FP1488001KT) for channel 2 and Opal 690 Reagent (AKOYA; no. FP1497001KT) for channel 3, diluted 1:800 in RNAscope Multiplex TSA Buffer (ACD; no. 322809). The assay was performed on FFPE mouse intestinal jejunum sections prepared as described above with a thickness of 7  $\mu\text{m}$  and standard pretreatment conditions with protease III (ACD; no. 322340) as recommended by the manufacturer.

After completion of the RNAscope assay, sections were stained using DAPI (ROTH, 6335.1) to visualize the nuclei, washed with PBS three times and mounted with ProLong Diamond Antifade Mountant (Life Technologies; no. P36970). Samples were visualized using a Leica SP5 confocal microscope (LAS AF software v.2.6.0-7266).

**Gene expression analysis from bulk sorted cells.** Statistical bulk transcriptome analyses were performed using the statistical programming environment R implemented in CARMAweb<sup>61,62</sup>. Gene-wise testing for differential expression was carried out employing the limma *t*-test. Sets of regulated genes were defined by raw  $P < 0.01$ , FC  $> 1.3\times$  and average expression in at least one group  $> 32$ . Enriched canonical pathways were analysed through the use of QIAGEN's ingenuity pathway analysis (<https://www.qiagen.com/ingenuity>). Microarray data are available at Gene Expression Omnibus (GEO).

**Bioinformatics and statistical analysis of MALDI-MSI data.** MATLAB R2014b (v.7.10.0, Mathworks) was used as MALDI spectral pre-processing tool for the subsequent data bioinformatics analysis. Peak picking was performed using an adapted version of the LIMPIC algorithm<sup>63</sup>. In brief, the parameters of peak picking included  $m/z$  0.0005 minimal peak width, signal-to-noise threshold of 4 and intensity threshold of 0.01%. Isotopes were automatically identified and excluded. Statistical comparisons were performed with a Student's *t*-test (two-tailed). Metabolites were considered to be significant if they had an intensity FC  $\geq 2$  and a  $P$  value  $\leq 0.05$ . Metabolite annotation was performed by matching accurate mass with databases (mass accuracy  $\leq 4$  ppm, METLIN, <http://metlin.scripps.edu/>; Human Metabolome Database, <http://www.hmdb.ca/>; MassTRIX, <http://masstrix3.helmholtz-muenchen.de/masstrix3/>; METASPACE, <http://annotate.metaspacesoftware.com/>). Heat-map-based clustering and enrichment analysis of metabolic pathways were performed with MetaboAnalyst v.3.0 (<http://www.metaboanalyst.ca>).

**Integration of bulk transcriptome and metabolome.** Lists of genes participating in the candidate pathways from the ingenuity canonical pathways analysis of the microarray data of FVF<sup>low</sup>, FVF<sup>high</sup> and FVF<sup>neg</sup> cells were compiled. For each of the genes on the lists, information on the reaction it is involved in and the participating metabolites were extracted from the mouse specific BiGG databank (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2874806/>, <https://www.ncbi.nlm.nih.gov/pubmed/20959003>). If genes and metabolites within a reaction were also significant in their analyses as described above, the reaction and its pathway were said to be affected at both the metabolomics and the transcriptomics levels.

**Single-cell gene expression analysis by microfluidic qRT-PCR.** To assess the transcriptional profiles of single FVF<sup>low</sup> and FVF<sup>high</sup> cells, a nested single-cell qPCR design was used. FVF<sup>low</sup> or FVF<sup>high</sup> cells isolated from three CD- and HFHSD-fed FVF mice were sorted as described above directly into single wells of 96-well plates containing 5  $\mu\text{l}$  of a pre-amplification solution composed of 1.2  $\mu\text{l}$  5 $\times$  VIL0 reaction mix (Invitrogen, no. 11754-050), 0.3  $\mu\text{l}$  20 U  $\mu\text{l}^{-1}$  SUPERase-In (Ambion, no. AM2694), 0.25  $\mu\text{l}$  10% NP40 (Thermo Fisher Scientific, no. 28324), 0.25  $\mu\text{l}$  RNA spikes mix (Fluidigm, no. 100-5582) and 3  $\mu\text{l}$  of nuclease-free water (Promega, no. P119C). Cells were lysed by incubation at 65°C for 90 s and cDNA transcription from RNA was performed by reverse transcription cycling (25°C for 5 min, 50°C for 30 min, 55°C for 25 min, 60°C for 5 min and 70°C for 10 min) with 1  $\mu\text{l}$  reverse transcription mix solution containing 0.15  $\mu\text{l}$  10 $\times$  SuperScript enzyme mix (Invitrogen, no. 11754-050), 0.12  $\mu\text{l}$  T4 Gene 32 Protein (New England BioLabs, no. M0300S) and 0.73  $\mu\text{l}$  nuclease-free water. The efficiency and specificity of outer and inner primer pairs for target-specific cDNA amplification were tested in advance. Primers showing single peaks and single bands by melt curve analysis and by separation of qPCR products on a 2.5% agarose gel, respectively, were considered specific. Primer efficiency was analysed over a range of tenfold cDNA dilutions (1:1 to 1:100) and primers with 100  $\pm$  15% efficiency were qualified for further proceedings. Specific target amplification was performed with 9  $\mu\text{l}$  reaction mix containing 7.5  $\mu\text{l}$  TaqMan PreAmp Master Mix (Applied Biosystems, no. 4391128), 0.075  $\mu\text{l}$  0.5 M EDTA, pH 8.0 (Invitrogen, no. Am9260G), 1.5  $\mu\text{l}$  10 $\times$  outer primer mix (500 nM) under the following cycling conditions: enzyme activation step at 95°C for 10 min, 20 cycles of denaturation for 5 s at 96°C and 4 min annealing/extension at 60°C. Amplified cDNA samples were then treated with 6  $\mu\text{l}$  Exonuclease I reaction mix containing 0.6  $\mu\text{l}$  reaction buffer, 1.2  $\mu\text{l}$  Exonuclease I (New England BioLabs, no. M0293S) and 4.2  $\mu\text{l}$  nuclease-free

water. To increase target specificity, amplified single-cell cDNA samples were analysed with gene-specific inner primer pairs and SsoFast EvaGreen Supermix with Low ROX (Bio-Rad Laboratories, no. 172-5210) using the 96.96 Dynamic Array on the BioMark System (Fluidigm). BioMark Real-Time PCR Analysis software (Fluidigm) was used to calculate Ct values.

**Computational analyses of single-cell data.** A detailed description of the computational analyses of single-cell data is provided in the Supplementary information.

**Statistical analyses.** Data collection was performed using Microsoft office Excel 2016–2018 and statistical analysis was performed using GraphPad Prism 6 Software (GraphPad Software). All data are shown as mean  $\pm$  s.e.m. unless otherwise specified. In box-and-whiskers plots, data are represented as minimum and maximum with centre lines indicating the median. All samples represent biological replicates. For statistical significance testing of two independent groups, an unpaired two-tailed Student's *t*-test was used. For statistical comparison of longitudinal data (body weight curves, GTT and IST), two-way analysis of variance corrected by Sidak's multiple comparison test was used.  $P$  values of  $< 0.05$  % were considered statistically significant. For metabolic studies (GTT and IST), sample size was statistically determined; otherwise sample size estimates were not used. Studies were not blinded and investigators were not blinded during analysis.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data generated or analysed during this study are included in this article and its supplementary files. Microarray data have been submitted to NCBI/GEO (GSE148227). scRNA-seq data have been submitted to NCBI/GEO (GSE147319). Source data are provided with this paper.

## Code availability

Code and custom scripts for scRNA-seq analysis have been deposited on [https://github.com/theislab/2021\\_Aliuliev\\_Tritschler\\_gut\\_HFD](https://github.com/theislab/2021_Aliuliev_Tritschler_gut_HFD).

Received: 27 March 2020; Accepted: 13 August 2021;

Published online: 22 September 2021

## References

- Schauer, P. R. et al. Bariatric surgery versus intensive medical therapy in obese patients with diabetes. *N. Engl. J. Med.* **366**, 1567–1576 (2012).
- Evers, S. S., Sandoval, D. A. & Seeley, R. J. The physiology and molecular underpinnings of the effects of bariatric surgery on obesity and diabetes. *Annu. Rev. Physiol.* **79**, 313–334 (2017).
- Drucker, D. J. Mechanisms of action and therapeutic application of glucagon-like peptide-1. *Cell Metab.* **27**, 740–756 (2018).
- Gribble, F. M. & Reimann, F. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. *Nat. Rev. Endocrinol.* **15**, 226–237 (2019).
- Yilmaz, Ö. H. et al. mTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* **486**, 490–495 (2012).
- Beyaz, S. et al. High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58 (2016).
- Mihaylova, M. M. et al. Fasting activates fatty acid oxidation to enhance intestinal stem cell function during homeostasis and aging. *Cell Stem Cell* **22**, 769–778 (2018).
- Le Gall, M. et al. Intestinal plasticity in response to nutrition and gastrointestinal surgery. *Nutr. Rev.* **77**, 129–143 (2019).
- Lean, M. E. J. & Malkova, D. Altered gut and adipose tissue hormones in overweight and obese individuals: cause or consequence. *Int. J. Obes.* **40**, 622–632 (2016).
- Dailey, M. J. Nutrient-induced intestinal adaption and its effect in obesity. *Physiol. Behav.* **136**, 74–78 (2014).
- Verdam, F. J. et al. Small intestinal alterations in severely obese hyperglycemic subjects. *J. Clin. Endocrinol. Metab.* **96**, E379–E383 (2011).
- Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).
- Clevers, H. & Watt, F. M. Defining adult stem cells by function, not by phenotype. *Annu. Rev. Biochem.* **87**, 1015–1027 (2018).
- Petit, V. et al. Chronic high-fat diet affects intestinal fat absorption and postprandial triglyceride levels in the mouse. *J. Lipid Res.* **48**, 278–287 (2007).
- Beuling, E. et al. GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice. *Gastroenterology* **140**, 1219–1229 (2011).
- Aronson, B. E., Stapleton, K. A. & Krasinski, S. D. Role of GATA factors in development, differentiation, and homeostasis of the small intestinal epithelium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306**, G474–G490 (2014).

17. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
18. Moor, A. E. et al. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell* **175**, 1156–1167 (2018).
19. Beumer, J. et al. Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. *Nat. Cell Biol.* **20**, 909–916 (2018).
20. Kraiczky, J. et al. DNA methylation defines regional identity of human intestinal epithelial organoids and undergoes dynamic changes during development. *Gut* **68**, 49–61 (2019).
21. Middendorp, S. et al. Adult stem cells in the small intestine are intrinsically programmed with their location-specific function. *Stem Cells* **32**, 1083–1091 (2014).
22. Burtscher, I., Barkey, W. & Lickert, H. Foxa2-venus fusion reporter mouse line allows live-cell analysis of endoderm-derived organ formation. *Genesis* **51**, 596–604 (2013).
23. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
24. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, N. & Luo, L. A global double-fluorescent cre reporter mouse. *Genesis* **45**, 593–605 (2007).
25. Imuta, Y., Kiyonari, H., Jang, C. W., Behringer, R. R. & Sasaki, H. Generation of knock-in mice that express nuclear enhanced green fluorescent protein and tamoxifen-inducible Cre recombinase in the notochord from Foxa2 and T loci. *Genesis* **51**, 210–218 (2013).
26. Parker, H. E., Gribble, F. M. & Reimann, F. The role of gut endocrine cells in control of metabolism and appetite. *Exp. Physiol.* **99**, 1116–1120 (2014).
27. Clemmensen, C. et al. Gut–brain cross-talk in metabolic control. *Cell* **168**, 758–774 (2017).
28. Gehart, H. et al. Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. *Cell* **176**, 1158–1173 (2019).
29. Buczacck, S. J. A. et al. Intestinal label-retaining cells are secretory precursors expressing *Igr5*. *Nature* **495**, 65–69 (2013).
30. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
31. Piccand, J. et al. Rfx6 promotes the differentiation of peptide-secreting enteroendocrine cells while repressing genetic programs controlling serotonin production. *Mol. Metab.* **29**, 24–39 (2019).
32. Martin, A. M. et al. The diverse metabolic roles of peripheral serotonin. *Endocrinology* **158**, 1049–1063 (2017).
33. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Igr5*. *Nature* **449**, 1003–1007 (2007).
34. Chandel, N. S., Jasper, H., Ho, T. T. & Passequé, E. Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nat. Cell Biol.* **18**, 823–832 (2016).
35. Ito, K., Bonora, M. & Ito, K. Metabolism as master of hematopoietic stem cell fate. *Int. J. Hematol.* **109**, 18–27 (2019).
36. Dahly, E. M., Guo, Z. & Ney, D. M. Alterations in enterocyte proliferation and apoptosis accompany TPN-induced mucosal hypoplasia and IGF-I-induced hyperplasia in rats. *J. Nutr.* **132**, 2010–2014 (2002).
37. Mah, A. T. et al. Impact of diet-induced obesity on intestinal stem cells: hyperproliferation but impaired intrinsic function that requires insulin/IGF1. *Endocrinology* **155**, 3302–3314 (2014).
38. Mao, J. et al. Overnutrition stimulates intestinal epithelium proliferation through  $\beta$ -catenin signaling in obese mice. *Diabetes* **62**, 3736–3746 (2013).
39. Patel, P. & Woodgett, J. R. Glycogen synthase kinase 3: a kinase for all pathways? *Curr. Top. Dev. Biol.* **123**, 277–302 (2017).
40. Fevr, T., Robine, S., Louvard, D. & Huelsken, J. Wnt/ $\beta$ -catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. *Mol. Cell Biol.* **27**, 7551–7559 (2007).
41. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17**, 1709–1713 (2003).
42. Aichler, M. et al. N-acyl taurines and acylcarnitines cause an imbalance in insulin synthesis and secretion provoking  $\beta$  cell dysfunction in type 2 diabetes. *Cell Metab.* **25**, 1334–1347 (2017).
43. Shao, W. & Espenshade, P. J. Expanding roles for SREBP in metabolism. *Cell Metab.* **16**, 414–419 (2012).
44. Mihaylova, M. M., Sabatini, D. M. & Yilmaz, Ö. H. Dietary and metabolic control of stem cell function in physiology and cancer. *Cell Stem Cell* **14**, 292–305 (2014).
45. Gao, Y. et al. LKB1 represses ATOH1 via PDK4 and energy metabolism and regulates intestinal stem cell fate. *Gastroenterology* **158**, 1389–1401 (2020).
46. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
47. Nusse, R. & Clevers, H. Wnt/ $\beta$ -Catenin signaling, disease, and emerging therapeutic modalities. *Cell* **169**, 985–999 (2017).
48. Postic, C. & Girard, J. Contribution of de novo fatty acid synthesis to hepatic steatosis and insulin resistance: lessons from genetically engineered mice. *J. Clin. Investig.* **118**, 829–838 (2008).
49. Ipsen, D. H., Lykkesfeldt, J. & Tveden-Nyborg, P. Molecular mechanisms of hepatic lipid accumulation in non-alcoholic fatty liver disease. *Cell. Mol. Life Sci.* **75**, 3313–3327 (2018).
50. Kawano, Y. & Cohen, D. E. Mechanisms of hepatic triglyceride accumulation in non-alcoholic fatty liver disease. *J. Gastroenterol.* **48**, 434–441 (2013).
51. Namkung, J., Kim, H. & Park, S. Peripheral serotonin: a new player in systemic energy homeostasis. *Molecules Cells* **38**, 1023–1028 (2015).
52. Watanabe, H. et al. Serotonin improves high fat diet induced obesity in mice. *PLoS ONE* **11**, e0147143 (2016).
53. Poher, A. L., Tschöp, M. H. & Müller, T. D. Ghrelin regulation of glucose metabolism. *Peptides* **100**, 236–242 (2018).
54. Tschöp, M. et al. Circulating ghrelin levels are decreased in human obesity. *Diabetes* **50**, 707–709 (2001).
55. Papaetis, G. S. Incretin-based therapies in prediabetes: current evidence and future perspectives. *World J. Diabetes* **5**, 817–834 (2014).
56. Matthews, D. R. et al. Homeostasis model assessment: insulin resistance and  $\beta$ -cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
57. Nakajima, S., Hira, T. & Hara, H. Postprandial glucagon-like peptide-1 secretion is increased during the progression of glucose intolerance and obesity in high-fat/high-sucrose diet-fed rats. *Br. J. Nutr.* <https://doi.org/10.1017/S0007114515000550> (2015).
58. Chambers, A. P. et al. Weight-independent changes in blood glucose homeostasis after gastric bypass or vertical sleeve gastrectomy in rats. *Gastroenterology* **141**, 950–958 (2011).
59. Panaro, B. L. et al. Intestine-selective reduction of Gcg expression reveals the importance of the distal gut for GLP-1 secretion. *Mol. Metab.* **37**, 100990 (2020).
60. Andersson-Rolf, A., Fink, J., Mustata, R. C. & Koo, B. K. A video protocol of retroviral infection in primary intestinal Organoid culture. *J. Vis. Exp.* <https://doi.org/10.3791/51765> (2014).
61. R Development Core Team, R. Lecture Notes in Physics, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing <https://doi.org/10.1007/978-3-540-74686-7> (2011).
62. Rainer, J., Sanchez-Cabo, F., Stocker, G., Sturm, A. & Trajanoski, Z. CARMAweb: Comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res.* **34**, W498–W503 (2006).
63. Mantini, D. et al. LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinforma.* **8**, 101 (2007).

## Acknowledgements

We thank J. Schultheiss, K. Diemer, I. Kunze, S. Felicitas Baumgarten, A. Raducanu, A. Theis, A. Bettenbrock and S. Straub for excellent technical assistance. We thank B. Schubert and E. Spiegel for their input and help for the compositional analysis and M. Büttner, A. Wolf, L. Simon and D. Fischer for fruitful discussions on the computational analysis. We are grateful to H. Edlund for providing the *Ngn3* antibody. We thank M. Bakhti and R. Böttcher for valuable comments on the manuscript. This work was funded by the Helmholtz Alliance ICEMED: Imaging and Curing Environmental Metabolic Diseases (H.L., K.S. and M.H.T.), through the Initiative and Networking Fund of the Helmholtz Association (H.L.) and the Deutsche Forschungsgemeinschaft (SFB 824 C4 and CRC/Transregio 205/1) (A.W.). For financial support we also thank the Helmholtz Society, the Helmholtz Portfolio Theme ‘Metabolic Dysfunction and Common Disease’ (H.L., K.S. and M.H.T.), the Helmholtz Alliance ‘Aging and Metabolic Programming, AMPro’ (H.L. and J.B.). Further, this project was funded by ExNet-0041-Phase2–3 (‘SyNergy-HMGU’) through the Initiative and 354Network Fund of the Helmholtz Association (H.L. and F.J.T.), the German Research Foundation and the German Center for Diabetes Research (H.L., K.S. and M.H.T.), the Alexander von Humboldt Foundation (M.H.T.), Helmholtz Initiative on Personalized Medicine iMed by Helmholtz Association (M.H.T.), the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation Grants DFG-TS226/1-1, DFG-TS226/3-1 and DFG-TR152 (project no. 239283807) (M.H.T.) and DFG-SFB 1321 (K.S. and M.H.T.) and the European Research Council (AdG HypoFlam 695054) (M.H.T.). S.T. is supported by a DFG Fellowship through the Graduate School of Quantitative Biosciences Munich.

## Author contributions

A.A. performed all rodent experiments, intestinal histology and western blots, and analysed and interpreted all data, interpreted scRNA-seq data and co-wrote the manuscript. S.T. analysed and interpreted scRNA-seq data and co-wrote the manuscript. M.S. analysed microarray data. L.O. analysed data and performed in vivo experiments. J.H. performed intestinal histology and analysed histology data. T.G. performed RNAscope staining. M.I. and J.B. performed microarray analysis. N.S. and A.W. performed MALDI-MSI and analysed data. K.S. helped with metabolic profiling in mice. M.H.T. helped to interpret data. A.K. and J.K. integrated and analysed microarray data and metabolomics data. M.D.L. supported scRNA-seq analysis. F.J.T. conceptualized, supervised and interpreted the scRNA-seq analysis and helped to write the manuscript. H.L. acquired funding and conceptualized, designed, supervised and interpreted all studies, co-wrote the manuscript and is the lead contact for this work. A.B. conceptualized, designed, supervised and interpreted all studies, performed experiments and co-wrote the manuscript.

## Funding

Open access funding provided by Helmholtz Zentrum München - Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH) (4209).

## Competing interests

M.H.T. serves as a scientific advisory board member of ERX Pharmaceuticals, Inc, Cambridge, MA. F.J.T. reports receiving consulting fees from Cellarity, Inc and ownership interest in Cellarity, Inc. and Dermagnostix. S.T. is an employee of Cellarity, Inc. and has stake-holder interests; the present work was carried out as an employee of Helmholtz Munich. The remaining authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42255-021-00458-9>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42255-021-00458-9>.

**Correspondence and requests for materials** should be addressed to Fabian J. Theis, Heiko Lickert or Anika Böttcher.

**Peer review information** Primary handling editors: Elena Bellafante and Isabella Samuelson. *Nature Metabolism* thanks Shalev Itzkovitz and the other, anonymous, reviewers for their contribution to the peer review of this work.

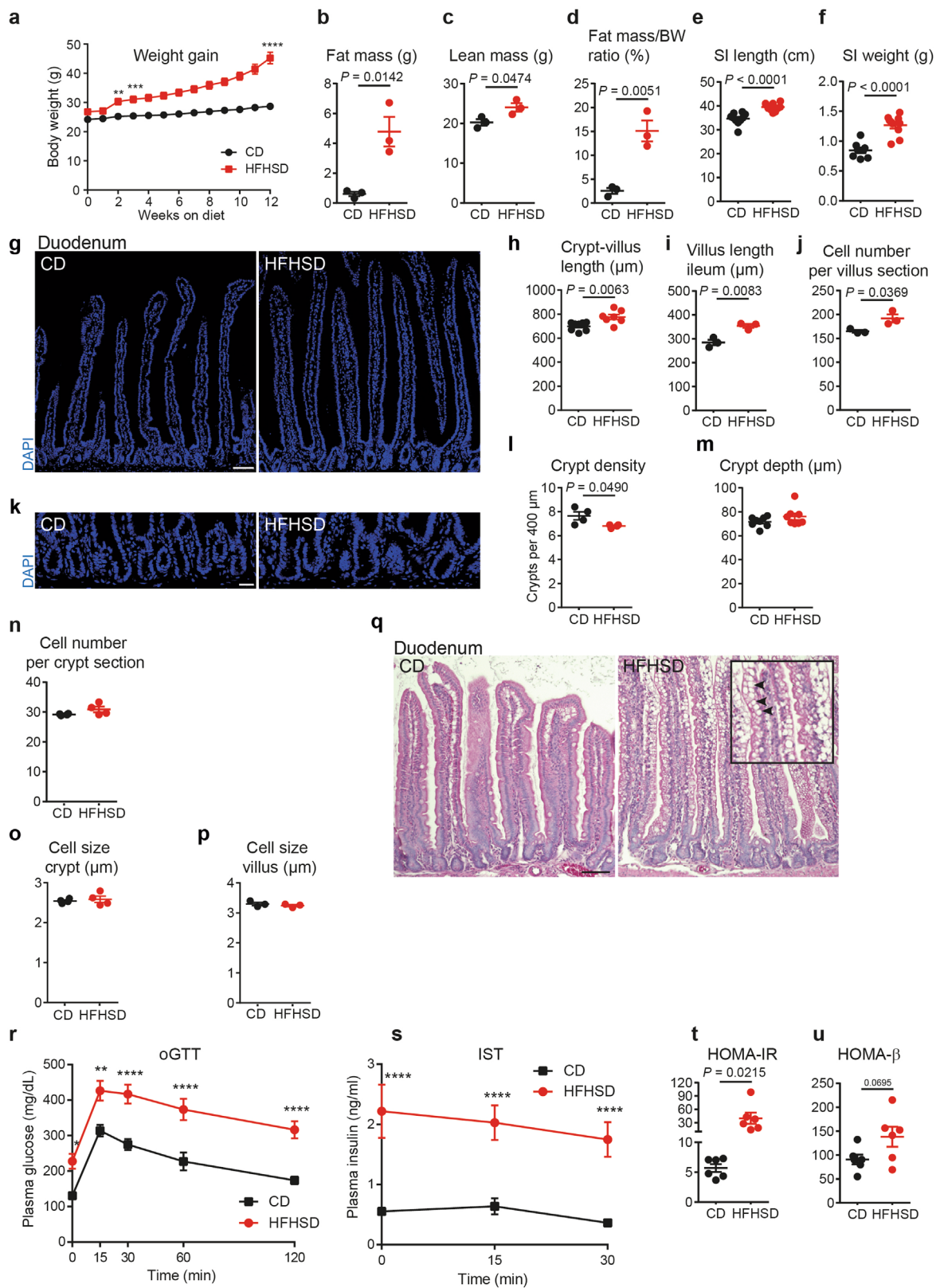
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



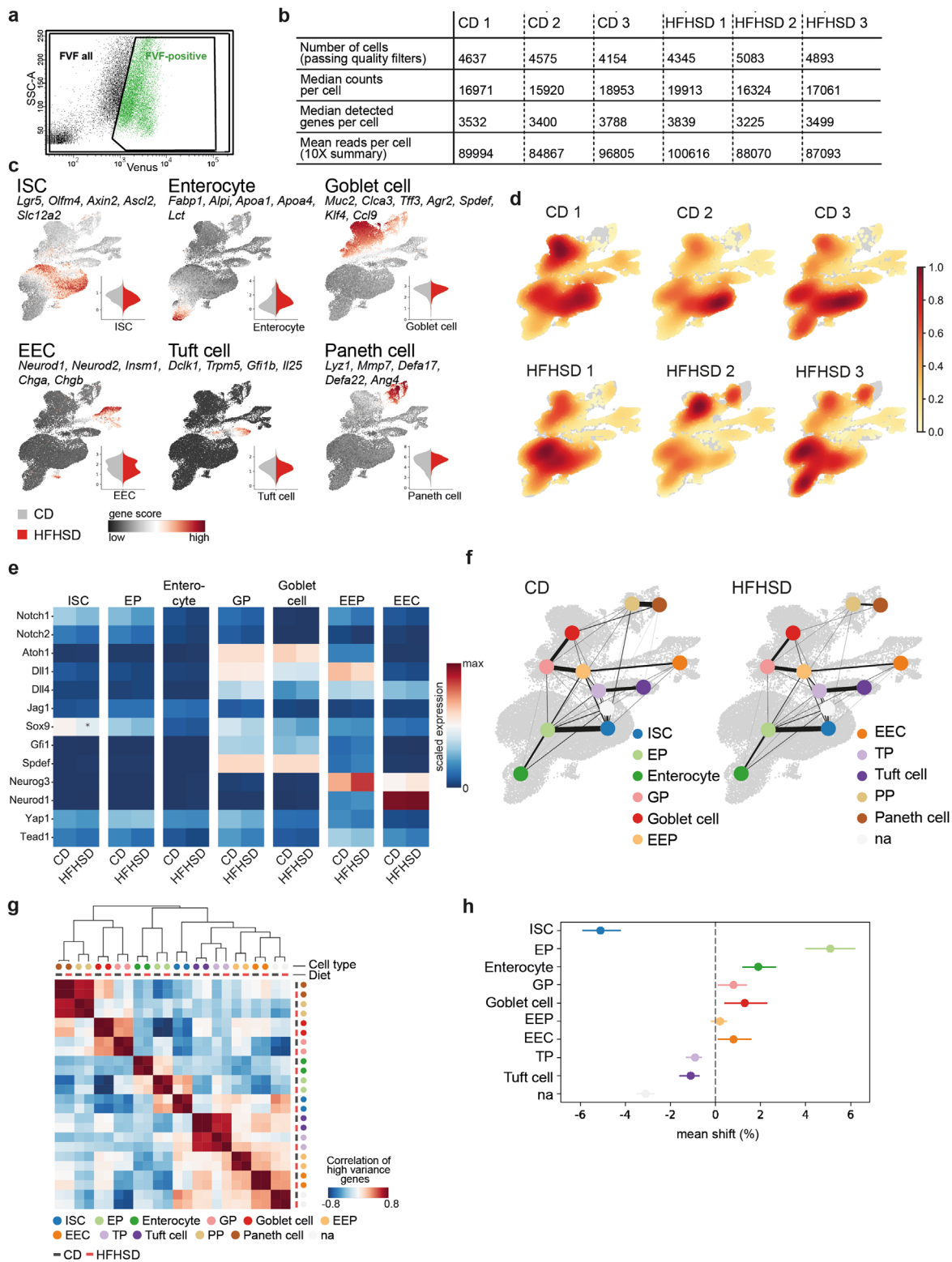
**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021



Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | HFHSD remodels the intestinal mucosa and leads to obesity and prediabetes.** **a**, Weekly assessment of body weight.  $n = 8$  mice per group. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-way ANOVA with Sidak's multiple comparison test. **b-d**, Analysis of fat (**b**) and lean mass (**c**) by nuclear magnetic resonance (NMR) and fat mass ratio (**d**).  $n = 3$  mice per group. Data are shown as mean  $\pm$  s.e.m.. Statistical significance was determined by two-tailed Student's t-test. **e, f**, Assessment of small intestinal length (**e**,  $n = 10$  mice per group) and weight (**f**,  $n = 8$  mice versus 10 mice per group). Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **g-m**, Representative LSM images of duodenal sections depicting crypt-villus units (**g**) or crypt regions (**k**). Assessment of crypt and villus morphology in duodenal and ileal sections. Duodenum: **h, j, l, m**. Ileum: **i**,  $n = 8$  mice versus 7 mice (**h**),  $n = 3$  mice (**i, j**),  $n = 4$  mice (**l**) and  $n = 8$  mice (**m**) per group. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. Scale bars,  $75 \mu\text{m}$  (**g**),  $25 \mu\text{m}$  (**k**). **n-p**, Cell number (**n**) and size (**o**) quantification in the crypts ( $n = 4$  mice per group) and cell size quantification in the villi (**p**,  $n = 3$  mice per group). Data are mean  $\pm$  s.e.m. of biologically independent samples. **q**, Histological assessment of duodenal tissue sections by haematoxylin and eosin staining. Black square depicts an enlarged villus section. Black arrowheads indicate fat inclusions. Scale bar,  $100 \mu\text{m}$ .  $n = 3$  biologically independent CD and HFHSD samples. **r, s**, Assessment of glucose tolerance (**r**,  $n = 8$  per group) by oral glucose tolerance test (oGTT) and insulin secretion (**s**,  $n = 4$  mice versus 5 mice per group) by insulin secretion test (IST). Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-way ANOVA with Sidak's multiple comparison test. **t, u**, Estimation of insulin resistance ((HOMA-IR) and  $\beta$ -cell function (HOMA- $\beta$ ).  $n = 6$  mice per group. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test.

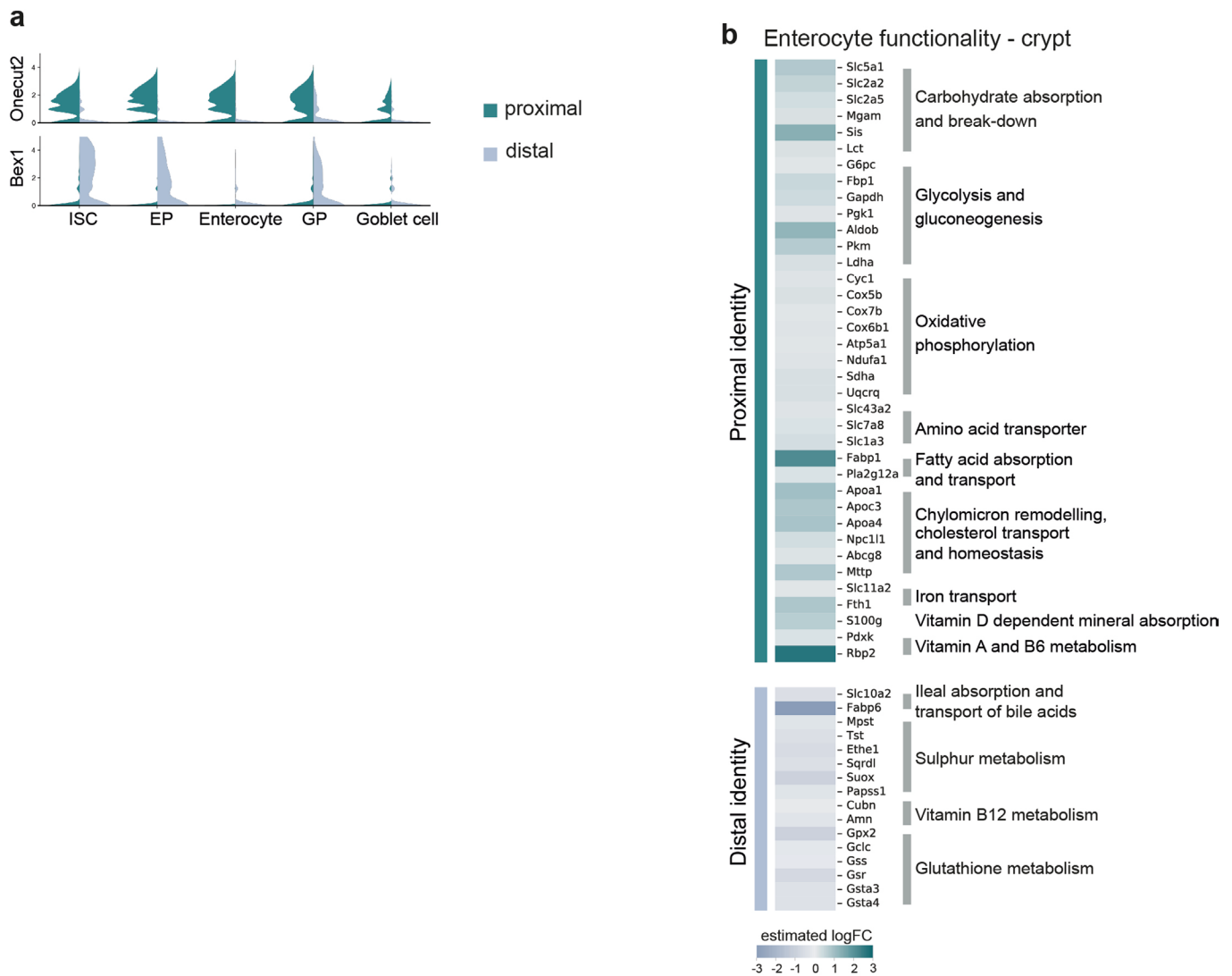


Extended Data Fig. 2 | See next page for caption.

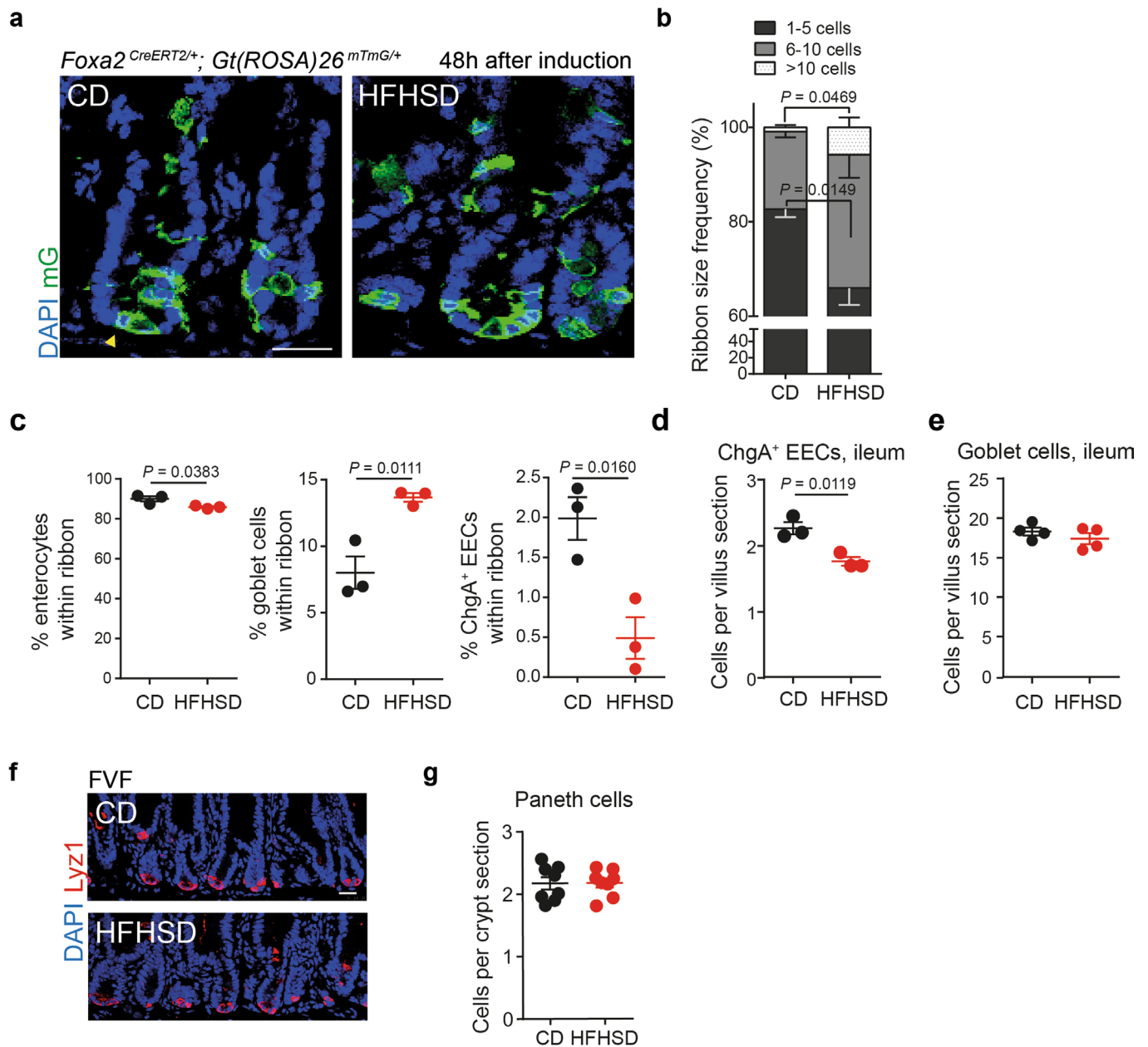
**Extended Data Fig. 2 | Characterization of small intestinal epithelial cell types and lineage relations in CD- and HFHSD-conditions by scRNA-seq.**

**a**, Representative flow cytometry plot for FVF-based enrichment of SI crypt cells for scRNA-seq. Gating strategy distinguishes FVF-positive cells from random live whole crypt cells (FVF all). **b**, Quality metrics for the scRNA-seq data. **c**, Levels of cell scores calculated based on the expression of a set of known marker genes specific for each cell type and used for cluster annotation. Violin plots show distribution of score levels assigned to the respective cell type in CD- and HFHSD-derived samples. **d**, UMAP plots of crypt cells colored by cell densities of individual CD and HFHSD mice. **e**, Mean expression levels of selected genes known to regulate lineage decisions in intestinal crypt cells of CD and HFHSD mice. \*indicates a significant change in HFHSD mice (limma, FDR < 0.01, logFC > 0.1). **f**, Lineage relations in CD- and HFHSD-derived SI crypt cells inferred based on a measure for cluster connectivity using PAGA. Edges are weighted by significance, node positions are cluster centres in UMAP space. **g**, Pearson correlation of mean expression profiles of high variance genes broken down by cell type and diet. Rows and columns are ordered using hierarchical clustering. **h**, Frequency shifts of each cluster from HFHSD-derived samples compared to CD-derived samples estimated using a Dirichlet-Multinomial model. Expected mean and the 95% highest posterior density interval (HPD-region) of the frequency shift for each cell type is shown. A credible shift is assumed if the HPD-region does not include 0.  $n = 3$  CD mice.  $n = 3$  HFHSD mice.



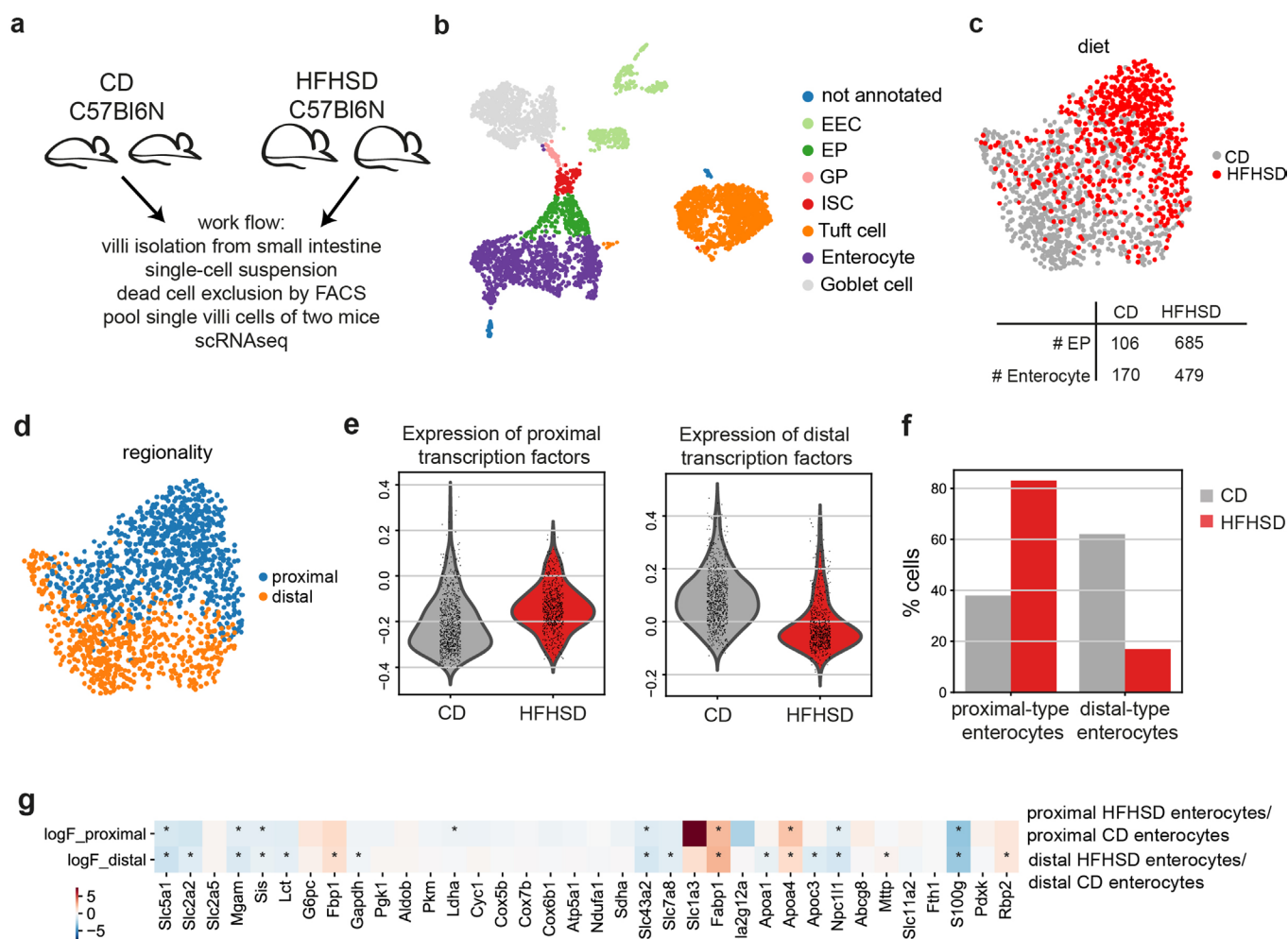


**Extended Data Fig. 3 | Proximal- and distal-type enterocytes exert different functions.** **a**, Violin plots showing expression distribution of regional markers *Onecut2* (proximal) and *Bex1* (distal) used to classify proximal and distal cells<sup>17</sup> (see Methods). **b**, Genes and their associated function characteristic for enterocytes with proximal or distal identity, respectively. Heatmap shows estimated fold change in proximal versus distal cells from CD-derived samples.

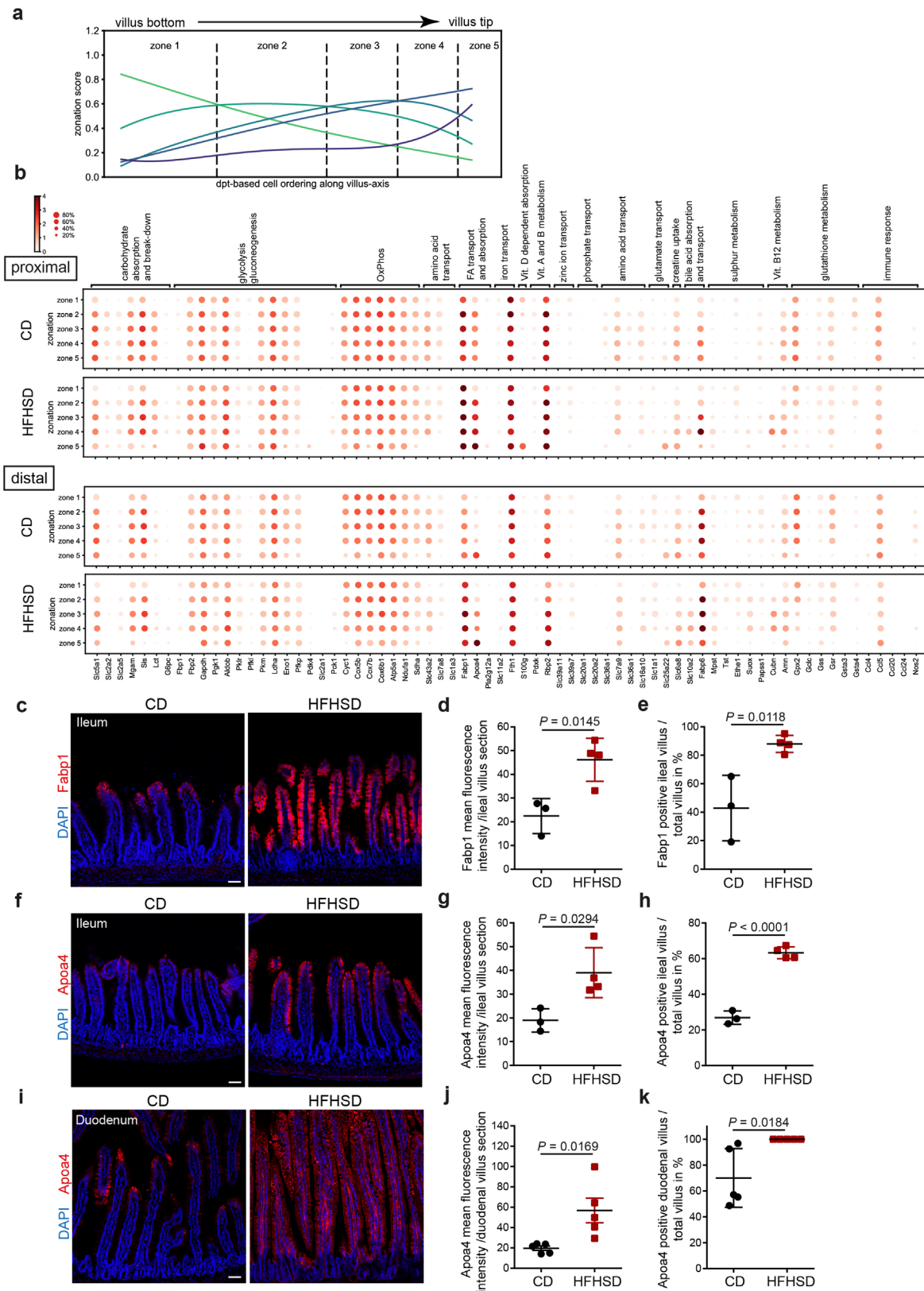


**Extended Data Fig. 4 | Analysis of *Foxa2* lineage positive cells and frequency of mature intestinal cell types in FVF mice in CD and HFHSD conditions.**

**a**, Representative LSM images of tamoxifen-induced recombination in duodenal crypts of CD- and HFHSD-fed *Foxa2<sup>CreERT2/+</sup>; Gt(Rosa26)<sup>mTmG/+</sup>* mice after 48 h. Yellow arrows indicate recombined mG<sup>+</sup> slender crypt base cells (CBC) at positions typical for ISCs, adjacent to pyramid-shaped Paneth cells. Scale bar, 25  $\mu$ m.  $n = 3$  biologically independent CD and HFHSD samples. **b**, Frequencies of coherent *Foxa2*-lineage ribbons of different sizes in the villi of CD- and HFHSD-fed *Foxa2<sup>CreERT2/+</sup>; Gt(Rosa26)<sup>mTmG/+</sup>* mice at 70 h after tamoxifen induction.  $n = 3$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **c**, Percentage of villin<sup>+</sup> enterocytes, Muc2<sup>+</sup> goblet cells and ChgA<sup>+</sup> EECs within *Foxa2* lineage ribbons.  $n = 3$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **d**, **e**, Quantification of ChgA<sup>+</sup> EECs (**d**,  $n = 3$  mice per group) and Muc2<sup>+</sup> goblet cells (**e**,  $n = 4$  mice per group) in ileal tissue sections of CD- and HFHSD-fed FVF mice. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **f**, **g**, Representative LSM images (**f**) and quantification (**g**) of Paneth cells (Lyz1<sup>+</sup>) in the duodenum. Scale bar, 25  $\mu$ m,  $n = 8$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples.

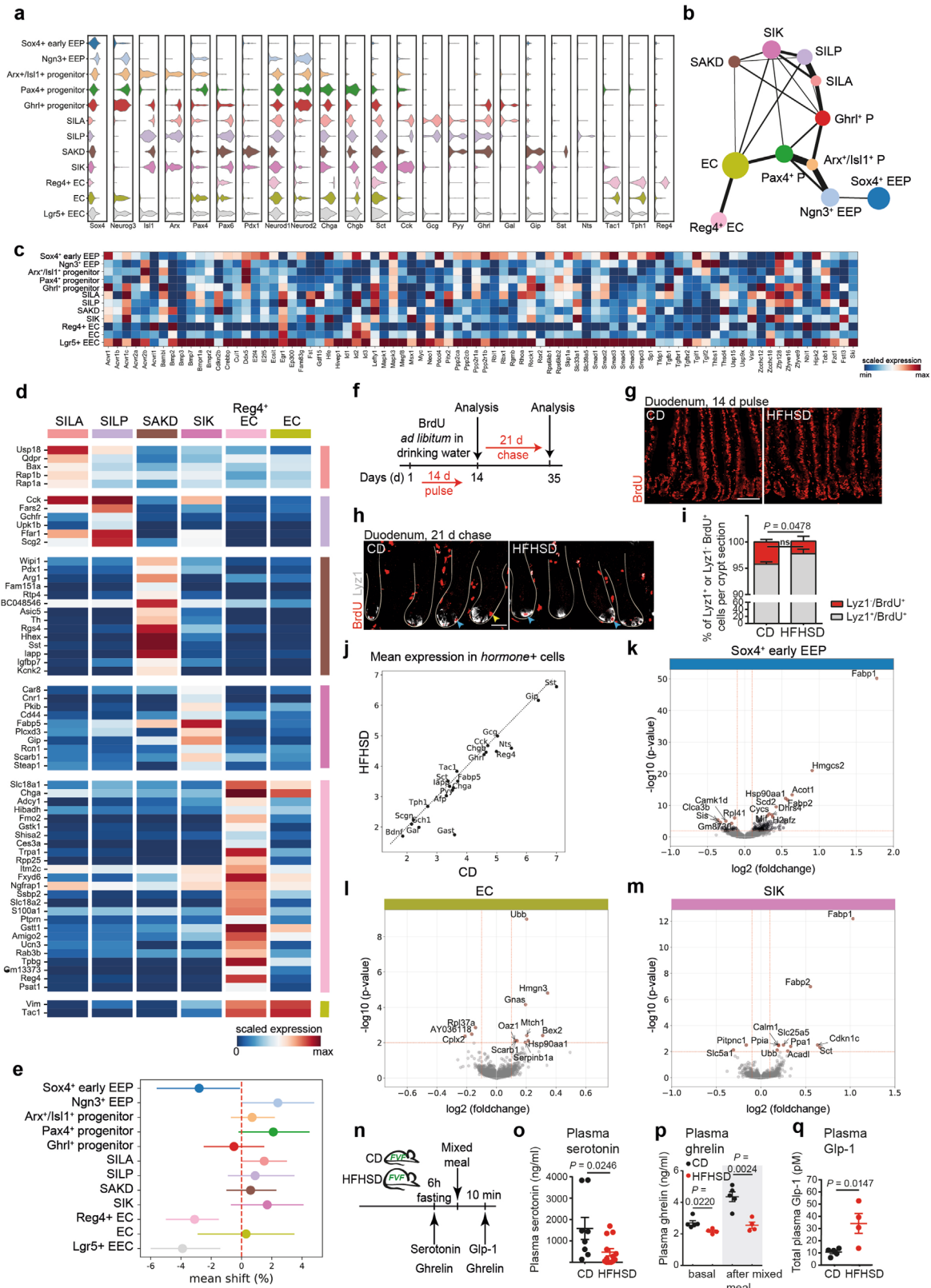


**Extended Data Fig. 5 | Regional identity of villi enterocytes is changed upon HFHSD.** **a**, Overview of the experimental setup for scRNA-seq of villi cells. **b**, UMAP plot of 4205 profiled single SI villi cells. Colors highlight clustering into major intestinal cell types based on expression of previously published marker genes<sup>17</sup>. **c**, UMAP plot of enterocyte progenitors (EP) and enterocytes computed using a set of 44 transcription factors previously described to be differential between proximal and distal gut regions<sup>17</sup>. Colors highlight diet. **d**, UMAP plot of EP and enterocytes computed using a set of 44 transcription factors previously described to be differential between proximal and distal gut regions<sup>17</sup>. Colors highlight regional annotation. **e**, Violin plots showing cell scores of the proximal and distal transcription factors used in **c**)<sup>17</sup> in CD and HFHSD-derived EPs and enterocytes. **f**, Proportions of enterocytes and EPs annotated as proximal- or distal-type enterocytes based on regional transcription factor scores in **c**). **g**, Heatmap showing the log<sub>2</sub>-fold change of the expression of selected functional proximal genes in HFHSD vs CD distal and proximal enterocyte cluster. \*indicates a significant change. The p-values are provided in Supplementary table 1.



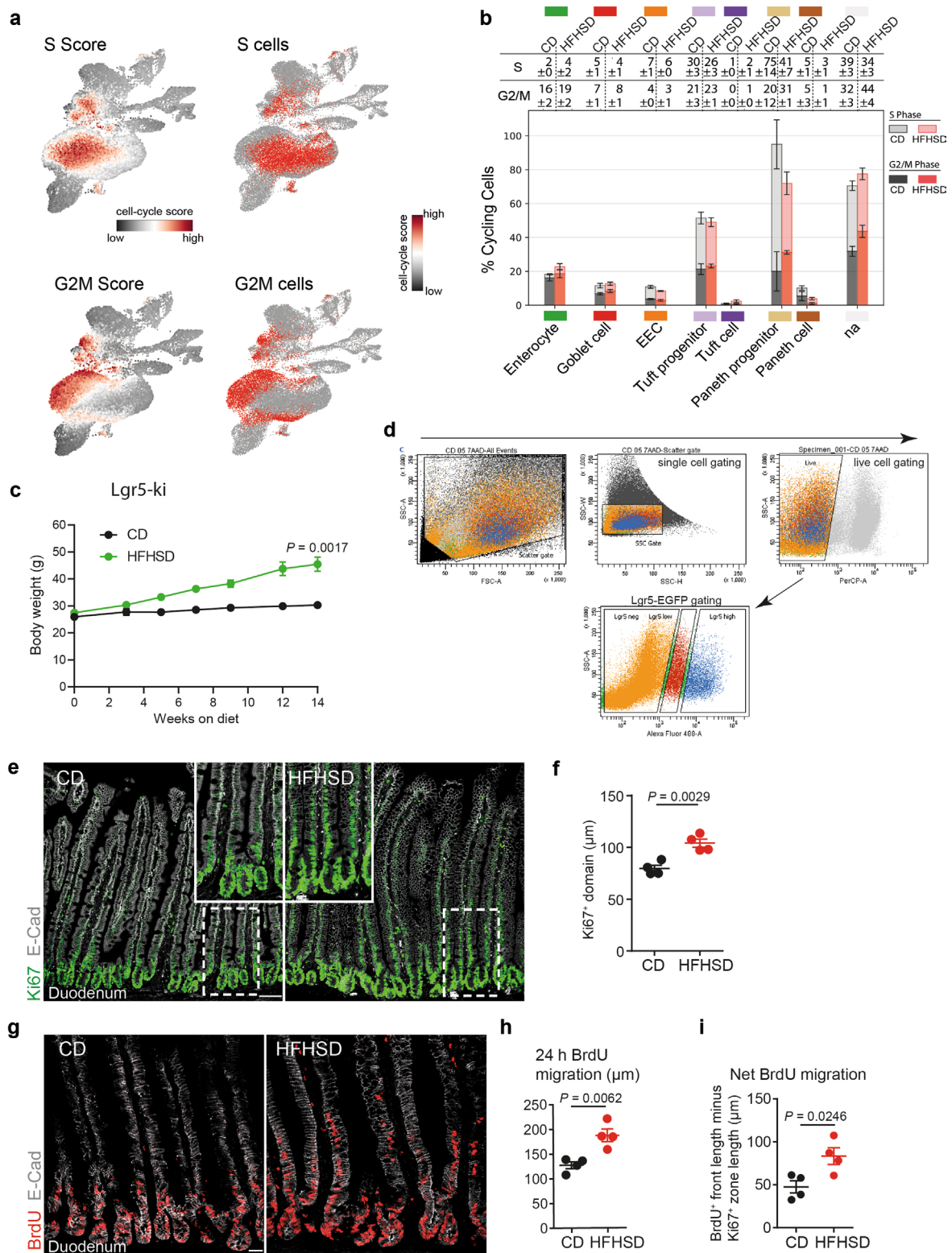
Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | Enterocyte zonation is altered upon HFHSD.** **a**, Activation of zonation scores along a pseudospacial ordering of EP and enterocytes from SI villi, which reflects the axis from villus bottom to tip. Cell scores for each zone were computed on reported markers<sup>18</sup> and approximated by polynomial regression fits along the pseudospacial axis. Crossing points of the fitted lines define the partitioning into the five zones. **b**, Dot plots showing the expression of selected markers of enterocyte function in distal and proximal enterocytes. Cells are partitioned into zones along the inferred pseudospacial axis from villus bottom (zone 1) to tip (zone 5). **c–e**, Representative LSM images (**c**) and quantification of Fabp1 mean fluorescence intensity in ileal villi (**d**) and determination of the length of the Fabp1 positive zone (**e**). Scale bar, 50  $\mu\text{m}$ ,  $n=3$  CD mice and  $n=4$  HFHSD mice. Data are shown as mean  $\pm$  s.d. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **f–h**, Representative LSM images (**f**) and quantification of ApoA4 mean fluorescence intensity in ileal villi (**g**) and determination of the length of the ApoA4 positive zone (**h**). Scale bar, 50  $\mu\text{m}$ ,  $n=3$  CD mice and  $n=4$  HFHSD mice. Data are shown as mean  $\pm$  s.d. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **i–k**, Representative LSM images (**i**) and quantification of ApoA4 mean fluorescence intensity in duodenal villi (**j**) and determination of the length of the ApoA4 positive zone (**k**). Scale bar, 50  $\mu\text{m}$ ,  $n=5$  mice per group. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test.



Extended Data Fig. 7 | See next page for caption.

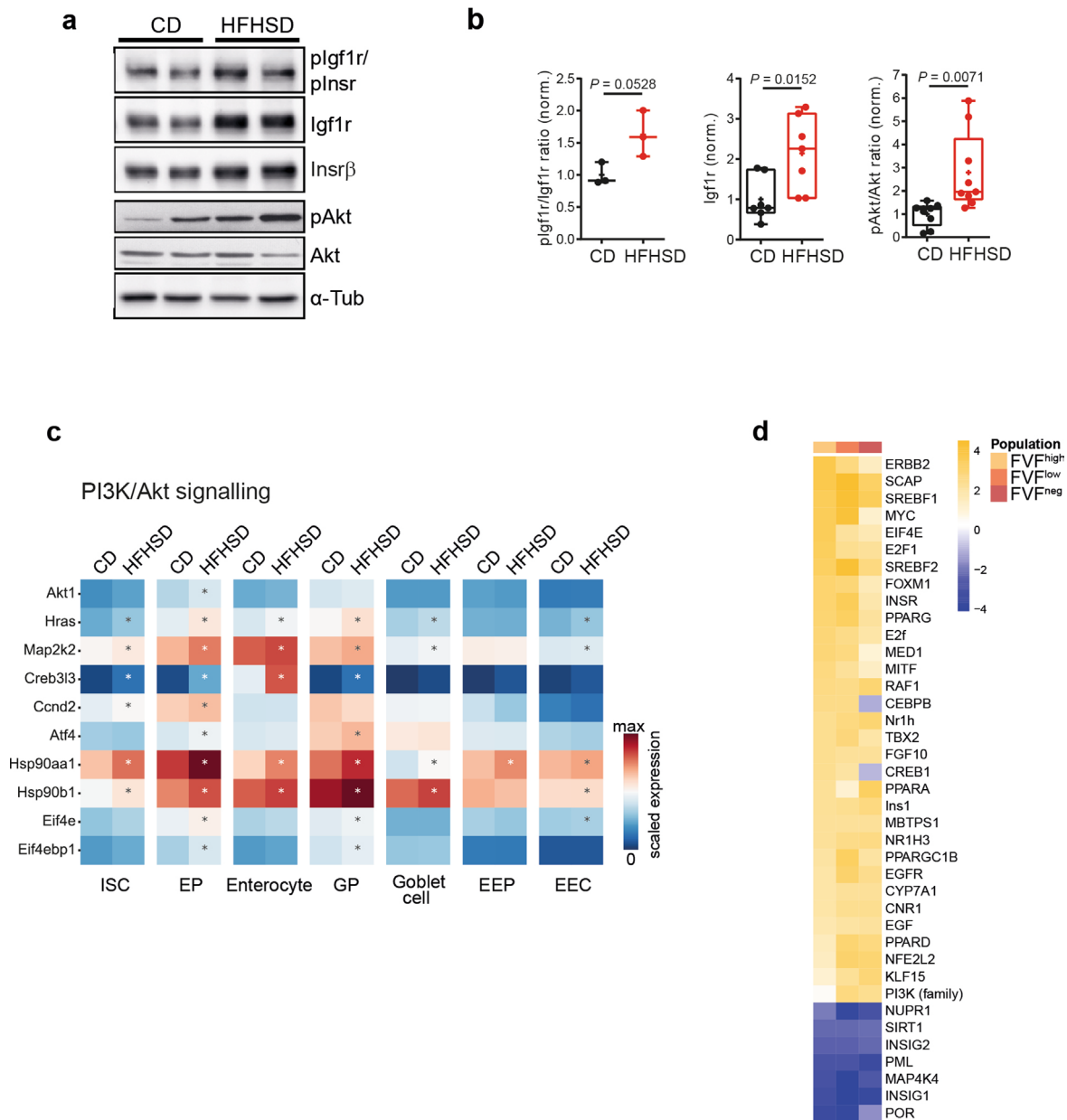
**Extended Data Fig. 7 | HFHSD changes transcription in EEC subsets and gut plasma hormone levels.** **a**, Violin plots showing the distribution of expression of selected genes in EEC clusters. **b**, PAGA plot showing lineage relationships for the endocrine lineage in CD-derived samples. **c**, Mean expression levels per cluster of genes associated with BMP signaling. **d**, Mean expression levels per cluster of subset-specific genes in EEC clusters. **e**, Frequency shifts of each cluster from HFHSD-derived samples compared to CD-derived samples estimated using a Dirichlet-Multinomial model. Expected mean and the 95% highest posterior density interval (HPD-region) of the frequency shift for each cell type is shown.  $n = 3$  mice per group. **f-i**, Abundance of label-retaining cells (LRCs). **f**, Schematic depicting the experimental strategy of the 5-bromo-2-deoxyuridine (BrdU) pulse-chase experiment to determine the LRC frequency. **g**, Representative LSM images of duodenal sections after 14 days of BrdU labelling. BrdU-labelled cells are stained in red. Scale bar,  $50 \mu\text{m}$ .  $n = 3$  biologically independent CD and HFHSD mice. **h**, Representative LSM images of duodenal crypts after a chase period of 21 days. Yellow arrowhead highlights a  $\text{BrdU}^+/\text{Lyz1}^-$  LRC. Blue arrowheads indicate  $\text{BrdU}^+/\text{Lyz1}^+$  cells. Scale bar,  $25 \mu\text{m}$ . **i**, Proportion of  $\text{Lyz1}^-$  and  $\text{Lyz1}^+$  BrdU-labelled cells of total  $\text{BrdU}^+$  cells in duodenal crypts.  $n = 4$  mice versus 5 mice per group. Data are presented as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **j**, Mean hormone expression levels in EECs. Mean was calculated only from cells with high expression levels ( $\log(\text{count} > 3)$ ). **k-m**, Differential gene expression in selected EEC clusters tested using limma. For each gene, the estimated logFC and its associated FDR is plotted. Red lines indicate thresholds used for significance level and gene expression change. Regulated genes are highlighted in black, of which annotated genes are the top 10 genes ranked by FDR. **n**, Experimental scheme for the assessment of selected hormone plasma levels. **o-q**, Plasma levels of serotonin ( $n = 4$  mice per group) (**o**), ghrelin ( $n = 5$  mice / 4 mice per group) (**p**) and Glp-1 ( $n = 5$  mice / 4 mice per group) (**q**). Data represent mean  $\pm$  s.e.m. of biologically independent mice. Statistical significance was determined by two-tailed Student's t-test.



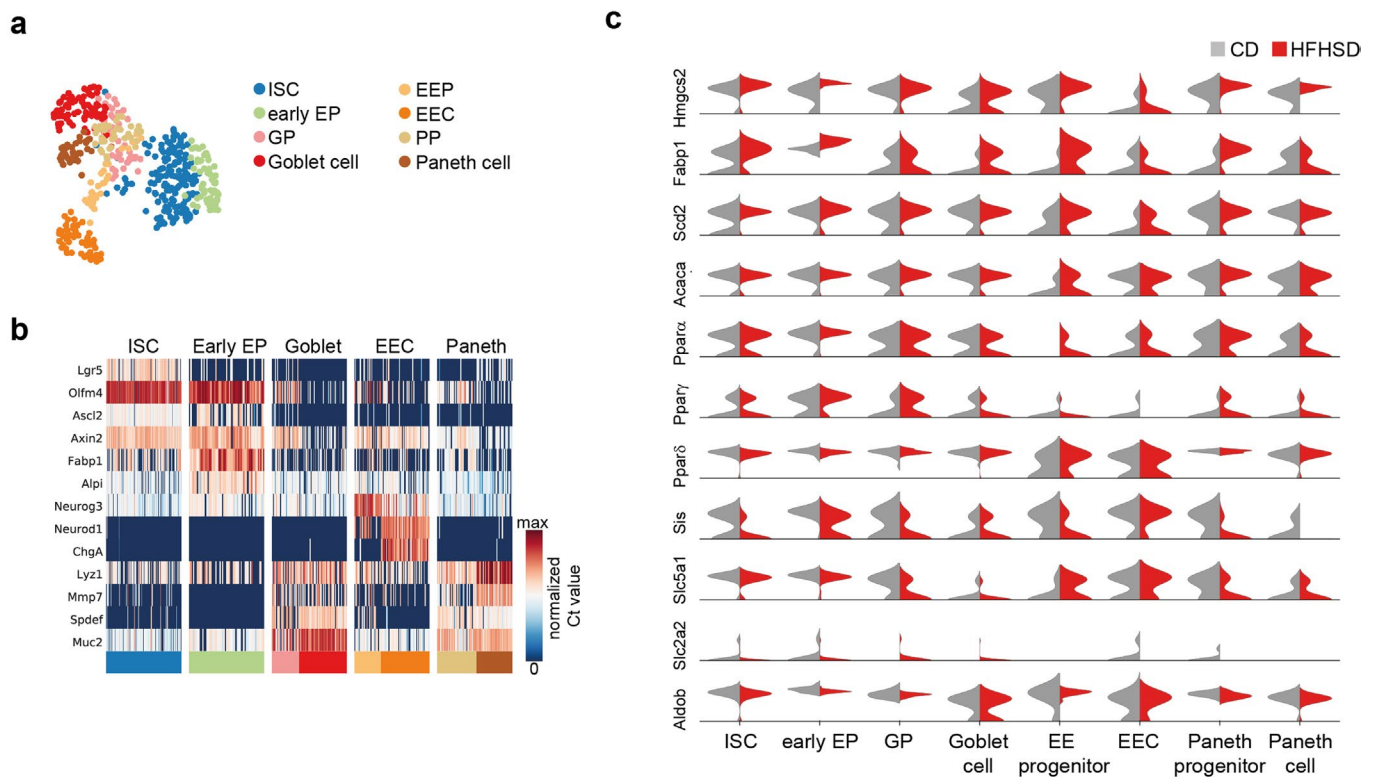
Extended Data Fig. 8 | See next page for caption.



**Extended Data Fig. 8 | HFHSD increases turnover of SI crypt cells.** **a**, Left: Cell score levels for cell-cycle phases S and G2/M calculated based on the expression of a set of genes related to cell-cycle are visualized in a UMAP plot. Right: Cells assigned to the corresponding cell-cycle phase are highlighted in red. **b**, Proportions of cycling cells assigned to S and G2/M cell-cycle phases in CD and HFHSD-derived in indicated clusters. Table indicates percentages. Data are shown as mean  $\pm$  s.e.m.,  $n=3$  mice per group. **c**, Assessment of body weight over the course of 14 weeks of CD- and HFHSD-fed Lgr5-ki mice.  $n=4$  per group. Data are shown as mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **d**, Representative FACS plots showing gating strategy to determine the proportion of Lgr5-EGFP<sup>hi</sup> cells (ISCs) in CD and HFHSD fed Lgr5-ki mice. **e, f**, Representative LSM images (**e**) and quantification (**f**) of Ki67<sup>+</sup> (green) proliferative regions. Scale bar, 75  $\mu$ m,  $n=4$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test. **g-i**, Representative LSM images of BrdU-labelled cells (BrdU<sup>+</sup> nuclei, red) along the crypt-villus axis 24 h after BrdU administration in duodenal sections from CD- and HFHSD-fed FVF mice (**g**). Quantification of cell migration length by measuring the distance from the crypt base to the highest labelled cell in the villus (**h**). Net BrdU migration (**i**), calculated as the average distance of BrdU-labelled migration front minus the Ki67<sup>+</sup> zone to correct for increased cell proliferation on HFHSD. Scale bar, 25  $\mu$ m.  $n=4$  mice per group. Data are mean  $\pm$  s.e.m. of biologically independent samples. Statistical significance was determined by two-tailed Student's t-test.



**Extended Data Fig. 9 | HFHSD stimulates the Insr/Igf1r/Akt pathway and fatty acid synthesis and Ppar signaling.** **a, b**, Representative immunoblots of phosphorylated (p)Igf1r/Insr (Tyr1158/Tyr1162/Tyr1163) and total Igf1r, Insrβ, pAkt (Ser473) and total Akt (**a**) and quantification of the protein levels (**b**). Protein expression was analysed in whole-cell lysates from SI crypts of CD- and HFHSD-fed FVF animals. For quantification, signal intensities of all protein bands were normalised to α-tubulin; phosphorylated proteins were additionally normalised to the corresponding non-phosphorylated protein.  $n = 3$  mice per group for pIgf1r/pInsr,  $n = 7$  mice per group for Igf1r,  $n = 9$  mice per group for (p)Akt and total Akt. Data are presented as box-and-whisker plots. The lower and upper boundaries of the boxes represent the 25th and 75th percentiles, respectively. The center lines indicate the medians, the crosses denote the mean values, and whiskers represent the maximum and minimum values. Statistical significance was determined by two-tailed Student's t-test. Circles represent biological independent samples. **c**, Mean expression levels per cluster of genes involved in PI3K/Akt signaling in CD and HFHSD. \* indicates a significant change in HFHSD mice (limma, FDR < 0.01, logFC > 0.1). The p-values are provided in Supplementary table 1. **d**, Predicted activation or inhibition of upstream regulators in bulk transcriptomics. Significantly regulated genes from each population (each HFHSD versus CD) were analysed by Ingenuity Pathway Software to predict the activation (yellow; z-score > 2) or inhibition (blue; z-score < -2) of upstream regulators. Shown are all genes with three z-scores available and from which at least two were significant (z-score > 1.2).



**Extended Data Fig. 10 | Confirmation of expression differences in metabolic genes in FVF-positive cells from CD- and HFHSD-fed mice by targeted single-cell qPCR. a, b**, UMAP plot of 465 profiled FVF positive crypt cells from 3 CD and 3 HFHSD-fed mice coloured by cell type (**a**). Cluster annotation was based on lineage marker gene expression depicted in the heatmap (**b**). **c**, Violin plots showing expression of selected metabolic genes related to fatty acid synthesis, Ppar signaling and glycolysis in indicated single-cell clusters split by diet.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

1. Microsoft office excel 2016-2018
2. qPCR was carried out using the Vii7 Real Time PCR System (Thermo Fisher Scientific).
3. Microarray data was obtained using Agilent 2100 Bioanalyzer and Affymetrix Mouse Gene 1.0 ST arrays.
4. MALDI-MSI measurements were obtained with a Bruker Solarix 7T FT-ICR-MS (Bruker Daltonics, Bremen, Germany) and after matrix removal scanned with a MIRAX DESK digital slide-scanning system (Carl Zeiss MicroImaging, Göttingen, Germany).
5. Western blot analysis was carried out using UVP Chem Studio SA (Analytik Jena AG, Biometra GmbH Jena Germany).
6. Tissue histology images were obtained with a Zeiss AXIO Scope A1 microscope (Carl Zeiss AG, Germany).
7. Fluorescent histological images were obtained with a Leica SP5 Confocal microscope (Leica Microsystems, Germany) using the LAS AF software v2.6.0-7266.
8. single-cell qRT-PCR were run on 96 × 96 Dynamic Array on the BioMark System (Fluidigm). Ct values for each gene in each cell was calculated using BioMark Real-Time PCR Analysis software v3 (Fluidigm).
9. single-cell libraries were sequenced on the HiSeq4000 (Illumina) with 150 bp paired-end sequencing of read 2.
10. BD FACS ARIA III and BD FACSDiva software v6.1.3
11. EchoMRI, Houston, TX, USA for body composition analysis

#### Data analysis

1. Prism 6 software (Graphpad) <http://www.graphpad.com/scientific-software/> N/A prism/.
2. Microarray was processed with Expression Console (v.1.3.0.187, Affymetrix),
3. Spectra processing of MALDI-MSI data was done with FlexImaging v. 4.2.
4. MATLAB R2014b (v.7.10.0, Mathworks, Inc., Natick, MA) was used as MALDI spectral pre-processing tool for the subsequent data bioinformatics analysis, LIMPIC algorithm was used for peak picking
5. Metabolite annotation was performed with databases (METLIN, <http://metlin.scripps.edu/>; Human Metabolome Database, <http://www.hmdb.ca/>; MassTRIX, <http://masstrix3.helmholtz-muenchen.de/masstrix3/>; METASPACE, <http://annotate.metaspacesoftware.com/>).
6. Heatmap-based clustering and enrichment analysis of metabolic pathways were performed with MetaboAnalyst 3.0 (<http://www.metaboanalyst.ca/>).
7. Western blots protein bands were quantified with the ImageJ software v1.51.

8. Histological images were analyzed using the LAS AF software program v2.6.0-7266..
9. scRNA-seq data were analyzed using python 3.5.4, 3.6.12 and 3.7.5. The raw reads were processed using the CellRanger analysis pipeline (v2.0.0) provided by 10X Genomics. Subsequent analyzes were carried out using custom scripts, Scanpy (<https://github.com/theislab/scanpy>, v1.0.4+92.g9a754bb and 1.4.4post for velocity estimation and scRNA-seq of villi cells), scVelo (0.1.26.dev7+g5e6d395, <https://github.com/theislab/scvelo> with scanpy v1.3.4), velocity (v0.17.7, <http://velocity.org>), limma (<http://bioinf.wehi.edu.au/limma/>, v3.28.10) via an rpy2 interface (v2.9.1) and gseapy (v0.9.3) implementation of EnrichR (2018/2019).
10. QIAGEN's Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)).
11. CARMAweb

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analysed during this study are included in this published article and its Supplementary Information files. Source data are provided with this paper.

Microarray data have been submitted to NCBI/GEO (GSE148227).

scRNAseq data have been submitted to NCBI/GEO (GSE147319).

Code and custom scripts for scRNAseq analysis have been deposited on [https://github.com/theislab/2021\\_Aliluev\\_Tritschler\\_gut\\_HFD](https://github.com/theislab/2021_Aliluev_Tritschler_gut_HFD).

Databases: <http://metlin.scripps.edu/>; Human Metabolome Database, <http://www.hmdb.ca/>; MassTRIX, <http://masstrix3.helmholtz-muenchen.de/masstrix3/>;

METASPACE, <http://annotate.metaspacesoftware.com/>

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<p>Sample size was calculated based on <math>\alpha</math>-error=0.05, power=0.8, and effect size=1.58. We obtained a sample size of <math>n \geq 8</math>/condition group for physiological measurements during the high-fat high sugar diet feeding based on published data.</p> <p>For single-cell RNA-seq a minimum expected requirement of ~3000 cells per experiment was used and by far exceeded in all cases for confident identification of rare cell subpopulations (~0.05% of cells). No statistical test or power analyzes were used to pre-determine sample size.</p> <p>Otherwise, we chose the sample size of individual experiments based on past experience on detecting differences with a given method and relevant literature (PMID: 26935695, PMID: 22722868, PMID: 29727683).</p>
Data exclusions	<p>For the single-cell RNAseq data cells were filtered using previously described standards for quality control. For each experiment the count matrix was filtered as follows: for scRNA-seq samples of crypt cells genes with expression in less than 20 cells were removed and cells with a fraction of UMI counts from mitochondrially encoded genes of 10% or more were excluded, for scRNA-seq samples of villi cells genes with expression in less than 5 cells were removed and cells with a fraction of UMI counts from mitochondrially encoded genes of 25% or more were excluded. These criteria were based on the cell quality within this study. Otherwise, no data were excluded. Some mice that were on a HFHSD did not gain weight (=non-responder) and were excluded from the analysis.</p>
Replication	<p>Experiments were performed at least in triplicates most with three or more biological replicates unless otherwise indicated. All attempts of replication were successful. Further details about the replication of data are stated in the figure legends.</p>
Randomization	<p>For dietary interventions mice were randomized into test groups based on body weight distribution. For microscopy, the fields of images were randomly selected. For FACS experiments, at least 50,000 events were randomly sampled by the FACS machine. Most other results were derived from computation therefore randomization was not required.</p>
Blinding	<p>The investigators were not blinded to group allocation during in vivo experiments as CD and HFHSD mice were clearly distinguishable by eye. Investigators were not blinded during data collection or analysis as most experiments/analyses were performed by automatic devices such as the FACS or qPCR cyclers. For cell-based assays such as Western blot or qPCR blinding was not possible because the experiments were performed by a single researcher.</p>

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

## Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

primary antibodies:  
 chicken anti-GFP (1:600, Aves Labs, USA, GFP-1020);  
 goat anti-ChgA (1:200, Santa Cruz, sc-1488);  
 rabbit anti-Lyz1 (1:1000, DAKO, M0776);  
 rabbit anti-Muc2 (1:1000, Santa Cruz, sc-7314);  
 rat anti-BrdU (1:200, Abcam, ab6326);  
 rabbit anti-5-HT (1:1000, Neuromics, RA20080);  
 anti-rabbit Ngn3 (1:100, a gift from Helena Edlund lab);  
 goat anti-villin (1:200, Santa Cruz, sc-7672);  
 goat anti-ghrelin (1:200, Santa Cruz, sc-10368);  
 rabbit anti-Ki67 (1:200, Abcam, ab15580);  
 rabbit anti-E-cadherin (extracellular domain) (1:1000, a gift from Dietmar Vestweber)  
 mouse anti-Srebp1 (1:1000, Novus Biologicals, NB600-582SS);  
 rabbit anti-Acc (1:1000, Cell Signaling Technology, 3676);  
 rabbit anti-Ppary (1:1000, Cell Signaling Technology, 2435);  
 goat anti-lamin (1:1000, Santa Cruz, sc-6217);  
 mouse anti- $\alpha$ -tubulin (1:1000, Sigma-Aldrich, T6199);  
 rabbit anti-Fasn 1:1000, Cell Signaling Technology, 3180);  
 rabbit anti-Scd1 (1:1000, Cell Signaling Technology, 2794);  
 mouse anti- $\beta$ -Catenin (1:1000, BD, 610154);  
 rabbit anti-Gsk3 $\beta$  (1:5000, Cell Signaling Technology, 12456);  
 rabbit anti-phospho Gsk3 $\beta$  (1:5000, Cell Signaling Technology, 5558)

Secondary antibodies:  
 donkey anti-chicken Alexa Fluor 488 (1:800, Dianova, 703-225-155);  
 donkey anti-mouse Cy5 (1:800, Dianova, 715-175-151);  
 donkey anti-goat Alexa Fluor 555 (1:800, Invitrogen, A21432);  
 donkey anti-rabbit Alexa Fluor 555 (1:800, Invitrogen, A31572);  
 donkey anti-rabbit Alexa Fluor 649 (1:800, Dianova, 711-605-152);  
 goat anti-mouse HRP (1:15000, Dianova, 115-036-062);  
 goat anti-rabbit HRP (1:15000, Dianova, 111-036-045);  
 rabbit anti-goat HRP (1:15000, Dianova, 305-035-045)

### Validation

primary antibodies: antibodies were validated by the company  
 chicken anti-GFP (1:600, Aves Labs, USA, GFP-1020); <https://www.aveslabs.com/products/green-fluorescent-protein-gfp-antibody>  
 goat anti-ChgA (1:200, Santa Cruz, sc-1488); <https://www.scbt.com/de/p/chr-a-antibody-c-20>  
 rabbit anti-Lyz1 (1:1000, DAKO, M0776); <https://www.labome.com/product/Dako/A0099.html>  
 rabbit anti-Muc2 (1:1000, Santa Cruz, sc-7314); <https://www.scbt.com/p/mucin-2-antibody-ccp58>  
 rat anti-BrdU (1:200, Abcam, ab6326); <https://www.abcam.com/brdu-antibody-bu175-icr1-proliferation-marker-ab6326.html>  
 rabbit anti-5-HT (1:1000, Neuromics, RA20080); <https://www.neuromics.com/RA20080>  
 anti-rabbit Ngn3 (1:100, a gift from Helena Edlund lab); as a control we used a secondary antibody only control, <https://pubmed.ncbi.nlm.nih.gov/31160421/>  
 goat anti-villin (1:200, Santa Cruz, sc-7672); <https://www.scbt.com/de/p/villin-antibody-c-19>  
 goat anti-ghrelin (1:200, Santa Cruz, sc-10368); <https://www.scbt.com/p/ghrelin-antibody-c-18>  
 rabbit anti-Ki67 (1:200, Abcam, ab15580); <https://www.abcam.com/ki67-antibody-ab15580.html>  
 rabbit anti-E-cadherin (extracellular domain) (1:1000, a gift from Dietmar Vestweber)  
 mouse anti-Srebp1 (1:1000, Novus Biologicals, NB600-582SS); [https://www.novusbio.com/products/srebp1-antibody-2a4\\_nb600-582](https://www.novusbio.com/products/srebp1-antibody-2a4_nb600-582)  
 rabbit anti-Acc (1:1000, Cell Signaling Technology, 3676); <https://www.cellsignal.de/products/primary-antibodies/acetyl-coa-carboxylase-c83b10-rabbit-mab/3676>

rabbit anti-Ppary (1:1000, Cell Signaling Technology, 2435); <https://www.cellsignal.de/products/primary-antibodies/pparg-c26h12-rabbit-mab/2435>  
 goat anti-lamin (Santa Cruz, sc-6217); <https://www.scbt.com/de/p/lamin-b-antibody-m-20>  
 mouse anti- $\alpha$ -tubulin (Sigma-Aldrich, T6199); <https://www.sigmaaldrich.com/catalog/product/sigma/t6199?lang=de&region=DE>  
 rabbit anti-Fasn 1:1000, Cell Signaling Technology, 3180); <https://www.cellsignal.de/products/primary-antibodies/fatty-acid-synthase-c20g5-rabbit-mab/3180>  
 rabbit anti-Scd1 (1:1000, Cell Signaling Technology, 2794); <https://www.cellsignal.de/products/primary-antibodies/scd1-c12h5-rabbit-mab/2794>  
 mouse anti- $\beta$ -Catenin (BD, 610154); <https://www.bdbiosciences.com/us/applications/research/stem-cell-research/cancer-research/human/purified-mouse-anti--catenin-14beta-catenin/p/610154>  
 rabbit anti-Gsk3 $\beta$  (1:5000, Cell Signaling Technology, 12456); <https://www.cellsignal.de/products/primary-antibodies/gsk-3b-d5c5z-xp-rabbit-mab/12456>  
 rabbit anti-phospho Gsk3 $\beta$  (1:5000, Cell Signaling Technology, 5558); <https://www.cellsignal.de/products/primary-antibodies/phospho-gsk-3b-ser9-d85e12-xp-rabbit-mab/5558>

#### Secondary antibodies:

donkey anti-chicken Alexa Fluor 488 (Dianova, 703-225-155); <https://www.dianova.com/en/shop/703-175-155-donkey-igg-anti-chicken-igy-hl-cy5-minx-bogogphshohumsrbrtsh/>  
 donkey anti-mouse Cy5 (Dianova, 715-175-151); <https://www.dianova.com/en/shop/715-175-151-donkey-igg-anti-mouse-igg-hl-cy5-minx-bogogphshohurbtrsh/>  
 donkey anti-goat Alexa Fluor 555 (Invitrogen, A21432); <https://www.thermofisher.com/antibody/product/Donkey-anti-Goat-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21432>  
 donkey anti-rabbit Alexa Fluor 555 (Invitrogen, A31572); <https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-31572>  
 donkey anti-rabbit Alexa Fluor 649 (Dianova, 711-605-152); <https://www.dianova.com/shop/711-605-152-esel-igg-anti-kaninchen-igg-hl-alex-fluor-647-minx-bogogphshohumsrtsh/>  
 goat anti-mouse HRP (1:15000, Dianova, 115-036-062); <https://www.dianova.com/shop/115-036-062-ziege-fab2-anti-maus-igg-hl-hrpo-minx-huboho/>  
 goat anti-rabbit HRP (1:15000, Dianova, 111-036-045); <https://www.dianova.com/shop/111-036-045-ziege-fab2-anti-kaninchen-igg-hl-hrpo-minx-hu/>  
 rabbit anti-goat HRP (1:15000, Dianova, 305-035-045); <https://www.dianova.com/shop/305-035-045-kaninchen-igg-anti-ziege-igg-hl-hrpo-minx-hu/>

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

### Laboratory animals

Homozygous Foxa2-Venus fusion (FVF) mice were generated as previously described and backcrossed to C57BL/6 background. Foxa2nEGFP-CreERT2 mice (CD1 background) were crossed with Gt(ROSA)26mTmG mice (mixed 129/SvJ, C57BL/6J background) to obtain heterozygous Foxa2nEGFP-CreERT2/+; Gt(ROSA)26 mTmG/+ animals and bred in our own facilities. Other mouse lines: Lgr5-EGFP-IRES-creERT232 (Lgr5-ki, C57BL/6J background), wild-type C57BL/6N (bred in our own facilities). Mice were housed in groups of two to four animals and maintained at 23  $\pm$  1  $^{\circ}$ C on a 12-hour dark/light cycle with ad libitum access to diet and water unless otherwise indicated. All experiments were performed using male animals 3 to 7 months of age.

All experiments were performed using 3-7-month-old male mice. Mice were housed in groups of two to four animals and maintained at 23  $\pm$  1  $^{\circ}$ C and 45-65 % humidity on a 12-hour dark/light cycle with ad libitum access to diet and water unless otherwise indicated.

### Wild animals

The study did not involve wild animals

### Field-collected samples

No field collected samples were used in this study.

### Ethics oversight

Animal experiments were carried out in compliance with the German Animal Protection Act and with the approved guidelines of the Society of Laboratory Animals (GV-SOLAS) and of the Federation of Laboratory Animal Science Associations (FELASA). This study was approved by the institutional Animal Welfare Officer (Helmholtz Center Munich) and by the Government of Upper Bavaria, Germany.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

### Sample preparation

Small intestines were removed and washed with cold PBS. Villi were scraped off with a glass slide. The remaining tissue was cut into 2-cm pieces, washed several times with cold PBS and incubated in 2 mM EDTA/PBS for 35 min at 4 °C on a tube roller. Subsequently, crypts were harvested by rigorous shaking and filtered through a 70- $\mu$ m mesh to remove villous contaminations/fragments.

Isolated crypts were incubated with TrypLE (Life technologies, #12605) for 5 min on ice and then 5 min at 37 °C and treated with 10  $\mu$ g/ml DNase in crypt complete medium (DMEM/F-12 containing 10 % FCS) for 5 min at 37 °C.

Single cell suspension was achieved by gentle repeated pipetting. Cells were washed twice with FACS buffer (2 % FCS, 2 mM EDTA in PBS) and pelleted at 300xg, 5 min, 4 °C. For flow cytometry, cells were collected in 1-2 ml FACS buffer supplemented with 10  $\mu$ M Rock-inhibitor (Sigma-Aldrich, #Y0503) and passed through the 40  $\mu$ m cell strainer caps of FACS tubes.

### Instrument

Single cells were analyzed by FACS-Aria III (BD) with a 100  $\mu$ m nozzle.

### Software

Data were analyzed with the FACS DIVA software v6.1.3.

### Cell population abundance

Abundance of intestinal cell populations (from live cells = 40-60%) are reported in this manuscript and source data (Fig. 4g, F, Extended Data Fig. 9a). When possible purity of post-sort fractions were checked by briefly re-running a small amount of sample. Purity was usually 90-95%.

### Gating strategy

Single cells were gated according to their FSC-A (front scatter area) and SSC-A (side scatter area). Singlets were gated dependent on the FSC-W (front scatter width) and FSC-H (front scatter height) and SSC-W and SSC-H and dead cells were excluded using the marker7AAD (eBioscience). To obtain FVF-enriched small intestinal crypt cell samples for single-cell RNA sequencing, 30,000 FVF+ (FVF<sub>low</sub> and FVF<sub>hi</sub>) cells were sorted followed by sorting 30,000 live crypt cells per sample.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.



## Appendix D

This is the published version of the article in *Molecular Metabolism* following peer review.

The article is open access thus the published version is inserted here. This article is published in *Molecular Metabolism* under a Creative Commons Attribution license (CC- BY-NC-ND 4.0).

**Sophie Tritschler**, Fabian J. Theis, Heiko Lickert, Anika Böttcher, **Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas**, *Molecular Metabolism*, Volume 6, Issue 9, Pages 974-990 (2017)

# Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas



Sophie Tritschler<sup>1,2,3</sup>, Fabian J. Theis<sup>3</sup>, Heiko Lickert<sup>1,2,4</sup>, Anika Böttcher<sup>1,2,4,\*</sup>

## ABSTRACT

**Background:** Diabetes mellitus is characterized by loss or dysfunction of insulin-producing  $\beta$ -cells in the pancreas, resulting in failure of blood glucose regulation and devastating secondary complications. Thus,  $\beta$ -cells are currently the prime target for cell-replacement and regenerative therapy. Triggering endogenous repair is a promising strategy to restore  $\beta$ -cell mass and normoglycemia in diabetic patients. Potential strategies include targeting specific  $\beta$ -cell subpopulations to increase proliferation or maturation. Alternatively, transdifferentiation of pancreatic islet cells (e.g.  $\alpha$ - or  $\delta$ -cells), extra-islet cells (acinar and ductal cells), hepatocytes, or intestinal cells into insulin-producing cells might improve glycemic control. To this end, it is crucial to systematically characterize and unravel the transcriptional program of all pancreatic cell types at the molecular level in homeostasis and disease. Furthermore, it is necessary to better determine the underlying mechanisms of  $\beta$ -cell maturation, maintenance, and dysfunction in diabetes, to identify and molecularly profile endocrine subpopulations with regenerative potential, and to translate the findings from mice to man. Recent approaches in single-cell biology started to illuminate heterogeneity and plasticity in the pancreas that might be targeted for  $\beta$ -cell regeneration in diabetic patients.

**Scope of review:** This review discusses recent literature on single-cell analysis including single-cell RNA sequencing, single-cell mass cytometry, and flow cytometry of pancreatic cell types in the context of mechanisms of endogenous  $\beta$ -cell regeneration. We discuss new findings on the regulation of postnatal  $\beta$ -cell proliferation and maturation. We highlight how single-cell analysis recapitulates described principles of functional  $\beta$ -cell heterogeneity in animal models and adds new knowledge on the extent of  $\beta$ -cell heterogeneity in humans as well as its role in homeostasis and disease. Furthermore, we summarize the findings on cell subpopulations with regenerative potential that might enable the formation of new  $\beta$ -cells in diseased state. Finally, we review new data on the transcriptional program and function of rare pancreatic cell types and their implication in diabetes.

**Major conclusion:** Novel, single-cell technologies offer high molecular resolution of cellular heterogeneity within the pancreas and provide information on processes and factors that govern  $\beta$ -cell homeostasis, proliferation, and maturation. Eventually, these technologies might lead to the characterization of cells with regenerative potential and unravel disease-associated changes in gene expression to identify cellular and molecular targets for therapy.

© 2017 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords**  $\beta$ -Cell heterogeneity; Single-cell analysis; Diabetes; Regeneration; Endocrine cells; Transdifferentiation; Dedifferentiation; Maturation; Subpopulations

## 1. INTRODUCTION

Diabetes mellitus is a complex and multifactorial disease characterized by progressive loss or dysfunction of the insulin-producing  $\beta$ -cells in the pancreas. This results in chronic hyperglycemia and systemic metabolic complications and, in the long-term, causes multi-organ damage including nephropathy, retinopathy, and enteropathy. Today, over 382 million people worldwide have been diagnosed with diabetes and the number is expected to rise to 592 million by 2035 [1]. Type 1 diabetes (T1D) is an autoimmune disorder caused by destruction of  $\beta$ -

cells through cytotoxic T-cells. Unlike in T1D, onset of the more prevalent type 2 diabetes (T2D) is usually in adulthood and is often consequence of genetic predisposition, obesity and lack of physical exercise. T2D is triggered by insulin resistance of the peripheral tissues, which is concomitant with  $\beta$ -cell mass expansion,  $\beta$ -cell exhaustion, and gradual loss of functional  $\beta$ -cell mass through  $\beta$ -cell dedifferentiation and/or  $\beta$ -cell death [2,3]. Thus, the common feature of both pathologies is loss of functional  $\beta$ -cells. Despite its high prevalence and increasing impact on global health, diabetes is still incurable and our knowledge of the underlying pathomechanisms is far

<sup>1</sup>Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, Am Parkring 11, 85748 Garching-Hochbrück, Germany <sup>2</sup>German Center for Diabetes Research, 85764 Neuherberg, Germany <sup>3</sup>Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany <sup>4</sup>Institute of Stem Cell Research, Helmholtz Zentrum München, 85764 Neuherberg, Germany

\*Corresponding author. Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, Am Parkring 11, 85748 Garching-Hochbrück (near Munich), Germany. E-mail: [anika.boettcher@helmholtz-muenchen.de](mailto:anika.boettcher@helmholtz-muenchen.de) (A. Böttcher).

Received April 7, 2017 • Revision received June 13, 2017 • Accepted June 19, 2017 • Available online 20 July 2017

<http://dx.doi.org/10.1016/j.molmet.2017.06.021>

from complete. Current treatments succeed in reducing symptoms; however, they fail to alleviate long-term complications and require lifelong compliance from patients. Therefore, intensive efforts in the field of diabetes research are put into the development of novel therapeutic strategies to stop the progression of the disease and restore functional  $\beta$ -cell mass.

Human islet transplantation from cadaveric donors has been successfully established as a therapeutic treatment for a subset of patients with “brittle” T1D that do not respond to standard conventional and intensive insulin therapies and suffer from kidney failure [4,5]. However, donor shortage and risks associated with life-long immunosuppression demand the development of alternative therapies. Two main strategies are currently extensively explored to replace lost and/or dysfunctional  $\beta$ -cells: i) *in vitro* differentiation of  $\beta$ -cells from stem cells and ii) endogenous  $\beta$ -cell regeneration. The former holds great promise for cell-replacement therapy and tissue engineering. In the past years, major advances have enabled the generation of mono-hormonal and glucose-responsive  $\beta$ -like cells from human embryonic stem cells and patient-derived induced pluripotent stem cells [6–8]. Importantly, these cells were able to secrete insulin and restored normoglycemia in diabetic mice [9]. Still, prior to application in humans, the differentiation efficiency and functionality of *in vitro* generated  $\beta$ -like cells needs to be improved. In this regard, the field would benefit greatly from a better understanding of the postnatal  $\beta$ -cell maturation process and the identification of biomarkers that label the different maturation stages and functional glucose-responsive  $\beta$ -cells. In addition, their immune-protection as well as safety must be guaranteed as not fully differentiated stem cells might have teratoma-initiating potential.

Stimulating regeneration of insulin-producing cells from cells residing within the adult pancreas or even in other metabolically active organs,

such as the liver or gut (not discussed in this review), is an appealing approach that could bypass the aforementioned hurdles. The main routes pursued to restore functional  $\beta$ -cell mass *in situ* include boosting the replication of remaining  $\beta$ -cells, maturation of immature (dedifferentiated)  $\beta$ -cell subpopulations, mobilization of putative precursors present in the adult pancreas and reprogramming of other cell types into insulin-producing  $\beta$ -like cells (Figure 1) [10]. Important in this respect is the existence of  $\beta$ -cell subpopulations that differ in their glucose responsiveness, proliferative activity, maturation state, or susceptibility to metabolic deregulation in animal models [11]. Moreover, adult exocrine and other endocrine cell types showed the ability to reprogram and produce insulin under certain conditions [12]. Further characterization of these candidate sources for the generation of new insulin-producing cells as well as the identification of biomarkers and therapeutic targets requires detailed dissection of the cellular heterogeneity within the pancreas and their underlying molecular mechanisms. To this end, single-cell studies might be paradigm changing. Single-cell technologies allow for simultaneously measuring the expression of tens to thousands of genes (e.g. single-cell RNA sequencing) or proteins (e.g. single-cell mass cytometry, flow cytometry) in individual cells with high-throughput and precision. Clustering of cells as per their expression profiles allows for unbiased detection and characterization of cell types and states including rare or unanticipated subpopulations that are masked in bulk analyses (Figure 2). By pooling many cells with partially correlated measurements, one can derive rich molecular profiles without prior knowledge of defining criteria and screen for subtype specific marker genes even if only a limited number of transcripts or proteins per cell are captured [13,14]. In addition, single-cell measurements provide an accurate temporal resolution of continuous processes, such as differentiation or reprogramming, as cells of all present (transient and stable) stages are

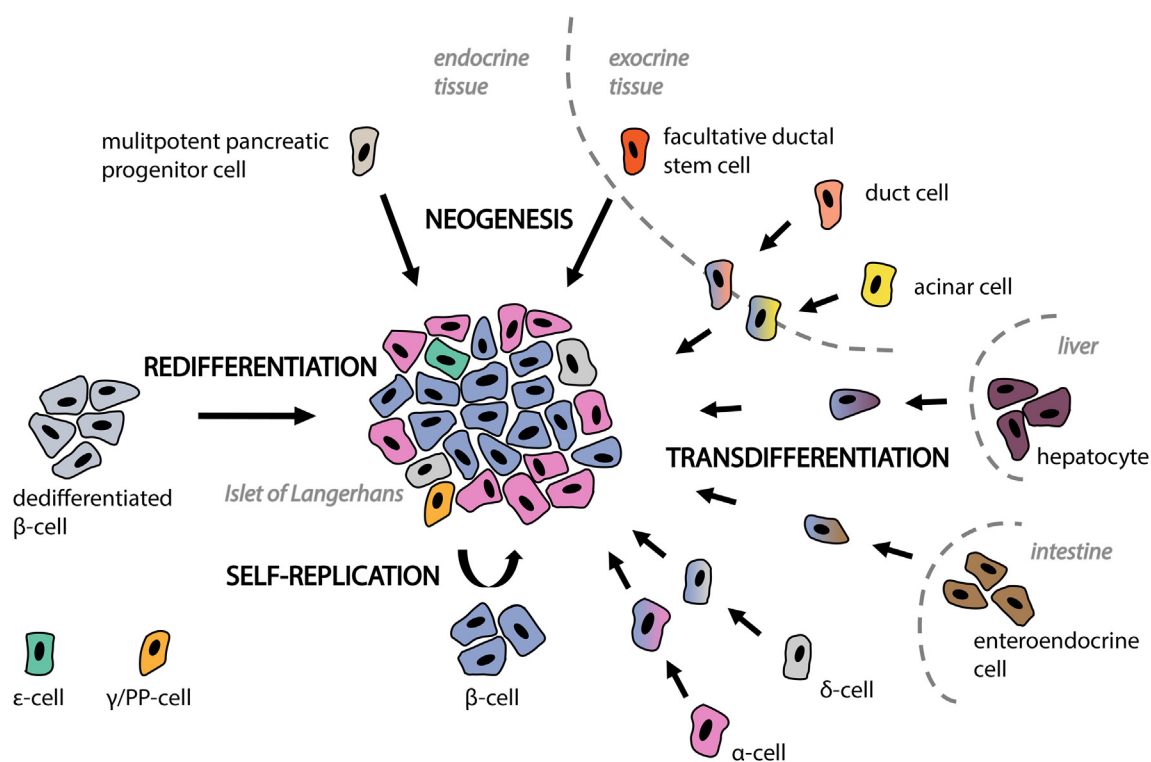
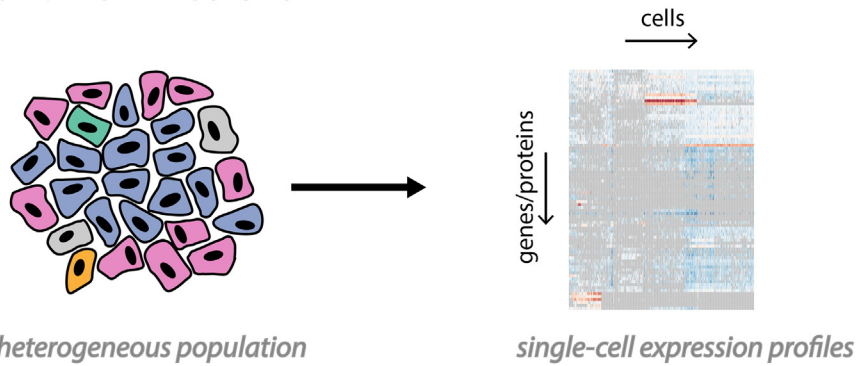
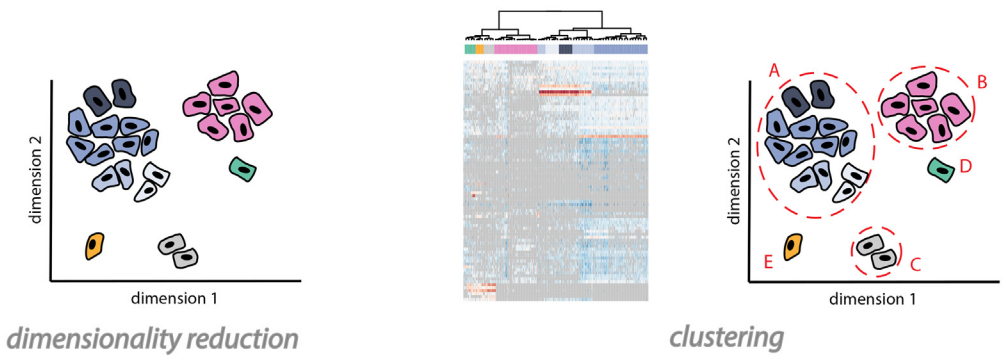


Figure 1: Main routes to restore functional  $\beta$ -cell mass *in situ*. Schematic summarizing the possible ways of  $\beta$ -cell regeneration that are discussed in the text.

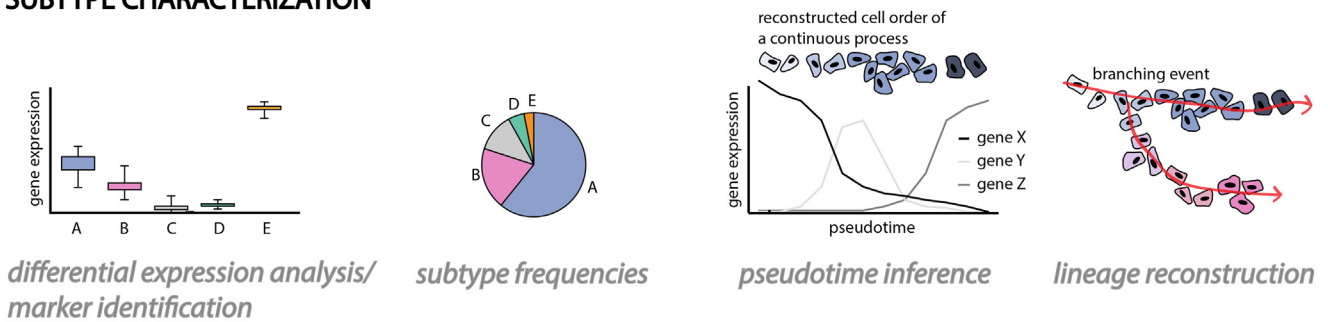
### SINGLE-CELL RESOLUTION



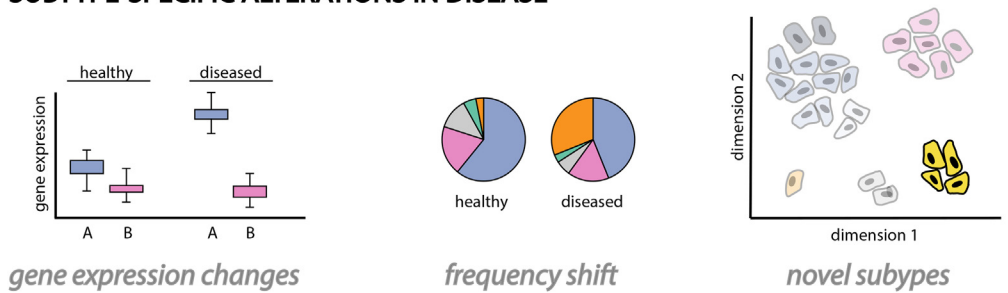
### VISUALIZATION AND SUBTYPE IDENTIFICATION



### SUBTYPE CHARACTERIZATION



### SUBTYPE-SPECIFIC ALTERATIONS IN DISEASE



**Figure 2: Identification and characterization of heterogeneity by single-cell analysis.** Single-cell approaches enable the identification and characterization of cellular subtypes in a heterogeneous population and their implications in disease based on rich expression profiles acquired at single-cell resolution. Computational analyses translate the high-dimensional data into low-dimensional cell maps and extract information on subtype composition and gene expression differences within and between subtypes in healthy and diseased conditions as well as reconstruct cellular trajectories of continuous processes.

captured simultaneously. The temporal order and lineage choices can be reconstructed from single-cell snapshot data using computational algorithms that infer a pseudotime and detect branching events [15–17]. This provides information on the genes most involved in determining the identity of a cell and on the factors that are expressed transiently. Finally, single-cell analyses have important implications in medicine since they allow parallel identification of disease-associated alterations in cellular composition as well as in gene expression of cell subtypes. Differences shed light on how cell-to-cell variability could lead to different cell function or fate of seemingly identical cells of a (sub)population, and on how these subpopulations respond to external cues or drugs, and therefore on etiology, pathomechanisms, and treatment efficiency (Figure 2).

Here we provide an overview of single-cell analyses of pancreatic cells and discuss how these new data add to our knowledge of pancreatic cell heterogeneity and improve our understanding of possible endogenous  $\beta$ -cell regeneration routes.

## 2. DETERMINATION OF THE CELLULAR COMPOSITION OF THE PANCREAS

The mature pancreas can be subdivided into the functionally distinct endocrine and exocrine compartment. The exocrine tissue comprises more than 95% of the pancreas mass and is composed of acinar and ductal cells. The endocrine tissue is organized into cell clusters, the so-called islets of Langerhans, that are dispersed within the exocrine tissue. The islets of Langerhans are comprised of five distinct endocrine cell types:  $\alpha$ -cells (secreting glucagon),  $\beta$ -cells (insulin),  $\gamma$ /PP-cells (pancreatic polypeptide),  $\delta$ -cells (somatostatin), and  $\epsilon$ -cells (ghrelin), together making up less than 5% of the pancreas mass.

Islet composition is determined during embryonic development, and has been extensively discussed in excellent reviews [18,19]. Briefly, in mice, the pancreas is first identified by the expression of the transcription factor pancreatic and duodenal homeobox 1 (Pdx1) in the foregut endoderm at embryonic day (E) 8.5. Pdx1 expression is followed by the expression of several other transcription factors that are required for pancreatic specification, including Pancreas specific transcription factor 1 (Ptf1a), Forkhead box protein A 1/2 (Foxa1/2), SRY box 9 (Sox9), Hepatocyte nuclear factor-1 $\beta$  (Hnf1 $\beta$ ), Gata4/6, Hairy and enhancer of split 1 (Hes1), and Nkx2.2 [18,19]. During the initial phase of murine pancreatic development, which has been referred to as primary transition, the Pdx1 multipotent pancreatic progenitor cells (MPCs) have high proliferative capacity and primarily give rise to a few glucagon-producing  $\alpha$ -cells [20,21]. The secondary transition, which encompasses the time between E12.5 and birth, is characterized by expansion and branching morphogenesis of the pancreatic epithelium and is closely linked to endocrine cell differentiation. During these complex morphogenetic events, MPCs become lineage restricted and adopt either tip or trunk identity. The trunk progenitors are bi-potential and further differentiate into ductal or endocrine precursor cells. The ductal versus endocrine fate decision is thought to be controlled by graded Notch activity, with high Notch signaling promoting ductal fate and localized inhibition of Notch signaling, allowing transient Ngn3 expression and endocrine differentiation [22]. Shortly after induction of Ngn3 expression, endocrine committed cells start to delaminate from the ductal epithelium and form nascent islets. Ngn3 activates the expression of several transcription factors implicated in endocrine lineage maintenance and specification including the pan-endocrine marker neuronal differentiation 1 (Neurod1), the LIM homeobox protein islet 1 (Isl1), paired box 4 and 6 (Pax4 and Pax6), and aristaless-related homeobox (Arx).

Specification into  $\alpha$ - and  $\beta$ -cell precursor relies on mutual repression between opposing lineage determinants. The transcription factors Pax4, Nkx2.2, and Nkx6.1 are direct transcriptional repressors of Arx, which specifies  $\alpha$ -cell fate and conversely, Arx represses Pax4, Nkx6.1, and Pdx1, which specify  $\beta$ -cell fate [23]. The subsequent maturation of insulin-positive cells into functional, glucose-responsive  $\beta$ -cells is characterized by a switch from MafB to MafA expression in mice [24–26]. Importantly, the transcription factors Pax6 and Nkx2.2 are not only crucial for  $\beta$ -cell lineage formation and function but they also actively repress non- $\beta$ -cell transcriptional programs, a requirement for maintenance of  $\beta$ -cell identity [27,28]. While the transcription factors determining  $\alpha$ - and  $\beta$ -cell fate are known, the formation of the less abundant  $\gamma$ /PP,  $\delta$ - and  $\epsilon$ -cells is not well understood and their fate might be determined by combinatorial action of different transcription factors.

Physiologically,  $\alpha$ - and  $\beta$ -cells act in concert to regulate blood glucose levels.  $\beta$ -cells release insulin in response to high blood glucose levels to stimulate glucose uptake into peripheral tissues such as liver, muscle, and fat tissue. In contrast,  $\alpha$ -cells release glucagon in response to low levels of glucose to stimulate glycogenolysis in the liver and in order to prevent hypoglycemia during fasting or exercise [29]. So far, the function of  $\delta$ -,  $\gamma$ /PP-, and  $\epsilon$ -cells remains largely elusive due to their low abundance. Recently, it has been shown that the peptide hormone urocortin3 which is co-released with insulin potentiates somatostatin secretion from  $\delta$ -cells and possibly in this way  $\delta$ -cells fine-tune insulin secretion [30]. Pancreatic polypeptide, which is released by  $\gamma$ /PP-cells, is believed to regulate pancreatic exocrine and endocrine secretions [31–33].

Interestingly, differences in islet composition and architecture between rodents and humans have been reported. Murine islets of Langerhans are comprised of up to 60–80% of insulin-producing  $\beta$ -cells that are clustered in the center of the islet and are surrounded by a mantle of endocrine  $\alpha$ -(15–20%),  $\delta$ -(< 10% of islet cells), and  $\gamma$ /PP-cells (< 1% of islet cells) [34]. Cell type composition and architecture of human islets seem to be more variable and are still matter of debate [35–37]. Yet, after birth, the majority is similarly organized as murine islets and show the characteristic mantle-core feature [37]. The human islet cell-type composition is on average ~50%  $\beta$ -cells, ~40%  $\alpha$ -cells, and ~10%  $\delta$ -cells, up to a few percent  $\gamma$ /PP cells, and very few  $\epsilon$ -cells [38–40]. The differences between rodents and humans in islet cell-type abundance might indicate additional functions of  $\alpha$ -cells and/or increased functional dependencies of  $\alpha$ - and  $\beta$ -cells in humans.

## 3. SINGLE-CELL ANALYSIS UNRAVELS THE TRANSCRIPTIONAL PROGRAM OF ADULT PANCREATIC CELL TYPES AND IDENTIFIES SPECIES-SPECIFIC DIFFERENCES

Recent technological advances in the field of single-cell ‘omics’ now allow for easily measuring gene or protein expression in thousands of cells. Importantly, pure cell populations can be derived during subsequent data analyses by clustering of cells based on their high-dimensional expression profiles, which is not possible through isolation from a population using a few markers. The systematic transcriptome analysis of adult human pancreatic cell types revealed that the expression of most lineage-determining transcription factors is conserved between human and mice (Figure 3). Strikingly, classifying hormone expression accounted for 50% of the total cellular transcripts in  $\beta$ -,  $\gamma$ /PP-, and  $\delta$ -cells. Whereas in  $\alpha$ - and  $\epsilon$ -cells the expression of *GCG* and *GHRL*, respectively, made up 20% of their transcriptomes [41]. Single-cell analysis also revealed that 15% of all  $\alpha$ - and  $\beta$ -cell enriched genes show distinct species-dependent expression [42].

pan-endocrine: PAX6, MORF4L1, NEUROD1, INSM1, NKX2.2

β-cell	α-cell	ε-cell	δ-cell	γ/PP-cell
PAX4	ARX	GHRLOS	HHEX	ARX
PDX1	MAFB	NPY1R	ETV1	ETV1
MAFA	IRX2	ASGR1	MAFB	UCN3
MAFB	GC	ARX	LEPR	FEV
BMI1	FEV		GHSR	
DLK1	TM4SF4		GPR120	
SIX2/3	GPR119		POUF3F1	
ID1	UCN3			
IAPP	FEV			
UCN3	DPP4			
OLIG1				

**Figure 3: Transcriptional program of human pancreatic endocrine cells revealed by single-cell transcriptomics.** Depicted genes are highly or exclusively expressed in the given endocrine cell type. Genes highlighted in blue show species-specific differences in cell-type expression.

Genes with robust expression in human, but not in mouse β-cells are for example *RBP4*, *DLK1*, *ADCYAP1*, *RGS16*, *SOX4*, *BMP5*, *TIMP2*, *TSPAN1*, *MAFB*, and *TFF3*. Genes with high expression in mouse but not in human β-cells are *UCN3*, *S100A1*, *ADH1C*, *FAM151A*, *COX6A2*, *BACE2*, *TTYH2*, *SYTL4*, *SDF2L1*, *FRZB*, and *PRLR* [42].

Besides differences in expression levels, some cell-type enriched genes also showed notable species differences. As mentioned above the transcription factors MafB and MafA regulate β-cell lineage formation and function in rodents [26]. During mouse endocrine lineage specification MafB expression becomes restricted to α-cells, whereas MafA is expressed in β-cells [24–26]. In contrast, in humans, *MAFB* expression has been detected in adult α-cells but also in β-cells and its expression pattern has now been expanded to δ-cells [41–44]. Interestingly, Li et al. found that half of the β-cells they studied expressed *MAFA* and *MAFB* concomitantly, suggesting that *MAFB* is a new marker describing β-cell heterogeneity in humans [44]. The Notch pathway component *delta like non-canonical Notch ligand 1* (*DLK1*), which has been associated with T1D in genome-wide association studies (GWAS), was found to be specifically expressed in human β-cells, whereas it is highly expressed in postnatal, immature α- and β-cells, and adult α-cells but not adult β-cells in mice [42,44–46]. Similarly, several groups confirmed previous findings by the Kaestner group that the gene *group-specific component (vitamin D binding protein)* (*GC*) is specifically expressed in α-cells in human tissue, whereas it is co-expressed with insulin in mouse islets (Figure 3) [44,47,48]. Vitamin D deficiency and common *GC* non-coding variants have been associated with T1D, T2D, and an increased risk of gestational diabetes mellitus [49,50]. These findings implicate a contribution of dysfunctional α-cells to the pathogenesis of diabetes. In-depth analysis of the genetic framework of β-cells might identify new transcription factors that drive functional maturation of β-cells. Indeed, several groups reported the expression of *SIX2* and *SIX3* in human β-cells [41,45,48,51]. These two transcription factors have recently been shown to elevate insulin content and insulin secretion in juvenile, immature β-cells suggesting a crucial role of *SIX2* and *SIX3* in β-cell maturation [52]. Importantly, *Six2* and *Six3* seem not to be expressed in adult mouse β-cells. Altogether, these findings call for caution when extrapolating findings from mice to humans and highlight the importance of studying human pancreatic cell types as well as indicate the requirement of cell-type specific mapping of genes.

So far, our understanding about the transcriptional program and function of pancreatic δ- and γ/PP-cells is limited due to their low

abundance. Only recently, DiGrucio et al. reported that murine δ-cells express several receptors, which they partially have in common with β-cells, such as *glucagon-like peptide 1 receptor* (*Glp1r*), the *alpha 2 adrenergic receptor* (*Adra2a*), and the *glucagon receptor* (*Gcgr*), whereas the expression of *somatostatin receptor* (*Sstr1*) and *ghrelin receptor* (*Ghsr*) seems to be restricted to δ-cells [53]. Ghrelin has a central role in energy and glucose metabolism and a well-known insulinostatic action [54], which can now be explained by its direct effect on somatostatin-releasing δ-cells. Several single-cell RNA sequencing studies confirmed the expression of *GHSR* in human δ-cells and additionally identified the leptin receptor (*LEPR*) as a δ-cell specific receptor [41,45,48,51]. Leptin is an adipose tissue-derived hormone that plays an important role in the maintenance of body weight and glucose homeostasis [55]. Until now, the peripheral target cell of leptin action was controversial. Several studies in mice have suggested a direct effect of leptin on β-cells. It has been shown that leptin suppressed insulin release from insulinoma cells and pancreas-specific leptin receptor knock-out mice exhibited improved glucose tolerance, increased early-phase insulin secretion and increased β-cell size [56,57]. However, leptin receptor deletion in insulin- or proglucagon-expressing cells had only a minor impact on glucose homeostasis [58]. Thus, the δ-cell-specific expression of the leptin receptor would explain the conflicting results obtained by a pancreas-specific (including δ-cells) versus β/α-cell-specific leptin-receptor deletion. Interestingly, several diabetes risk genes such as *PDX1*, *HADH*, and *UCP2* are expressed by β- and δ-cells, thus implicating the δ-cell type in islet dysfunction in rare and common forms of diabetes [51]. Taken together, these new data strongly suggest an important role of δ-cells in maintaining β-cell function and systemic regulation of appetite, food intake, and energy balance.

Little is known about the function of the rare γ/PP-cells. Human γ/PP-cells have now been reported to express high levels of the transcription factors *MEIS2*, *ETV1*, *ID4*, and the serotonergic transcription factor *FEVI/PET1* besides *TPH1*, *SERTM1*, *SPOCK1*, *ABCC9*, and *SLITRK6*, suggesting similarity of γ/PP-cells with neuronal cells [41,44,45,51]. To summarize, recent single-cell studies reveal the transcriptional program of rare islet cell-types and species-specific differences in gene expression important for the interpretation of conditional knock-out studies in mice and for consideration when designing new therapies for diabetes (Figure 3).

## 4. APPROACHES OF ENDOGENOUS β-CELL REGENERATION

### 4.1. Targeting β-cell subpopulations

Endogenous regenerative therapy aims to stimulate subset(s) of cells with the potential to compensate for the lost functional β-cell mass. A tempting source in this regard is the remaining (dedifferentiated) β-cells that survive disease for long periods. It has been reported that in T1D patients a small amount of functional β-cells remain for many years and escape immune attack [59–61]. Similarly, not all β-cells undergo cell death in T2D patients. A fraction of cells regress to a more immature state, creating a pool of possible precursor cells that can potentially be re-differentiated [3,62,63]. In addition, in an early stage of disease, there are still functional β-cells, which may be expanded by inducing their proliferation. The molecular differences that cause cells to adopt distinct fates and make them survive, die, or proliferate are still elusive. Thus, to directly target specific β-cell subpopulations and trigger their proliferation and/or functional maturation, it is crucial to decipher β-cell heterogeneity and to identify the underlying molecular mechanism driving it. It is long known that β-cells are a heterogeneous population of cells that differ in morphology, glucose responsiveness,

insulin secretion, proliferative capacity and maturation state [11,64–67]. Regardless of these extensive studies of phenotypic and functional  $\beta$ -cell heterogeneity first described over 50 years ago, heterogeneity has only recently moved into the spotlight of regenerative diabetes research [11,68,69]. Besides pathological dedifferentiation, heterogeneity may be caused by differences in the microenvironment, islet architecture, and distinct origins of  $\beta$ -cells. Markers that are differentially expressed in  $\beta$ -cells in homeostasis include insulin [67,70], the transcription factor Pdx1 [71], the membrane transporter glucose transporter Glut2 [72], and signaling components such as WNT antagonist DKK3 [73] among many others [11].  $\beta$ -cell heterogeneity has been mostly studied in rodents. The detection of  $\beta$ -cell heterogeneity within human pancreatic cell populations remains currently a challenging task due to large donor-to-donor variability [41,74], specific loss of  $\beta$ -cells during the experimental procedure [41], and a requirement to analyze a large number of cells. Nonetheless, recent single-cell studies confirm that murine and human  $\beta$ -cells differ in their proliferative capacity, insulin secretion, and response to diabetogenic cues as discussed below (Figure 4).

#### 4.1.1. Triggering $\beta$ -cell proliferation

Pancreatic islets possess the functional flexibility to adapt rapidly to environmental changes by  $\beta$ -cell mass expansion and enhanced insulin secretion. Exploiting natural proliferation cues to increase cell number in disease, therefore, is evident.  $\beta$ -cell proliferation peaks during the early postnatal period, in which  $\beta$ -cell mass is determined. Qiu et al. showed that 25% of the  $\beta$ -cells are proliferative on postnatal day (P) 3 in mice based on the expression of cell cycle-related genes. However, from P9 onwards proliferation decreases rapidly and the adult  $\beta$ -cell turnover is minimal (below 1% in mice) but stable [46,74,75]. Understanding the molecular mechanisms that drive the early postnatal  $\beta$ -cell proliferation boost and induce cell cycle arrest of most  $\beta$ -cells thereafter might be helpful for therapeutic approaches. Two recent single-cell RNA-sequencing studies aimed to reconstruct the developmental trajectory of pancreatic  $\beta$ -cells to gain insight into the regulation of postnatal proliferation and maturation. Zeng et al. used the *mInsl1-H2B-mCherry* reporter mouse line to isolate  $\beta$ -cells whereas Qiu et al. made use of the *Insl1-RFP* and *Gcg-Cre; Rosa-RFP* reporter mouse lines to isolate  $\beta$ -cells and  $\alpha$ -cells, respectively [46,76]. In both studies, the maturation trajectory was reconstructed by ordering  $\beta$ -cells isolated from multiple postnatal time points based on their transcriptional similarity. Qiu et al. reported 664 genes and 448 genes to be dynamically regulated during the  $\beta$ -cell maturation and  $\alpha$ -cell maturation processes, respectively. Interestingly, their results suggest that  $\beta$ -cells mature primarily through the upregulation of genes. In contrast,  $\alpha$ -cells seem to mature through downregulation of genes expressed in immature  $\alpha$ -cells [46]. Furthermore, pseudotemporal cell ordering revealed the signatures of immature, proliferative  $\beta$ -cells and associated expression changes of genes regulating amino acid uptake and metabolism, mitochondrial respiration, and reactive oxygen species (ROS) production with postnatal  $\beta$ -cell development. Precisely, the results of the Sander group suggest that amino acid deprivation due to downregulation of amino acid transporter genes, as well as reduced ROS level and downregulation of *serum response factor* (*Srf*) and its target genes during  $\beta$ -cell maturation might contribute to the postnatal decline in  $\beta$ -cell proliferation [76]. Other pathways that are regulated in pseudotime and implicated in proliferation are MAPK and PDGF signaling [76]. In addition, the transcription factor *Foxm1* and several members of the pro-proliferative *E2F* transcription factor family and their targets are highly expressed in postnatal proliferative  $\beta$ -cells [46,76]. In that





respect, it is also worth mentioning that cell cycle exit/proliferation seems to be differently regulated in immature and mature  $\alpha$ -/ $\beta$ -cells as the cell cycle inhibitors p57 and p18 are highly expressed in immature and mature islet cells, respectively [46]. Thus, targeting the ROS/SRF/MAPK/PDGF pathways and amino acid availability might reactivate and stimulate proliferation in adult human  $\beta$ -cells.

As a boost of  $\beta$ -cell proliferation is observed under high metabolic demand, such as pregnancy or obesity, it would be interesting to investigate if pathways implicated in the regulation of early postnatal  $\beta$ -cell proliferation are reactivated under these conditions [77]. However, increased  $\beta$ -cell proliferation upon metabolic demand shows that at least a subset of adult  $\beta$ -cells maintains the ability to replicate.

Unlike tissues with rapid cell turnover, such as skin, blood, or gut, the presence of stem cells in the pancreas is controversial [10]. Even though neogenesis of ductal facultative stem cells residing outside the islets may contribute to formation of new  $\beta$ -cells [78,79], lineage-tracing studies indicate that self-replication of pre-existing  $\beta$ -cells is the major source of new  $\beta$ -cells *in vivo* in homeostasis and upon injury [80]. Others reported the existence of a small (< 1%) adult pancreatic multipotent progenitor (PMP) population within both mouse and human islets. The described PMP cells are insulin-positive and show high proliferative capacity, can give rise to all endocrine cell types *in vitro*, and may contribute to  $\beta$ -cell compensation under stress and hyperglycemia [72,81,82]. The presence of PMPs would have major implications for regenerative therapy; however, due to the very low cell turnover in the islet the concept needs verification. Identification of marker genes for the isolation of PMPs as well as stem cell assays to test potency and the ability of self-renewal are warranted [10].

To efficiently stimulate  $\beta$ -cell replication, it is important to delineate heterogeneities in the replicative potential of cells. The presence of adult, proliferative human  $\beta$ -cells in homeostasis is supported by a single-cell mass cytometry study, in which islets of 20 human donors covering ages from birth to adulthood were profiled [74]. In accordance with previous studies, cell percentage and proliferation of all major endocrine cells was highest neonatally and declined after childhood with some basal proliferation remaining in adults. Moreover, hierarchical clustering revealed three, distinct  $\beta$ -cell states. Proliferative cells segregated into two of these subgroups. The cluster containing most of the Ki67-positive  $\beta$ -cells also showed higher levels of the proliferation marker Ki67 and an upregulation of signaling components involved in  $\beta$ -cell proliferation (PDGFRA, pERK1/2, pSTAT3, and pSTAT5) [83]. The surface markers CD44 and CD49F were identified among the proteins highest expressed in proliferative  $\beta$ -cells, which have been associated with pancreatic progenitor cells and pancreatic cancer-initiating cells [84,85]. Strikingly, the number of proliferative cells is decreased in T2D donors [74]. Nevertheless, treatment with the drug harmine, which has been previously demonstrated to enhance human  $\beta$ -cell proliferation [83,86,87], had similar effects on various endocrine cells from T2D and healthy donors indicating that even in an impaired metabolic state endocrine mass can be expanded.

So far, single-cell RNA sequencing captured proliferative  $\beta$ -cells only from early postnatal stages [46,76]. This is possibly due to the very low replication rate inherent to adult  $\beta$ -cells [74] and high sensitivity of proliferative  $\beta$ -cells to flow cytometry, on which many single-cell technologies rely [41]. In addition, most studies to date analyzed a limited number of cells or lacked an in-depth analysis of  $\beta$ -cells, and, therefore, rare cell types might have been missed. Still, several single-cell RNA sequencing studies found indications for the presence of proliferative  $\beta$ -cell subpopulations in adult islets. The Sandberg group identified five clusters within the  $\beta$ -cell population, of which three showed elevated expression of *inhibition of differentiation* (*ID*) genes

PHENOTYPE		mature	immature	proliferative	other	genes/ proteins
 β-cells	<b>GPR120</b> <sup>41</sup>	■	■	-		<i>GPR120</i>
	<b>RBP4</b> <sup>41,45,48,61</sup>	■	■	-	downregulated in resistant cells in T1D	<i>RBP4</i> , β-cell markers (e.g. <i>UCN3</i> , <i>MAFA</i> , <i>CHGA</i> )
	<b>ER stress</b> <sup>45,48,76</sup>	■	■	91	absent in T2D, upregulated in early postnatal β-cells	<i>HERPUD1</i> , <i>DDIT3/Ddit3</i> , <i>TRIB3</i> , <i>HSPA5</i> , <i>PP1R15A</i> , <i>LURAP1L</i> , <i>ATF3/Atf3</i> , <i>FTH1P3</i> , <i>SQSTM1</i> , <i>FTH1</i> , <i>FTL</i> , <i>SRXN1</i> , <i>Atf4</i> , <i>C/EBP</i>
	<b>ID genes</b> <sup>41,76</sup>	■	88,89,90	88,89,90	absent in T2D, upregulated in early postnatal β-cells	<i>ID1</i> , <i>ID3</i>
	<b>CD44/CD49F</b> <sup>74</sup>	■	-	■	absent in T2D	<i>Ki67</i> , <i>PDGFRA</i> , <i>pERK1/2</i> , <i>pSTAT3</i> , <i>pSTAT5</i>
 α-cells	<b>Ucn3</b> <sup>110</sup>	■	■	-	intermediate stage in α- to β-cell/β- to α-cell conversion, neogenic niche at islet edge, no glucose sensing	<i>Ucn3</i> , <i>G6pc2</i> , <i>Eroib</i> , <i>Glut2</i>
	<b>SHH</b> <sup>41,63</sup>	-	-	■		<i>Ki67</i> , <i>GLI</i> factors, <i>DYRK1A</i> , <i>GSK3β</i> , <i>STMN1</i>
 acinar cells	<b>STMN1</b> <sup>41,116</sup>	-	-	■		<i>STMN1</i> , <i>SOX9</i>
	<b>REG3A</b> <sup>48</sup>	■	48,129	129,130	spatial patches	<i>REG3A</i>
	<b>MHC class II</b> <sup>41</sup>	■	-	-	inflammatory	<i>CD74</i> , <i>HLA-DMA</i> , <i>HLA-DRA</i> , and <i>HLA-DRB1</i> , <i>CXCL1</i> , <i>CXCL6</i> , <i>CCL2</i> , <i>CCL20</i> , and <i>CX3CL1</i> , <i>IL17C</i> , <i>IL18</i> , <i>STAT1</i> , <i>NFKBIA</i> , <i>NFKBIZ</i> , <i>HIF1A</i> , <i>SOX4</i> , <i>ONECUT2</i>
 duct cells	<b>CEACAM6</b> <sup>141</sup>	-	■	-	progenitor-like linked to α-, δ-cells	<i>CEACAM6</i>
	<b>FTH1/FTL</b> <sup>141</sup>	-	■	-	progenitor-like linked to β-γ/PP-, acinar cells co-expression with insulin	<i>FTH1</i> , <i>FTL</i>
	<b>MUC1</b> <sup>45</sup>	■	-	-	mucous secretion, centroacinar	<i>MUC1</i> , <i>MUC20</i> , <i>TFF1</i> , <i>TFF2</i> , <i>PLAT</i>
	<b>CFTR</b> <sup>45</sup>	■	-	-	ion transport, terminal duct	<i>CFTR</i> , <i>CD44</i>

■	upregulation/high expression
■	downregulation/low expression
■	no expression
xxx	link to phenotype based on literature
-	no link in discussed single-cell publications

**Figure 4: New heterogeneities in pancreatic cell types.** Table summarizing the new concepts and markers of cellular heterogeneity in the pancreas that are discussed in the text.

[41]. ID transcription factors have been shown to be involved in the regulation of cell proliferation and inhibition of differentiation in several cell types [88]. Specifically, upregulation of IDs has also been associated with cancerous pancreatic tissue, expansion of ductal progenitors, and inhibition of endocrine cell differentiation [89,90]. Interestingly, *Id1* and *Id3* are also highly expressed in early postnatal β-cells, in which they are implicated in the regulation of ROS, an important driver of postnatal β-cell proliferation [76]. Further, two groups independently detected a human β-cell subpopulation expressing genes that function in the unfolded protein response (UPR), also known as endoplasmic reticulum (ER) stress [45,48]. Recent work has linked high insulin demand and β-cell proliferation to mild ER

stress. Sharma et al. suggest that β-cells sense insulin demand via the UPR mechanism, which triggers proliferation [91]. Importantly, the set of genes upregulated in this novel subtype overlap between the studies (e.g. *HERPUD1*, *DDIT3*, *TRIB3*, *PP1R15A*, *LURAP1L*, *ATF3*), which indicates their biological significance. In agreement, Zeng et al. observe a downregulation of ER stress related genes (e.g. *Atf4*, *C/EBP*, *Ddit3*) concomitant with downregulation of genes associated with proliferation during postnatal β-cell maturation in mice, further supporting a role of mild ER stress in the control of the proliferative response [76]. Together, these studies hint at proliferative β-cell subpopulations owing the potential to restore part of the β-cell mass and at underlying pathways including UPR/ER stress pathway, SRF- and ROS-mediated



signaling, as well as PDGF and JAK/STAT signaling, but, clearly, more detailed knowledge is required to target  $\beta$ -cell self-replication for therapeutic purposes.

#### 4.1.2. Triggering $\beta$ -cell maturation

The most stringent prerequisite to effectively restore metabolic homeostasis is to achieve full functionality and maturity of the newly generated  $\beta$ -cells. As part of naturally reversible events and aging,  $\beta$ -cells are expected to undergo a spectrum of changes in adulthood equivalent to varying differentiation states [92]. A mature  $\beta$ -cell phenotype is generally associated with high levels of Insulin (Ins) and expression of  $\beta$ -cell-specific glucose transporter 2 (Glut2) as well as of the transcription factors MafA and Nkx6.1 among others. In contrast, immature cells in mice are described by elevated expression of genes involved in early  $\beta$ -cell development (*MafB*, *Pax4*, *Pax6*, and, in extreme cases, *Ngn3*) and decreased insulin secretion, in brief a loss of key maturation factors needed for optimal function [3,93]. The identification of markers and a detailed characterization of the different stages would be of great benefit, since hyperglycemia in diabetes is thought to cause a fraction of  $\beta$ -cells to move toward a more primitive state, a process often termed dedifferentiation [3,92]. Recent work suggests that even in T1D a subpopulation expressing immature features persists long-term immune attack (for more detail see section 4.1.3) [61].

The presence of immature  $\beta$ -cells in adult islets has been reported in mice and human. Szabat et al. detected two stable subpopulations of PDX1 positive  $\beta$ -cells with distinct insulin levels [71]. 25% of human and mouse  $\beta$ -cells were PDX1<sup>+</sup>/Insulin<sup>low</sup> and displayed an immature expression profile together with an increased replication rate and diminished insulin secretion. Moreover, a significant fraction of these cells transition into a more mature state in culture. The recent findings from our lab further confirm the presence of immature or pre- $\beta$ -cells and link active WNT/planar cell polarity (PCP) signaling to  $\beta$ -cell maturation [69]. We showed that Flattop (Fltp), a downstream effector and reporter of the WNT/PCP pathway is heterogeneously expressed among pancreatic endocrine cells and subdivides  $\beta$ -cells in mice into Fltp Venus reporter (FVR)-negative proliferative cells (20%) and FVR-positive metabolically active cells (80%) [69,94]. Intriguingly, in-depth analysis revealed that FVR-negative  $\beta$ -cells show characteristics similar to the proposed PMPs (Insulin<sup>+</sup>, Glut2<sup>low</sup>), which suggests that proliferative potential varies among  $\beta$ -cells in the adult islet and indicates that pre- $\beta$ -cells might be enriched in the FVR-negative population. Moreover, a genome-wide transcript profiling array of FVR-positive and FVR-negative endocrine populations showed differential expression of WNT and MAPK signaling components,  $\beta$ -cell maturation markers, genes associated with G-protein coupled receptor (GPCR), hormones, proliferation markers, and glycolysis enzymes, which suggests that  $\beta$ -cell subpopulations indeed can be selectively targeted [69]. Interestingly, WNT and MAPK signaling components are highly expressed in both immature, postnatal  $\beta$ -cells, and FVR-negative endocrine cells [46,69]. However, Qui et al. did not detect differences in the transcriptional profile of *Fltp mRNA*-positive and *Fltp mRNA*-negative  $\beta$ -cells [46]. The reason is due to the transient expression of the *Fltp mRNA* during WNT/PCP acquisition [[69,94] and Böttcher et al. in preparation]. As already speculated by the authors, the WNT/PCP pathway acts preferentially at the post-transcriptional level to establish planar polarization, which is accompanied by the acquisition of distinct physiological features. Therefore, while differences at the transcriptional level are minor, post-translational WNT/PCP signaling is essential to trigger a mature  $\beta$ -cell phenotype. These studies emphasize that the analysis of proteins and post-translational

modifications in single cells is warranted to reveal in-depth knowledge of physiology and disease.

Together, these results demonstrate that  $\beta$ -cells pass through different maturation states in adult islets. This implies that promotion of both replication and maturation is required to obtain functional, insulin-producing  $\beta$ -cells. The FVR-negative cells described above show a combined proliferative and immature phenotype and eventually mature into insulin-secreting cells, which is triggered by active WNT/PCP signaling [69]. Similarly, a number of recent single-cell studies reported on heterogeneities in expression of genes involved in insulin regulation and  $\beta$ -cell development. Apart from differential expression of *ID* genes (see above) the Sandberg group detected two clusters with high levels of *serum retinol-binding protein (RBP4)* and *GPR120* (also known as *FFAR4*) [41]. The adipokine RBP4 promotes insulin resistance and is increased in obese and T2D individuals [95], whereas engagement of GPR120 induces insulin secretion [96]. Strikingly, Baron et al. also identified a  $\beta$ -cell state with elevated expression of *RBP4* concomitant with the expression of marker genes of mature  $\beta$ -cells [45]. Further, Muraro et al. described a  $\beta$ -cell subpopulation expressing ER stress genes indicative of less mature cells [48]. Looking at the expression fold change in the published data of these cells shows a slight decrease of *RBP4*, which is, however, not significant. Interestingly, *RBP4* expression is also downregulated in a  $\beta$ -cell subpopulation resistant to the immune response associated with T1D [61]. Together, this suggests a role of RBP4 also in homeostatic regulation of insulin secretion, which is potentially impaired in diabetes and contributes to dysfunction and disease. While *RBP4* shows different expression levels, *GPR120* seems to be completely absent in a fraction of cells [41]. Given its function in insulin release, GPR120 is an interesting candidate marker for functional and mature  $\beta$ -cells. In this context, it is important to note that cells might not clearly separate into distinct maturation states due to the continuous nature of the maturation process and the presence of transitioning cells. Indeed, Baron et al. describe rather gradual changes in gene expression over the whole  $\beta$ -cell population, instead of clear on/off states [45]. Taken together, single-cell analysis identified new potential markers of mature  $\beta$ -cells. However, to molecularly dissect  $\beta$ -cell heterogeneity and to better understand and drive  $\beta$ -cell maturation we need more information on regulatory elements such as transcription factors and signaling molecules, which are low in expression but might have a strong effect on cell fate and state (Figure 4).

#### 4.1.3. $\beta$ -Cell heterogeneity in the pathogenesis of diabetes

The available data clearly support the presence of  $\beta$ -cell heterogeneity in adult islets; its role in the pathogenesis of diabetes mellitus, however, is not fully understood. Single-cell studies considerably extended the list of genes with aberrant expression in  $\beta$ -cells of diabetic patients. Most of the recent single-cell studies, however, failed to detect  $\beta$ -cell subpopulations and thus ignore the evidently present heterogeneity [42,51,63]. Distinct  $\beta$ -cell subpopulations might respond differently to diabetic stressors. Likewise, not all cells are equally susceptible to therapeutic interventions. In addition, metabolic stress itself might contribute to heterogeneity, as novel dysfunctional  $\beta$ -cell subtypes can arise or mature cells dedifferentiate to a more immature state. Indeed, Baron et al. found indication of a disappearing  $\beta$ -cell subtype under hyperglycemic conditions [45]. They detected downregulation of ER stress marker genes such as *Herpud1* in  $\beta$ -cells in the diabetic state that were associated to a  $\beta$ -cell subpopulation in healthy pancreatic islets [45]. Accordingly, Dorell et al. describe two surface markers, ST8SIA1 and CD9, that discriminate four  $\beta$ -cell subpopulations in the human adult islets [68]. All four populations

expressed common  $\beta$ -cell markers but displayed differences in insulin secretion rates and in their gene expression profiles. Importantly, the subtype distribution was altered in T2D islets highlighting the relevance of the described  $\beta$ -cell heterogeneity and its implication in disease. In line with these results, three distinct  $\beta$ -cell groups were identified that shift in number in T2D or with age and BMI [74]. All three studies could not clarify whether altered partitions of cells were a cause or consequence of  $\beta$ -cell dysfunction. In addition, it remains unclear whether these subtypes have a temporal relationship or are independent lineages and whether distorted proportions result from the selective loss of a  $\beta$ -cell subtype and/or from cellular transitions. More detailed molecular profiles of these  $\beta$ -cell subgroups are required to determine their biological significance, function, ontogeny, and implication in disease. A challenge in this regard is the high donor-to-donor variation that was detected by multiple single-cell studies [41,45,74]. Conclusions on islet cell composition in homeostasis and disease, therefore, should only be drawn if they are robust over multiple donors.

Records of subpopulation-specific expression profiles also would help to solve the debate on the contribution of partial dedifferentiation to the reduction of functional  $\beta$ -cell mass in T2D.  $\beta$ -cell dedifferentiation is characterized by diminished expression of  $\beta$ -cell specific transcription factors (e.g. *PDX1*, *NKX6.1*, *MAFA*), reactivation of developmental genes (e.g. *Ngn3*) and, surprisingly, pluripotency genes (e.g. *Oct4*, *Nanog*, *L-myc*), as well as decreased insulin production [3,62,92,97] and expression of disallowed genes that are silenced in healthy  $\beta$ -cells including *lactate dehydrogenase A (Ldha)*, the *monocarboxylate carrier (Mct1)*, as well as *glucose-6-phosphatase (G6pc)*, and *hexokinase I (Hk1)* [98,99]. Recent studies demonstrated that acquiring and maintaining  $\beta$ -cell identity and function also requires sustained repression of other endocrine gene programs and, thus, give an explanation on how adult  $\beta$ -cells can partially dedifferentiate or become reprogrammed into other islet endocrine cells [27,28,100]. Indeed, Wang et al. found evidence of dedifferentiated cells in diabetic islets when comparing their transcriptional profiles to cells of healthy patients [63].  $\beta$ -cells of adult T2D individuals exhibit transcriptional patterns of juvenile endocrine cells indicating partial regression to an immature state characterized by a less well-defined  $\alpha$ - and  $\beta$ -cell gene signature and expression of *CDKN2B*, *BARD1*, *JUNB*, and *PRKD1*. Segerstolpe et al. reported significantly lower *INS* mRNA levels in T2D  $\beta$ -cells, a feature of dedifferentiated  $\beta$ -cells [41]. However, the authors of both studies did not comment on the expression of *NGN3*, *FOXO1*, or other factors associated with  $\beta$ -cell dedifferentiation.

In a recent study, a novel murine  $\beta$ -cell subpopulation was described that develops during progression of T1D in response to immune cell activity and persists for a long time [61]. These cells resist immune assault, exhibit increased proliferation rates, and acquire stem-like and immature features while the expression of mature  $\beta$ -cell marker genes, insulin content, and diabetes antigens is decreased. A similar, less-differentiated  $\beta$ -cell subpopulation appeared when human islets were co-cultured with allogeneic lymphoid cells. These changes in gene expression likely account for the long-term survival of a few  $\beta$ -cells in T1D patients [61]. Together, these studies indicate that a group of  $\beta$ -cells adapt to metabolic and immune stressors in T2D and T1D, respectively, by regression to an immature state. To what degree dedifferentiation occurs in diabetes and whether  $\beta$ -cells revert to a multipotent precursor or a novel, reversible dedifferentiated state will have to be established [62,101]. Nevertheless, dedifferentiation clearly contributes to the etiology of diabetes together with  $\beta$ -cell death and creates a pool of cells that can possibly be triggered to redifferentiate

and reestablish islet function. Since  $\beta$ -cell dedifferentiation is associated with and possibly caused by hyperglycemia, reconstitution of normoglycemia might restore a normal mature  $\beta$ -cell phenotype [92]. In summary,  $\beta$ -cell subpopulations show varying responses to metabolic stress; thus, heterogeneity needs to be considered to gain an understanding of the pathomechanisms of diabetes and to identify therapeutic targets (Figure 4).

#### 4.2. Intra-islet cell transdifferentiation

Recent findings in mice have revealed unexpected plasticity and potential of intra-islet (e.g.  $\alpha$ -cell) and extra-islet (acinar and ductal) cells to transdifferentiate and produce insulin. Transdifferentiation is defined as a process whereby a differentiated cell is converted into another type of cell either directly via a double hormone-positive intermediate cell or through a dedifferentiated state. Examples for both forms exist in the pancreas and are reviewed in Puri et al. [12].

##### 4.2.1. $\alpha$ - to $\beta$ -cell conversion

$\alpha$ -cells secrete the hormone glucagon, which induces glycogenolysis to elevate blood glucose levels. Importantly,  $\alpha$ -cells are more resistant to metabolic stress than  $\beta$ -cells, and their number does not significantly change in T1D and T2D patients [41,51,63,102]. Considering the close developmental and physiological relationship between  $\alpha$ - and  $\beta$ -cells as well as the big overlap of their transcriptome,  $\alpha$ -cell transdifferentiation draws much attention as a source for  $\beta$ -cell regeneration. Transdifferentiation of  $\alpha$ -cells into  $\beta$ -cells has been observed by several groups in different experimental settings [47,103–109]. Single gene manipulations suffice to induce  $\alpha$ -cell conversion towards the  $\beta$ -cell fate, as first shown by Collombat et al. through ectopic overexpression of Pax4 [103]. Interestingly, besides forced expression of key transcription factors,  $\beta$ -cell depletion after puberty drives the conversion of remaining  $\alpha$ -cells albeit by a different mechanism; nevertheless indicating a natural regeneration mechanism that could be triggered in diabetic conditions [104]. The recent identification of urocortin3-negative, insulin-expressing  $\beta$ -cells (1–2% of all  $\beta$ -cells) that are transcriptionally (lack expression of e.g. *G6pc2*, *Eroib*, and *Glut2*) and functionally (do not sense glucose) immature suggests a naturally occurring  $\alpha$ - into  $\beta$ -cell conversion as part of islet homeostasis [109]. Lineage-tracing studies revealed that these cells represent an intermediate stage in both the transdifferentiation of  $\alpha$ -cells to mature  $\beta$ -cells as well as the inverse transition from  $\beta$ -cells to functional  $\alpha$ -cells. Interestingly, transdifferentiation takes place within a specialized neogenic niche at the periphery of healthy islets, which might exist in humans, at least at a younger age, as well [109]. Other studies also reported on ‘intermediate’ cells expressing both  $\alpha$ - and  $\beta$ -cell markers in human islets [70]. The transcriptional profile of double hormone-positive cells could give insights into the mechanisms underlying  $\alpha$ - to  $\beta$ -cell conversion and unravel driving factors. Indeed, such rare double hormone-positive cells were captured by single-cell RNA sequencing studies but not further commented on or excluded as doublet cells [41,63].

The relative ease with which  $\alpha$ -cells can be transdifferentiated into  $\beta$ -cells may be partly explained by their plastic epigenomic state.  $\alpha$ -cells have hundreds of bivalent activating and repressing histone marks on developmental genes, strikingly similar to the histone modification map of human embryonic stem cells, indicating an undifferentiated multipotent epigenomic state [105]. Indeed, inhibition of histone methyltransferases resulted in the appearance of bi-hormonal, insulin-, and glucagon-positive cells. Analysis of open chromatin in  $\alpha$ - and  $\beta$ -cells using the assay for transposase-accessible chromatin with high th-

roughput sequencing (ATAC-Seq) confirmed a bivalent chromatin state in  $\alpha$ -cells [47]. In addition,  $\alpha$ -cells have more open chromatin regions compared to  $\beta$ -cells, of which many are associated with  $\beta$ -cell signature genes. Interestingly, simultaneous inactivation of the  $\alpha$ -cell regulator *Arx* and DNA methyltransferase 1 (*Dnmt1*) is sufficient to promote rapid conversion of  $\alpha$ -cells into  $\beta$ -like cells capable of insulin production and secretion [110], which substantiates that the conversion process involves epigenetic changes in addition to modulation of cell-type specific master regulators. Single-cell RNA sequencing and functional evaluation revealed strikingly little difference between converted and native  $\beta$ -cells but uncovered cells retaining  $\alpha$ -cell character suggesting that not all  $\alpha$ -cells are equally susceptible to reprogramming [110].

Two recently published studies report that stimulation of  $\gamma$ -aminobutyric acid (GABA) signaling can induce  $\beta$ -cell regeneration and reverse severe diabetes in mice [106,107]. Ben-Othman et al. describe that long-term administration of GABA induces  $\alpha$ -cell mediated  $\beta$ -like cell neogenesis *in vivo* [106]. Precisely, GABA triggers conversion of  $\alpha$ -cells to functional  $\beta$ -like cells through downregulation of *Arx*, the transcriptional repressor of *Pax4*. This, in turn, stimulates  $\alpha$ -cell regeneration mechanisms involving the reactivation of *Ngn3*-controlled endocrine developmental processes and increases proliferation of duct-lining progenitor cells, which gives rise to new  $\alpha$ -cells. The newly generated  $\alpha$ -cells are subsequently converted into  $\beta$ -cells upon prolonged GABA exposure. GABA treatment results in increased islet number and size, and  $\beta$ -cell mass could be repeatedly replaced in chemically induced diabetes. Most important, human islets responded similarly to GABA, indicating that the findings in mice indeed might be translatable to humans. In a complementary study, Li et al. found that the antimalarial drug artemether exerts similar effects by binding to gephyrin, a protein associated with the GABA receptor complex [107]. Together these studies suggest a therapeutic potential of GABA pathways, but, clearly, further research is needed to confirm these results and to clarify how GABA acts in diabetic human islets. Although,  $\alpha$ -cell mass and metabolism seem not to be strongly affected by diabetes, gene expression is clearly changed. Up to 200 differentially expressed genes between healthy and T2D islets were identified by single-cell transcriptomics, of which approximately 35% overlap between different studies [41,42,51,63]. This suggests altered  $\alpha$ -cell states, which might influence GABA treatment efficiency. Thus, whether adaptive neogenesis of  $\alpha$ -cells induced by GABA is possible in humans remains to be shown.

Besides  $\alpha$ -cell replacement from exocrine tissue, a reserve  $\alpha$ -cell pool that converts into  $\beta$ -cells upon metabolic stress could also be established by triggering  $\alpha$ -cell proliferation. Strikingly, of all major endocrine cell types,  $\alpha$ -cells show the highest basal proliferation rate as well as the most robust mitotic response to the mitogen harmine across healthy and T2D donors [74]. Concurrently with these findings, single-cell transcriptomics identified and enabled transcriptional profiling of rare proliferating  $\alpha$ -cells [41,46,63]. The Kaestner group identified a single proliferating  $\alpha$ -cell out of 190 annotated  $\alpha$ -cells, which showed high expression of the proliferation marker *Ki67*, activation of cell cycle pathways and inhibition of cell cycle checkpoint control genes [63] (Figure 4). In addition, both *DYRK1A* and *GSK3 $\beta$*  were downregulated, which is consistent with their suggested role as inhibitors of endocrine cell replication [86,111]. Common targets of both proteins are the *GLI* transcription factors, which implicates modulation of the Sonic Hedgehog (SHH) signaling pathway in replicating  $\alpha$ -cells [112–114]. Segerstolpe et al. also detected a small population of proliferative  $\alpha$ -cells with increased expression of

proliferation-associated genes and slightly reduced levels of several  $\alpha$ -cell specific markers [41]. This proliferating subtype was further distinguished by 439 significantly differentially expressed genes. Interestingly, when looking more closely at the list of the differential gene expression analysis, we see that the transcriptional regulator of the SHH pathway *GLI2* is enriched, however not significantly. Moreover, the expression of *Stathmin (STMN1)*, which is associated with dividing acinar cells (see below) and progenitor cells in other organs, is significantly increased in proliferative  $\alpha$ -cells [115]. Taken together, the high-dimensional single-cell profiles provide information on activated and repressed pathways in replicating cells and potential targets to trigger  $\alpha$ -cell proliferation.

The accumulating evidence supports the model that  $\alpha$ -cells represent a natural source of new  $\beta$ -cells.  $\alpha$ - to  $\beta$ -cell transdifferentiation can be induced by rather simple genetic and epigenetic manipulations or drug treatment and to some extent occurs spontaneously in homeostasis and upon  $\beta$ -cell depletion or in T1D. Together with the higher rate of proliferation, the bivalent histone modifications at loci of developmental genes and open-chromatin regions in  $\beta$ -cell genes, this makes  $\alpha$ -cells a tempting target for future clinical applications in T1D and T2D and imply that  $\alpha$ -cells might have an important function besides glucagon secretion.

### 4.3. Extra-islet cell conversion

#### 4.3.1. Acinar to $\beta$ -cell conversion

Acinar cells are the major constituent of the adult pancreatic tissue and responsible for the production and secretion of digestive enzymes. Owing to their abundance and origin from common pancreatic progenitor cells, acinar cells represent an appealing pool for  $\beta$ -cell replacement. Several studies reported successful conversion of acinar cells into endocrine cells *in vivo* and *in vitro* either via a pancreatic progenitor state or by direct lineage switching [12]. Acinar-to- $\beta$ -cell reprogramming can be stimulated by ectopic expression of specific transcription factors [116,117], administration of signaling and growth factors [118–120] and co-transplantation with fetal pancreatic cells [121–123]. The Melton group was the first to show successful *in situ* formation of cells with  $\beta$ -cell like function and morphology in mice by the expression of  $\beta$ -cell transcription factors (*Pdx1*, *Ngn3*, *MafA*) in acinar cells [116]. Importantly, human acinar cells also can be induced to transdifferentiate into insulin-producing  $\beta$ -cells [124]. The contribution of acinar cells to  $\beta$ -cell regeneration *in vivo*, for example after injury, remains unclear. Moreover, inconsistencies exist as to whether acinar cells assume a  $\beta$ -cell state through transdifferentiation or dedifferentiation to a facultative progenitor or both [125]. Genetic-lineage tracing experiments using acinar specific promoters could not yet clearly identify the origin of new  $\beta$ -cells after triggering cell conversion [119,120]. To harness acinar cell reprogramming towards  $\beta$ -cell production in diabetic patients, we need a better understanding of the factors that control the lineage conversion process and clarify if and to what extent it is a natural process in regeneration. To this end, high-dimensional analyses at the single-cell level can give important insights. Until now, the existence of possible heterogeneity in acinar cells that might influence reprogramming efficiency has been mostly neglected in conversion protocols. Indeed, distinct acinar subclusters have been identified in single-cell studies [41,48]. In a very recent study, Wollny et al. found a progenitor-like acinar subpopulation in rodents with the capacity for long term self-renewal in homeostasis by lineage-tracing and organoid-formation assays [115]. Single-cell RNA sequencing of acinar cells and immunohistochemistry supported these

results and revealed a small subpopulation of dividing cells characterized by high expression of STMN1 (Figure 4). Stathmin plays an important role in the regulation of the cytoskeleton by destabilizing microtubules and has been identified as a marker for progenitor cells in other organs [126]. A subset of the STMN1-positive cells also expressed the transcription factor Sox9 that has been associated with exocrine progenitor cells of the pancreas [127]. Strikingly, the number of STMN1-positive acinar cells is increased upon injury, suggesting transient acquisition of proliferative capacity by normally non-proliferative cells. In line with these results, Segerstolpe et al. found a small subset of acinar cells (~4%) expressing proliferation markers [41]. They did not investigate these cells in detail, and it is not known whether they represent the proliferating acinar cells described by Wollny et al. [115]. A progenitor-like exocrine subtype could be a target for endogenous  $\beta$ -cell regeneration, and it would be interesting to see if the transcriptional profile of the proliferating acinar cells resembles a cell state of the acinar- $\beta$ -cell reprogramming. Other groups also reported on acinar cell heterogeneity in single-cell transcriptomics data. Muraro et al. described differential expression of a member of the regenerating (Reg) protein family, REG3A, among acinar cells [48]. Reg proteins, first discovered in pancreatic inflammation and islet regeneration, exert anti-inflammatory, anti-apoptotic, and mitogenic effects in various physiological and disease contexts and are involved in differentiation and proliferation of various cell types [128]. REG3A in particular has been suggested to be involved in tumorigenesis of pancreatic cells by regulation of key genes and pathways implicated in cell growth [129]. Strikingly, REG3A-positive acinar cells formed cell patches close to the islets and showed lower expression levels of the acinar cell marker PRSS1 and key acinar genes encoding secretory enzymes (*CEL*, *CELA3A*, *CELA3B*, *AMY*). The reduced expression of functional acinar genes but increased level of proliferative signaling might suggest a more immature cell state potentially harboring the actively dividing acinar cells. More detailed analyses are required to confirm the existence of these cell subtypes and fully characterize them, to draw connections between them, and to investigate their function and plasticity. Proliferative or progenitor-like exocrine cells also may provide clues on regeneration mechanisms in endocrine cells since they are derived from a common progenitor. Furthermore, recent evidence indicates that acinar reprogramming efficiency is significantly reduced upon pancreatic inflammation or hyperglycemia but improved by inhibition of contact-mediated signaling [125,130–132]. This highlights the importance of investigating extrinsic factors that potentially hamper or enhance cell conversion *in vivo*. For example, acinar cells have been demonstrated to produce different inflammatory mediators as part of the first immune response to injury [133]. Consistently, Segerstolpe et al. detected a functionally distinct subpopulation that showed increased inflammatory markers in single-cell data [41]. If and how they influence reprogramming is an open question.

Taken together, acinar cells show plasticity *in vivo* and *in vitro* and could serve as a pool for  $\beta$ -cell regeneration. Importantly, results from rodents were translatable to human. However, the often contradictory experimental outcomes suggest that heterogeneity, contact-mediated lateral signals, the microenvironment, and physiological conditions also influence reprogramming and need to be understood to efficiently and reliably convert acinar into  $\beta$ -cells. Of note, genetic reprogramming and loss of acinar cell identity are critical early drivers of pancreatic ductal adenocarcinoma (PDA) formation, one of the deadliest malignancies [134,135]. Therefore, caution must be taken when manipulating acinar cell fate or triggering acinar cell proliferation.

#### 4.3.2. Duct cell conversion into $\beta$ -cells

The pancreatic duct drains the exocrine fluid into the duodenum and produces bicarbonate to regulate the luminal pH of acinar and duct cells [136]. Strikingly, ductal cells might have an important function beyond their physiological implication by providing a reserve pool of progenitor cells with the potential to give rise to endocrine cells, a process known as neogenesis. During the neonatal period in rodents, the  $\beta$ -cell pool further expands by  $\beta$ -cell replication as well as by neogenesis from ductal cells [79]. Neogenesis has also been reported to occur in certain experimental conditions such as ectopic expression of Pax4 in  $\alpha$ -cells, upon which duct-lining cells sense glucagon-shortage and reactivate Ngn3 expression and associated developmental pathways to generate  $\alpha$ -cells that eventually transdifferentiate into  $\beta$ -cells [103,137]. Xu et al. found that duct cells can give rise to new  $\beta$ -cells upon pancreatic duct ligation [138]. Furthermore, cultured murine adult pancreatic duct-like cells could be directly reprogrammed to insulin-producing  $\beta$ -like cells by adenoviral delivery of *Pdx1*, *Ngn3*, and *MafA* [139].

Whether adult duct lining cells also contribute to homeostatic renewal of  $\beta$ -cells is still under debate. Dor et al. concluded from their studies that replication and not neogenesis is the mechanism of  $\beta$ -cell expansion in adult mice [80]. In contrast, Seaberg et al. showed that besides islet cells, adult duct cells contain a population of pancreatic multipotent progenitors (PMP) that can give rise to all pancreatic lineages as well as neural lineages [81]. In agreement, Grün et al. identified two ductal cell clusters with a high multipotency score in the adult human pancreas using their newly developed StemID algorithm to detect potential stem cell populations within heterogeneous cell populations [140]. The inferred pancreatic lineage tree implies that distinct subtypes of ductal cells give rise to different endocrine subtypes and acinar cells. Precisely, they found that the cell cluster characterized by high *CEACAM6* expression, is linked to  $\alpha$ - and  $\delta$ -cells, whereas a cell cluster highly expressing the ferritin subunits *FTH1* and *FTL* is linked to  $\beta$ -,  $\gamma$ /PP- and acinar cells (Figure 4). Moreover, the authors also observed ductal as well as  $\beta$ -cells that co-express insulin and FTL, suggesting that ferritin-positive ductal cells differentiate into mature  $\beta$ -cells and implying that neogenesis contributes to homeostatic endocrine cell renewal. The ductal cell subpopulations reported by Baron et al. that are characterized by the expression of *CFTR* or *TFF1*, *TFF2*, and *MUC1*, most likely reflect their localization within the duct and form the terminal duct or connect to the acinus, respectively [45] (Figure 4). How these two duct cell populations relate to ductal PMPs or to the ductal subpopulations described by Grün et al. is not clear. Taken together, genetically manipulated mice show that neogenesis can be induced upon injury. Now, there is new evidence that ductal neogenesis not only occurs in the neonatal period and to some extent in homeostasis and upon injury in rodents but might indeed be part of homeostatic renewal of endocrine cells in humans [78]. Further work is necessary to identify the signals that mobilize ductal progenitors to provide a route to replenish  $\beta$ -cells *in situ*. In rodents, ductal cells convert either into  $\beta$ - or  $\alpha$ -cells depending on the injury mode. To identify the signals and transcriptional program that specifically drive ductal into  $\beta$ -cell conversion, we need to study the behavior of ductal cells in different injury models at the single-cell level. Notably, the presence of ductal progenitor cell giving rise to  $\beta$ -cells could also explain part of the observed  $\beta$ -cell heterogeneity.

## 5. CONCLUSION

Here, we reviewed recent literature on single-cell analysis of pancreatic cell types and discussed the findings with respect to

endogenous  $\beta$ -cell regeneration routes. Single-cell genomics in the human pancreas is still at an early stage. Nevertheless, it has already provided new insights into the transcriptional program of pancreatic cells including those of rare cell types such as  $\delta$ - and  $\gamma$ /PP cells. Expression of the leptin receptor and diabetes risk genes, such as *PDX1*, *HADH*, and *UCP2* in  $\delta$ -cells for instance strongly suggests an important role of  $\delta$ -cells in the maintenance of glucose homeostasis despite their low number and demands to study  $\delta$ -cells in more detail [41,45,48,51].

One of the most promising strategies to restore  $\beta$ -cell mass is stimulation of  $\beta$ -cell replication and/or maturation. Single-cell transcriptomics data on postnatal, murine  $\beta$ -cells revealed the signature of immature  $\beta$ -cells and implicate ROS, ER stress, SRF, MAPK, TGF- $\beta$ , WNT, and PDGF signaling in the regulation of postnatal  $\beta$ -cell proliferation and maturation [46,76]. Strikingly, PDGF and MAPK signaling as well as ER stress have also been associated with potentially proliferative and regenerative adult human  $\beta$ -cell subpopulations [45,48,74]. Thus, modulating these pathways could be a strategy for reactivating and promoting the expansion and maturation of residual  $\beta$ -cells in diabetic patients.

Reprogramming of intra-islet cells into  $\beta$ -cells is an alternative approach considered for  $\beta$ -cell regeneration. New data implicate a contribution of  $\alpha$ - into  $\beta$ -cell conversion as well as ductal neogenesis to homeostatic  $\beta$ -cell renewal in mice [109,140]. Since severe  $\beta$ -cell depletion is known to trigger transdifferentiation of  $\alpha$ -cells into  $\beta$ -cells in mice and  $\alpha$ -cells are the most proliferative cell-type among the endocrine cells, employing  $\alpha$ -cells might be a promising approach for  $\beta$ -cell regeneration [74,104]. Therefore, identifying the signals that drive  $\alpha$ -cell proliferation and  $\alpha$ -cell into  $\beta$ -cell conversion in homeostasis and upon injury is crucial. Proliferative adult  $\alpha$ -cells have been captured by single-cell RNA sequencing and SHH signaling has been identified as a candidate pathway regulating  $\alpha$ -cell proliferation [41,63,74].

Additionally, single-cell transcriptomics revealed species-specific differences in gene expression important to consider when designing new therapies for diabetes.

In conclusion, single-cell studies on pancreatic cells identified several genes and pathways critical for driving proliferation and maturation of  $\beta$ -cells and shed light on intra-endocrine cell heterogeneity, hint at underlying molecular mechanisms, and describe potential regenerative subpopulations and therapeutic targets.

## 6. FUTURE PERSPECTIVE

To understand the pathomechanisms of T1D and T2D and find potential routes to restore  $\beta$ -cell mass - the basis for diagnosis and therapy - we must determine the factors that shape  $\beta$ -cell identity, drive differentiation, and maintain function and plasticity in the pancreas. At the cellular level, these include the molecular phenotype and response mechanisms to external stimuli that regulate function, the spatial position and local environment, and the developmental history and maturation state [141]. Current studies almost exclusively have focused on descriptive analyses and statistical pattern identification, but they lack mechanistic insights. Single-cell analysis can infer function through correlation but will not substitute thorough functional validation by interventional analysis and perturbation experiments. New CRISPR-based technologies for gene manipulation in single-cells are promising approaches to obtain functional answers and fill the gap between the molecular profile and actual phenotype of a cell [142–145]. In addition, the transcriptome only hints at active processes and pathways but needs to be connected to a cell's proteome and epigenome to make conclusive statements on its

functional state and reveal post-translational mechanisms that can be targeted in therapy. Finally, and most importantly, the single-cell information needs to be mapped back onto healthy and diseased tissue as well as onto the temporal trajectory of differentiation. Imaging methods such as single-molecule RNA fluorescence *in situ* hybridization (FISH) [146,147] and imaging mass-cytometry [148] enable spatially resolved quantification of single mRNA molecules and proteins and connect a cell's phenotype to neighboring cells and the environment. Time-resolved data elucidate the genes controlling every single step in differentiation and maturation, important information lacking in end-point analyses. Single-cell methods and lineage-tracing in combination with organ on a chip approaches or *in vitro* differentiation approaches [7,149] will allow to follow the temporal progression of a cell, either live or in pseudotime reconstructed by machine learning algorithms.

Insights from pioneering single-cell studies of pancreatic cells have illustrated the potential of single-cell data. Now, we need to go beyond collecting data and simple descriptive, correlative analyses and integrate cellular profiles over multiple omics layers in space and time. This will pave the way to understand the mechanisms underlying diabetes and to design strategies for *in vivo* regeneration of functional islets.

## ACKNOWLEDGMENTS

We apologize to all colleagues that we could not mention in this review due to the space constraints. The authors' research is funded by an Emmy-Noether Fellowship; the European Union (grant agreement No. 602587); the Helmholtz Alliance 'Imaging and Curing Environmental Metabolic Diseases'; the Helmholtz Association; the Helmholtz Society; the Helmholtz Portfolio Theme 'Metabolic Dysfunction and Common Disease'; the German Research Foundation; and the German Center for Diabetes Research. S.T. is supported by a DFG Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). We would like to thank Ralph Böttcher, Pallavi Mahaddalkar, Adriana Migliorini, and Alexandra Aliluev for helpful comments on the manuscript.

## CONFLICT OF INTEREST

None declared.

## REFERENCES

- [1] Guariguata, L., Whiting, D.R., Hambleton, I., Beagley, J., Linnenkamp, U., Shaw, J.E., 2014. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Research and Clinical Practice* 103(2):137–149. <http://dx.doi.org/10.1016/j.diabres.2013.11.002>.
- [2] Butler, A.E., Janson, J., Bonner-Weir, S., Ritzel, R., Rizza, R.A., Butler, P.C., 2003. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* 52(1):102–110. <http://dx.doi.org/10.2337/diabetes.52.1.102>.
- [3] Talchai, C., Xuan, S.H., Lin, H.V., Sussel, L., Accili, D., 2012. Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. *Cell* 150(6):1223–1234. <http://dx.doi.org/10.1016/j.cell.2012.07.029>.
- [4] Shapiro, A.M.J., Lakey, J.R.T., Ryan, E.A., Korbitt, G.S., Toth, E., Warnock, G.L., et al., 2000. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *The New England Journal of Medicine* 343(4):230–238. <http://dx.doi.org/10.1056/Nejm200007273430401>.
- [5] Bruni, A., Gala-Lopez, B., Pepper, A.R., Abualhassan, N.S., James Shapiro, A.M., 2014. Islet cell transplantation for the treatment of type 1 diabetes: recent advances and future challenges. *Diabetes, Metabolic*

- Syndrome and Obesity: Targets and Therapy 7:211–223. <http://dx.doi.org/10.2147/DMSO.S50789>.
- [6] D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., et al., 2006. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature Biotechnology* 24(11):1392–1401. <http://dx.doi.org/10.1038/nbt1259>.
- [7] Pagliuca, F.W.W., Millman, J.R.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H.H., et al., 2014. Generation of functional human pancreatic  $\beta$  cells in vitro. *Cell* 159(2):428–439. <http://dx.doi.org/10.1016/j.cell.2014.09.040>.
- [8] Millman, J.R., Xie, C., Van Dervort, A., Gürtler, M., Pagliuca, F.W., Melton, D.A., 2016. Generation of stem cell-derived  $\beta$ -cells from patients with type 1 diabetes. *Nature Communications* 7:11463. <http://dx.doi.org/10.1038/ncomms11463>.
- [9] Russ, H.A., Parent, A.V., Ringler, J.J., Hennings, T.G., Nair, G.G., Shveygert, M., et al., 2015. Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *The EMBO Journal* 34(13):e201591058. <http://dx.doi.org/10.15252/emj.201591058>.
- [10] Migliorini, A., Bader, E., Lickert, H., 2014. Islet cell plasticity and regeneration. *Molecular Metabolism* 3(3):268–274. <http://dx.doi.org/10.1016/j.molmet.2014.01.010>.
- [11] Roscioni, S.S., Migliorini, A., Gegg, M., Lickert, H., 2016. Impact of islet architecture on beta-cell heterogeneity, plasticity and function. *Nature Reviews Endocrinology* 12(12):695–709. <http://dx.doi.org/10.1038/nrendo.2016.147>.
- [12] Puri, S., Folias, A.E., Hebrok, M., 2015. Plasticity and dedifferentiation within the pancreas: development, homeostasis, and disease. *Cell Stem Cell*, 18–31. <http://dx.doi.org/10.1016/j.stem.2014.11.001>.
- [13] Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., Teichmann, S.A., 2015. The technology and biology of single-cell RNA sequencing. *Molecular Cell* 58(4):610–620. <http://dx.doi.org/10.1016/j.molcel.2015.04.005>.
- [14] Stegle, O., Teichmann, S.A., Marioni, J.C., 2015. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics* 16(3):133–145. <http://dx.doi.org/10.1038/nrg3833>.
- [15] Haghverdi, L., Buttner, M., Wolf, F.A., Büttner, F., Theis, F.J., 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods* 13(10). <http://dx.doi.org/10.1038/nmeth.3971>, 845–+.
- [16] Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., et al., April 2016. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature Biotechnology* 34:1–14. <http://dx.doi.org/10.1038/nbt.3569>.
- [17] Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* 32(4):381–386. <http://dx.doi.org/10.1038/nbt.2859>.
- [18] Pan, F.C., Wright, C., 2011. Pancreas organogenesis: from bud to plexus to gland. *Developmental Dynamics* 240(3):530–565. <http://dx.doi.org/10.1002/dvdy.22584>.
- [19] Shih, H.P., Wang, A., Sander, M., 2013. Pancreas organogenesis: from lineage determination to morphogenesis. *Annual Review of Cell and Developmental Biology* 29(1):81–105. <http://dx.doi.org/10.1146/annurev-cellbio-101512-122405>.
- [20] Zhou, Q., Law, A.C., Rajagopal, J., Anderson, W.J., Gray, P.A., Melton, D.A., 2007. A multipotent progenitor domain guides pancreatic organogenesis. *Developmental Cell* 13(1):103–114. <http://dx.doi.org/10.1016/j.devcel.2007.06.001>.
- [21] Johansson, K.A., Dursun, U., Jordan, N., Gu, G., Beermann, F., Gradwohl, G., et al., 2007. Temporal control of Neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. *Developmental Cell* 12(3):457–465. <http://dx.doi.org/10.1016/j.devcel.2007.02.010>.
- [22] Shih, H.P., Kopp, J.L., Sandhu, M., Dubois, C.L., Seymour, P.A., Grapin-Botton, A., et al., 2012. A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. *Development* 139(14):2488–2499. <http://dx.doi.org/10.1242/dev.078634>.
- [23] Collobat, P., Mansouri, A., Hecksher-Sorensen, J., Serup, P., Krull, J., Gradwohl, G., et al., 2003. Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes & Development* 17(20):2591–2603. <http://dx.doi.org/10.1101/gad.269003>.
- [24] Nishimura, W., Kondo, T., Salameh, T., El Khattabi, I., Dodge, R., Bonner-Weir, S., et al., 2006. A switch from MafB to MafA expression accompanies differentiation to pancreatic beta-cells. *Developmental Biology* 293(2):526–539. <http://dx.doi.org/10.1016/j.ydbio.2006.02.028>.
- [25] Artner, I., Hang, Y., Guo, M., Gu, G., Stein, R., 2008. MafA is a dedicated activator of the insulin gene in vivo. *The Journal of Endocrinology* 198(2):271–279. <http://dx.doi.org/10.1677/JOE-08-0063>.
- [26] Artner, I., Hang, Y., Mazur, M., Yamamoto, T., Guo, M., Lindner, J., et al., 2010. MafA and MafB regulate genes critical to beta-cells in a unique temporal manner. *Diabetes* 59(10):2530–2539. <http://dx.doi.org/10.2337/db10-0190>.
- [27] Gutierrez, G.D., Bender, A.S., Cirulli, V., Mastracci, T.L., Kelly, S.M., Tsigos, A., et al., 2017. Pancreatic beta cell identity requires continual repression of non-beta cell programs. *The Journal of Clinical Investigation* 127(1):244–259. <http://dx.doi.org/10.1172/Jci88017>.
- [28] Swisa, A., Avrahami, D., Eden, N., Zhang, J., Feleke, E., Dahan, T., et al., 2017. PAX6 maintains  $\beta$  cell identity by repressing genes of alternative islet cell types. *The Journal of Clinical Investigation* 127(1):230–243. <http://dx.doi.org/10.1172/Jci88015>.
- [29] Röder, P.V., Wu, B., Liu, Y., Han, W., 2016. Pancreatic regulation of glucose homeostasis. *Experimental & Molecular Medicine* 48(3):e219. <http://dx.doi.org/10.1038/emm.2016.6>.
- [30] van der Meulen, T., Donaldson, C.J., Cáceres, E., Hunter, A.E., Cowing-Zitron, C., Pound, L.D., et al., 2015. Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. *Nature Medicine* 21(7):769–776. <http://dx.doi.org/10.1038/nm.3872>.
- [31] Gettys, T.W., Tanaka, I., Taylor, I.L., 1992. Modulation of pancreatic exocrine function in rodents by treatment with pancreatic polypeptide. *Pancreas* 7(6):705–711.
- [32] Batterham, R.L., Le Roux, C.W., Cohen, M.A., Park, A.J., Ellis, S.M., Patterson, M., et al., 2003. Pancreatic polypeptide reduces appetite and food intake in humans. *The Journal of Clinical Endocrinology & Metabolism* 88(8):3989–3992. <http://dx.doi.org/10.1210/jc.2003-030630>.
- [33] Katsuura, G., Asakawa, A., Inui, A., 2002. Roles of pancreatic polypeptide in regulation of food intake. *Peptides* 23(2):323–329.
- [34] Steiner, D.J., Kim, A., Miller, K., Hara, M., 2010. Pancreatic islet plasticity: interspecies comparison of islet architecture and composition. *Islets*. <http://dx.doi.org/10.4161/isl.2.3.11815>.
- [35] Bosco, D., Armanet, M., Morel, P., Niclauss, N., Sgroi, A., Müller, Y.D., et al., 2010. Unique arrangement of  $\alpha$ - and  $\beta$ -cells in human islets of Langerhans TL - 59. *Diabetes* 59(5):1202–1210. <http://dx.doi.org/10.2337/db09-1177>.
- [36] Kharouta, M., Miller, K., Kim, A., Wojcik, P., Kilimnik, G., Dey, A., et al., 2009. No mantle formation in rodent islets-The prototype of islet revisited. *Diabetes Research and Clinical Practice* 85(3):252–257. <http://dx.doi.org/10.1016/j.diabres.2009.06.021>.
- [37] Bonner-Weir, S., Sullivan, B.A., Weir, G.C., 2015. Human islet morphology revisited: human and rodent islets are not so different after all. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society* 63(8):604–612. <http://dx.doi.org/10.1369/0022155415570969>.
- [38] Brissova, M., Fowler, M.J., Nicholson, W.E., Chu, A., Hirshberg, B., Harlan, D.M., et al., 2005. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society* 53(9):1087–1097. <http://dx.doi.org/10.1369/jhc.5C6684.2005>.

- [39] Cabrera, O., Berman, D.M., Kenyon, N.S., Ricordi, C., Berggren, P.-O., Caicedo, A., 2006. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proceedings of the National Academy of Sciences of the United States of America* 103(7):2334–2339. <http://dx.doi.org/10.1073/pnas.0510790103>.
- [40] Blodgett, D.M., Nowosielska, A., Afik, S., Pechhold, S., Cura, A.J., Kennedy, N.J., et al., 2015. Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. *Diabetes* 64(9):3172–3181. <http://dx.doi.org/10.2337/db15-0039>.
- [41] Segerstolpe, A., Palasantza, A., Eliasson, P., Andersson, E.M., Andreasson, A.C., Sun, X.Y., et al., 2016. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metabolism* 24(4):593–607. <http://dx.doi.org/10.1016/j.cmet.2016.08.020>.
- [42] Xin, Y.R., Kim, J., Okamoto, H., Ni, M., Wei, Y., Adler, C., et al., 2016. RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metabolism* 24(4):608–615. <http://dx.doi.org/10.1016/j.cmet.2016.08.018>.
- [43] Benner, C., van der Meulen, T., Caceres, E., Tigyi, K., Donaldson, C.J., Huisling, M.O., et al., 2014. The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. *BMC Genomics* 15(1):620. <http://dx.doi.org/10.1186/1471-2164-15-620>.
- [44] Li, J., Klughammer, J., Farlik, M., Penz, T., Spittler, A., Barbieux, C., et al., 2015. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Reports* 17(2):1–10. <http://dx.doi.org/10.15252/embr.201540946>.
- [45] Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., et al., 2016. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Systems* 3(4):346–360. <http://dx.doi.org/10.1016/j.cels.2016.08.011>.
- [46] Qiu, W.-L., Zhang, Y.-W., Feng, Y., Li, L.-C., Yang, L., Xu, C.-R., 2017. Deciphering pancreatic islet beta cell and alpha cell maturation pathways and characteristic features at the single-cell level. *Cell Metabolism* 25(5):1194–1205. <http://dx.doi.org/10.1016/j.cmet.2017.04.003> e4.
- [47] Ackermann, A.M., Wang, Z.P., Schug, J., Naji, A., Kaestner, K.H., 2016. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Molecular Metabolism* 5(3):233–244. <http://dx.doi.org/10.1016/j.molmet.2016.01.002>.
- [48] Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., et al., 2016. A single-cell transcriptome atlas of the human pancreas. *Cell Systems* 3(4):385–394. <http://dx.doi.org/10.1016/j.cels.2016.09.002>.
- [49] Mathieu, C., 2015. Vitamin D and diabetes: where do we stand? *Diabetes Research and Clinical Practice*, 201–209. <http://dx.doi.org/10.1016/j.diabres.2015.01.036>.
- [50] Wang, Y., Wang, O., Li, W., Ma, L., Ping, F., Chen, L., et al., 2015. Variants in vitamin D binding protein gene are associated with gestational diabetes mellitus. *Medicine* 94(40):e1693. <http://dx.doi.org/10.1097/MD.0000000000001693>.
- [51] Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., et al., 2016. Single cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Research*. <http://dx.doi.org/10.1101/gr.212720.116>.
- [52] Arda, H.E., Li, L., Tsai, J., Torre, E.A., Rosli, Y., Peiris, H., et al., 2016. Age-dependent pancreatic gene regulation reveals mechanisms governing human?? cell function. *Cell Metabolism* 23(5):909–920. <http://dx.doi.org/10.1016/j.cmet.2016.04.002>.
- [53] DiGrucio, M.R., Mawla, A.M., Donaldson, C.J., Noguchi, G.M., Vaughan, J., Cowing-Zitron, C., et al., 2016. Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. *Molecular Metabolism* 5(7):449–458. <http://dx.doi.org/10.1016/j.molmet.2016.04.007>.
- [54] Müller, T.D., Nogueiras, R., Andermann, M.L., Andrews, Z.B., Anker, S.D., Argente, J., et al., 2015. Ghrelin. *Molecular Metabolism*, 437–460. <http://dx.doi.org/10.1016/j.molmet.2015.03.005>.
- [55] Friedman, J., 2016. The long road to leptin. *The Journal of Clinical Investigation* 126(12):4727–4734. <http://dx.doi.org/10.1172/JCI91578>.
- [56] Kulkarni, R.N., Wang, Z.L., Wang, R.M., Hurley, J.D., Smith, D.M., Ghatei, M.A., et al., 1997. Leptin rapidly suppresses insulin release from insulinoma cells, rat and human islets and, in vivo, in mice. *The Journal of Clinical Investigation* 100(11):2729–2736. <http://dx.doi.org/10.1172/JCI119818>.
- [57] Morioka, T., Asilmaz, E., Hu, J., Dishinger, J.F., Kurpad, A.J., Elias, C.F., et al., 2007. Disruption of leptin receptor expression in the pancreas directly affects  $\beta$  cell growth and function in mice. *The Journal of Clinical Investigation* 117(10):2860–2868. <http://dx.doi.org/10.1172/JCI30910.2860>.
- [58] Soedling, H., Hodson, D.J., Adrianssens, A.E., Gribble, F.M., Reimann, F., Trapp, S., et al., 2015. Limited impact on glucose homeostasis of leptin receptor deletion from insulin- or proglucagon-expressing cells. *Molecular Metabolism* 4(9):619–630. <http://dx.doi.org/10.1016/j.molmet.2015.06.007>.
- [59] Keenan, H.A., Sun, J.K., Levine, J., Doria, A., Aiello, L.P., Eisenbarth, G., et al., 2010. Residual insulin production and pancreatic beta-cell turnover after 50 years of diabetes: Joslin medalist study. *Diabetes* 59(11):2846–2853. <http://dx.doi.org/10.2337/db10-0676>.
- [60] Oram, R.A., Jones, A.G., Besser, R.E.J., Knight, B.A., Shields, B.M., Brown, R.J., et al., 2014. The majority of patients with long-duration type 1 diabetes are insulin microsecretors and have functioning beta cells. *Diabetologia* 57(1):187–191. <http://dx.doi.org/10.1007/s00125-013-3067-x>.
- [61] Rui, J., Deng, S., Arazi, A., Perdigo, A.L., Liu, Z., Herold, K.C., 2017.  $\beta$  cells that resist immunological attack develop during progression of autoimmune diabetes in NOD mice. *Cell Metabolism* 25(3):727–738. <http://dx.doi.org/10.1016/j.cmet.2017.01.005>.
- [62] Wang, Z.Y., York, N.W., Nichols, C.G., Remedi, M.S., 2014. Pancreatic beta cell dedifferentiation in diabetes and redifferentiation following insulin therapy. *Cell Metabolism* 19(5):872–882. <http://dx.doi.org/10.1016/j.cmet.2014.03.010>.
- [63] Wang, Y.J., Schug, J., Won, K.J., Liu, C.Y., Naji, A., Avrahami, D., et al., 2016. Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* 65(10):3028–3038. <http://dx.doi.org/10.2337/db16-0405>.
- [64] Salomon, D., Meda, P., 1986. Heterogeneity and contact-dependent regulation of hormone secretion by individual B cells. *Experimental Cell Research* 162(2):507–520.
- [65] Stefan, Y., Meda, P., Neufeld, M., Orci, L., 1987. Stimulation of insulin secretion reveals heterogeneity of pancreatic B cells in vivo. *The Journal of Clinical Investigation* 80(1):175–183. <http://dx.doi.org/10.1172/JCI113045>.
- [66] Heimberg, H., De Vos, A., Vandercammen, A., Van Schaffingen, E., Pipeleers, D., Schuit, F., 1993. Heterogeneity in glucose sensitivity among pancreatic beta-cells is correlated to differences in glucose phosphorylation rather than glucose transport. *The EMBO Journal* 12(7):2873–2879.
- [67] Katsuta, H., Aguayo-Mazzucato, C., Katsuta, R., Akashi, T., Hollister-Lock, J., Sharma, A.J., et al., 2012. Subpopulations of GFP-marked mouse pancreatic beta-cells differ in size, granularity, and insulin secretion. *Endocrinology* 153(11):5180–5187. <http://dx.doi.org/10.1210/en.2012-1257>.
- [68] Dorrell, C., Schug, J., Canaday, P.S., Russ, H.A., Tarlow, B.D., Grompe, M.T., et al., 2016. Human islets contain four distinct subtypes of beta cells. *Nature Communications* 7:11756. <http://dx.doi.org/10.1038/ncomms11756>.
- [69] Bader, E., Migliorini, A., Gegg, M., Moruzzi, N., Gerdes, J., Roscioni, S.S., et al., 2016. Identification of proliferative and mature beta-cells in the islets of Langerhans. *Nature* 535(7612). <http://dx.doi.org/10.1038/nature18624>, 430–+.
- [70] Katsuta, H., Akashi, T., Katsuta, R., Nagaya, M., Kim, D., Arinobu, Y., et al., 2010. Single pancreatic beta cells co-express multiple islet hormone genes in mice. *Diabetologia* 53(1):128–138. <http://dx.doi.org/10.1007/s00125-009-1570-x>.

- [71] Szabat, M., Luciani, D.S., Piret, J.M., Johnson, J.D., 2009. Maturation of adult beta-cells revealed using a Pdx1/insulin dual-reporter Lentivirus. *Endocrinology* 150(4):1627–1635. <http://dx.doi.org/10.1210/en.2008-1224>.
- [72] Smukler, S.R., Arntfield, M.E., Razavi, R., Bikopoulos, G., Karpowicz, P., Seaberg, R., et al., 2011. The adult mouse and human pancreas contain rare multipotent stem cells that express insulin. *Cell Stem Cell* 8(3):281–293. <http://dx.doi.org/10.1016/j.stem.2011.01.015>.
- [73] Hermann, M., Pirkebner, D., Draxl, A., Berger, P., Untergasser, G., Margreiter, R., et al., 2007. Dickkopf-3 is expressed in a subset of adult human pancreatic beta cells. *Histochemistry and Cell Biology* 127(5):513–521. <http://dx.doi.org/10.1007/s00418-007-0278-6>.
- [74] Wang, Y.J., Golson, M.L., Schug, J., Traum, D., Liu, C.Y., Vivek, K., et al., 2016. Single-cell mass cytometry analysis of the human endocrine pancreas. *Cell Metabolism* 24(4):616–626. <http://dx.doi.org/10.1016/j.cmet.2016.09.007>.
- [75] Teta, M., Long, S.Y., Wartschow, L.M., Rankin, M.M., Kushner, J.A., 2005. Very slow turnover of beta-cells in aged adult mice. *Diabetes* 54(9):2557–2567. <http://dx.doi.org/10.2337/diabetes.54.9.2557>.
- [76] Zeng, C., Mulas, F., Sui, Y., Guan, T., Miller, N., Tan, Y., et al., 2017. Pseudotemporal ordering of single cells reveals metabolic control of postnatal  $\beta$  cell proliferation. *Cell Metabolism* 25(5):1160–1175. <http://dx.doi.org/10.1016/j.cmet.2017.04.014>, e11.
- [77] Parsons, J.A., Brelje, T.C., Sorenson, R.L., 1992. Adaptation of islets of Langerhans to pregnancy: increased islet cell proliferation and insulin secretion correlates with the onset of placental lactogen secretion. *Endocrinology* 130(3):1459–1466. <http://dx.doi.org/10.1210/endo.130.3.1537300>.
- [78] Bonner-Weir, S., Toschi, E., Inada, A., Reitz, P., Fonseca, S.Y., Aye, T., et al., 2004. The pancreatic ductal epithelium serves as a potential pool of progenitor cells. *Pediatric Diabetes* (5 Suppl 2):16–22. <http://dx.doi.org/10.1111/j.1399-543X.2004.00075.x>.
- [79] Bonner-Weir, S., Guo, L., Li, W.-C., Ouziel-Yahalom, L., Weir, G.C., Sharma, A., et al., 2012. Islet neogenesis: a possible pathway for beta-cell replenishment. *The Review of Diabetic Studies* 9(4):407–416. <http://dx.doi.org/10.1900/RDS.2012.9.407>.
- [80] Dor, Y., Brown, J., Martinez, O.I., Melton, D.A., 2004. Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* 429(6987):41–46. <http://dx.doi.org/10.1038/nature02520>.
- [81] Seaberg, R.M., Smukler, S.R., Kieffer, T.J., Enikolopov, G., Asghar, Z., Wheeler, M.B., et al., 2004. Clonal identification of multipotent precursors from adult mouse pancreas that generate neural and pancreatic lineages. *Nature Biotechnology* 22(9):1115–1124. <http://dx.doi.org/10.1038/nbt1004>.
- [82] Razavi, R., Najafabadi, H.S., Abdullah, S., Smukler, S., Arntfield, M., van der Kooy, D., 2015. Diabetes enhances the proliferation of adult pancreatic multipotent progenitor cells and biases their differentiation to more beta-cell production. *Diabetes* 64(4):1311–1323. <http://dx.doi.org/10.2337/db14-0070>.
- [83] Bernal-Mizrachi, E., Kulkarni, R.N., Scott, D.K., Mauvais-Jarvis, F., Stewart, A.F., Garcia-Ocana, A., 2014. Human  $\beta$ -cell proliferation and intracellular signaling Part 2: still driving in the dark without a road map. *Diabetes* 63(3):819–831. <http://dx.doi.org/10.2337/db13-1146>.
- [84] Li, C.W., Heidt, D.G., Dalerba, P., Burant, C.F., Zhang, L.J., Adsay, V., et al., 2007. Identification of pancreatic cancer stem cells. *Cancer Research* 67(3):1030–1037. <http://dx.doi.org/10.1158/0008-5472.Can-06-2030>.
- [85] Sugiyama, T., Rodriguez, R.T., McLean, G.W., Kim, S.K., 2007. Conserved markers of fetal pancreatic epithelium permit prospective isolation of islet progenitor cells by FACS. *Proceedings of the National Academy of Sciences of the United States of America* 104(1):175–180. <http://dx.doi.org/10.1073/pnas.0609490104>.
- [86] Wang, P., Alvarez-Perez, J.C., Felsenfeld, D.P., Liu, H.T., Sivendran, S., Bender, A., et al., 2015. A high-throughput chemical screen reveals that harmine-mediated inhibition of DYRK1A increases human pancreatic beta cell replication. *Nature Medicine* 21(4):383. <http://dx.doi.org/10.1038/nm.3820>.
- [87] Dirice, E., Walpita, D., Vetere, A., Meier, B.C., Kahraman, S., Hu, J., et al., 2016. Inhibition of DYRK1A stimulates human  $\beta$ -cell proliferation. *Diabetes* 65(6):1660–1671. <http://dx.doi.org/10.2337/db15-1127>.
- [88] Roschger, C., Cabrele, C., 2017. The Id-protein family in developmental and cancer-associated pathways. *Cell Communication and Signaling* 15(1):7. <http://dx.doi.org/10.1186/s12964-016-0161-y>.
- [89] Kleeff, J., Maruyama, H., Ishiwata, T., Sawhney, H., Friess, H., Buchler, M.W., et al., 1999. Bone morphogenetic protein 2 exerts diverse effects on cell growth in vitro and is expressed in human pancreatic cancer in vivo. *Gastroenterology* 116(5):1202–1216. [http://dx.doi.org/10.1016/S0016-5085\(99\)70024-7](http://dx.doi.org/10.1016/S0016-5085(99)70024-7).
- [90] Hua, H., Zhang, Y.Q., Dabernat, S., Kritzik, M., Dietz, D., Sterling, L., et al., 2006. BMP4 regulates pancreatic progenitor cell expansion through Id2. *The Journal of Biological Chemistry* 281(19):13574–13580. <http://dx.doi.org/10.1074/jbc.M600526200>.
- [91] Sharma, R.B., O'Donnell, A.C., Stamateris, R.E., Ha, B., McCloskey, K.M., Reynolds, P.R., et al., 2015. Insulin demand regulates beta cell number via the unfolded protein response. *The Journal of Clinical Investigation* 125(10):3831–3846. <http://dx.doi.org/10.1172/Jci79264>.
- [92] Weir, G.C., Aguayo-Mazzucato, C., Bonner-Weir, S., 2013.  $\beta$ -cell dedifferentiation in diabetes is important, but what is it? *Islets* 5(5):233–237. <http://dx.doi.org/10.4161/isl.27494>.
- [93] Laybutt, D.R., Sharma, A., Sgroi, D.C., Gaudet, J., Bonner-Weir, S., Weir, G.C., 2002. Genetic regulation of metabolic pathways in beta-cells disrupted by hyperglycemia. *The Journal of Biological Chemistry* 277(13):10912–10921. <http://dx.doi.org/10.1074/jbc.M111751200>.
- [94] Gegg, M., Bottcher, A., Burtscher, I., Hasenoeder, S., Van Campenhout, C., Aichler, M., et al., 2014. Flattop regulates basal body docking and positioning in mono- and multiciliated cells. *Elife* 3:e03842. <http://dx.doi.org/10.7554/eLife.03842>.
- [95] Yang, Q., Graham, T.E., Mody, N., Preitner, F., Peroni, O.D., Zabolotny, J.M., et al., 2005. Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature* 436(7049):356–362. <http://dx.doi.org/10.1038/nature03711>.
- [96] Moran, B.M., Abdel-Wahab, Y.H.A., Flatt, P.R., McKillop, A.M., 2014. Evaluation of the insulin-releasing and glucose-lowering effects of GPR120 activation in pancreatic beta-cells. *Diabetes Obesity & Metabolism* 16(11):1128–1139. <http://dx.doi.org/10.1111/dom.12330>.
- [97] Guo, S., Dai, C., Guo, M., Taylor, B., Harmon, J.S., Sander, M., et al., 2013. Inactivation of specific  $\beta$  cell transcription factors in type 2 diabetes. *The Journal of Clinical Investigation* 123(8):3305–3316. <http://dx.doi.org/10.1172/JCI65390>.
- [98] Pullen, T.J., Khan, A.M., Barton, G., Butcher, S.A., Sun, G., Rutter, G.A., 2010. Identification of genes selectively disallowed in the pancreatic islet. *Islets* 2(2):89–95. <http://dx.doi.org/10.4161/isl.2.2.11025>.
- [99] Thorrez, L., Laudadio, I., Van Deun, K., Quintens, R., Hendrickx, N., Granvik, M., et al., 2011. Tissue-specific disallowance of housekeeping genes: the other face of cell differentiation. *Genome Research* 21(1):95–105. <http://dx.doi.org/10.1101/gr.109173.110>.
- [100] Ediger, B.N., Lim, H.-W., Juliana, C., Groff, D.N., Williams, L.T., Dominguez, G., et al., 2016. LIM domain-binding 1 maintains the terminally differentiated state of pancreatic  $\beta$  cells. *The Journal of Clinical Investigation* 127(1):215–229. <http://dx.doi.org/10.1172/JCI88016>.
- [101] Brereton, M.F., Iberl, M., Shimomura, K., Zhang, Q., Adriaenssens, A.E., Proks, P., et al., 2014. Reversible changes in pancreatic islet structure and function produced by elevated blood glucose. *Nature Communications* 5:4639. <http://dx.doi.org/10.1038/ncomms5639>.
- [102] Marroqui, L., Masini, M., Merino, B., Grieco, F.A., Millard, I., Dubois, C., et al., 2015. Pancreatic  $\alpha$  cells are resistant to metabolic stress-induced apoptosis



- in type 2 diabetes. *EBioMedicine* 2(5):378–385. <http://dx.doi.org/10.1016/j.ebiom.2015.03.012>.
- [103] Collobat, P., Xu, X.B., Ravassard, P., Sosa-Pineda, B., Dussaud, S., Billestrup, N., et al., 2009. The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. *Cell* 138(3):449–462. <http://dx.doi.org/10.1016/j.cell.2009.05.035>.
- [104] Thorel, F., Nepote, V., Avril, I., Kohno, K., Desgraz, R., Chera, S., et al., 2010. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* 464(7292):1149–1154. <http://dx.doi.org/10.1038/nature08894>.
- [105] Bramswig, N.C., Everett, L.J., Schug, J., Dorrell, C., Liu, C.Y., Luo, Y.P., et al., 2013. Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. *The Journal of Clinical Investigation* 123(3):1275–1284. <http://dx.doi.org/10.1172/Jci66514>.
- [106] Ben-Othman, N., Vieira, A., Courtney, M., Record, F., Gjernes, E., Avolio, F., et al., 2017. Long-term GABA administration induces alpha cell-mediated beta-like cell neogenesis. *Cell* 168(1–2):73–85. <http://dx.doi.org/10.1016/j.cell.2016.11.002>, e11.
- [107] Li, J., Casteels, T., Frogne, T., Ingvorsen, C., Honoré, C., Courtney, M., et al., 2017. Artemisinins target GABA A receptor signaling and impair  $\alpha$  cell identity. *Cell* 168(1–2):86–100. <http://dx.doi.org/10.1016/j.cell.2016.11.010>, e15.
- [108] Ye, L.H., Robertson, M.A., Hesselton, D., Stainier, D.Y.R., Anderson, R.M., 2015. Glucagon is essential for alpha cell transdifferentiation and beta cell neogenesis. *Development* 142(8):1407–1417. <http://dx.doi.org/10.1242/dev.117911>.
- [109] van der Meulen, T., Mawla, A.M., DiGruccio, M.R., Adams, M.W., Nies, V., Dölleman, S., et al., 2017. Virgin beta cells persist throughout life at a neogenic niche within pancreatic islets. *Cell Metabolism* 25(4):911–926. <http://dx.doi.org/10.1016/j.cmet.2017.03.017>, e6.
- [110] Chakravarthy, H., Gu, X., Enge, M., Dai, X., Wang, Y., Damond, N., et al., 2017. Converting adult pancreatic islet  $\alpha$  cells into  $\beta$  cells by targeting both Dnm1 and Arx. *Cell Metabolism* 25(3):622–634. <http://dx.doi.org/10.1016/j.cmet.2017.01.009>.
- [111] Shen, W.J., Taylor, B., Jin, Q.H., Nguyen-Tran, V., Meeusen, S., Zhang, Y.Q., et al., 2015. Inhibition of DYRK1A and GSK3B induces human beta-cell proliferation. *Nature Communications* 6:8372. <http://dx.doi.org/10.1038/ncomms9372>.
- [112] Mill, P., Mo, R., Hu, M.C., Dagnino, L., Rosenblum, N.D., Hui, C., 2005. Shh controls epithelial proliferation via independent pathways that converge on N-Myc. *Developmental Cell* 9(2):293–303. <http://dx.doi.org/10.1016/j.devcel.2005.05.009>.
- [113] Kise, Y., Morinaka, A., Teglund, S., Miki, H., 2009. Sufu recruits GSK3 $\beta$  for efficient processing of Gli3. *Biochemical and Biophysical Research Communications* 387(3):569–574. <http://dx.doi.org/10.1016/j.bbrc.2009.07.087>.
- [114] Schneider, P., Miguel Bayo-Fina, J., Singh, R., Kumar Dhanyamraju, P., Holz, P., Baier, A., et al., 2015. Identification of a novel actin-dependent signal transducing module allows for the targeted degradation of GLI1. *Nature Communications* 6:8023. <http://dx.doi.org/10.1038/ncomms9023>.
- [115] Wollny, D., Zhao, S., Everlien, I., Lun, X.K., Brunken, J., Brune, D., et al., 2016. Single-cell analysis uncovers clonal acinar cell heterogeneity in the adult pancreas. *Developmental Cell* 39(3):289–301. <http://dx.doi.org/10.1016/j.devcel.2016.10.002>.
- [116] Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., Melton, D.A., 2008. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 455(7213). <http://dx.doi.org/10.1038/nature07314>, 627–U30.
- [117] Li, W.D., Nakanishi, M., Zumsteg, A., Shear, M., Wright, C., Melton, D.A., et al., 2014. In vivo reprogramming of pancreatic acinar cells to three islet endocrine subtypes. *Elife* 3:e01846. <http://dx.doi.org/10.7554/eLife.01846>.
- [118] Baeyens, L., De Breuck, S., Lardon, J., Mfopou, J.K., Rooman, I., Bouwens, L., 2005. In vitro generation of insulin-producing beta cells from adult exocrine pancreatic cells. *Diabetologia* 48(1):49–57. <http://dx.doi.org/10.1007/s00125-004-1606-1>.
- [119] Minami, K., Okuno, M., Miyawaki, K., Okumachi, A., Ishizaki, K., Oyama, K., et al., 2005. Lineage tracing and characterization of insulin-secreting cells generated from adult pancreatic acinar cells. *Proceedings of the National Academy of Sciences of the United States of America* 102(42):15116–15121. <http://dx.doi.org/10.1073/pnas.0507567102>.
- [120] Baeyens, L., Lemper, M., Leuckx, G., De Groef, S., Bonfanti, P., Stange, G., et al., 2014. Transient cytokine treatment induces acinar cell reprogramming and regenerates functional beta cell mass in diabetic mice. *Nature Biotechnology* 32(1). <http://dx.doi.org/10.1038/nbt.2747>, 76–+.
- [121] Beattie, G.M., Otonkoski, T., Lopez, A.D., Hayek, A., 1997. Functional B-cell mass after transplantation of human fetal pancreatic cells - differentiation or proliferation? *Diabetes* 46(2):244–248. <http://dx.doi.org/10.2337/diabetes.46.2.244>.
- [122] Hao, E., Tyrberg, B., Itkin-Ansari, P., Lakey, J.R.T., Geron, I., Monosov, E.Z., et al., 2006. Beta-cell differentiation from nonendocrine epithelial cells of the adult human pancreas. *Nature Medicine* 12(3):310–316. <http://dx.doi.org/10.1038/nm1367>.
- [123] Zhang, W.J., Xu, S.Q., Cai, H.Q., Men, X.L., Wang, Z., Lin, H., et al., 2013. Evaluation of islets derived from human fetal pancreatic progenitor cells in diabetes treatment. *Stem Cell Research & Therapy* 4:141. <http://dx.doi.org/10.1186/scrt352>.
- [124] Lemper, M., Leuckx, G., Heremans, Y., German, M.S., Heimberg, H., Bouwens, L., et al., 2015. Reprogramming of human pancreatic exocrine cells to beta-like cells. *Cell Death and Differentiation* 22(7):1117–1130. <http://dx.doi.org/10.1038/cdd.2014.193>.
- [125] de Back, W., Zimm, R., Brusch, L., 2013. Transdifferentiation of pancreatic cells by loss of contact-mediated signaling. *BMC Systems Biology* 7:77. <http://dx.doi.org/10.1186/1752-0509-7-77>.
- [126] Cassimeris, L., 2002. The oncoprotein 18/stathmin family of microtubule destabilizers. *Current Opinion in Cell Biology* 14(1):18–24. [http://dx.doi.org/10.1016/S0955-0674\(01\)00289-7](http://dx.doi.org/10.1016/S0955-0674(01)00289-7).
- [127] Furuyama, K., Kawaguchi, Y., Akiyama, H., Horiguchi, M., Kodama, S., Kuhara, T., et al., 2011. Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nature Genetics* 43(1). <http://dx.doi.org/10.1038/ng.722>, 34–U52.
- [128] Parikh, A., Stephan, A.-F., Tzanakakis, E.S., 2012. Regenerating proteins and their expression, regulation, and signaling. *BioMolecular Concepts* 3(1):57–70. <http://dx.doi.org/10.1515/bmc.2011.055>.
- [129] Xu, Q.Q., Fu, R., Yin, G.X., Liu, X.L., Liu, Y., Xiang, M., 2016. Microarray-based gene expression profiling reveals genes and pathways involved in the oncogenic function of REG3A on pancreatic cancer cells. *Gene* 578(2):263–273. <http://dx.doi.org/10.1016/j.gene.2015.12.039>.
- [130] Clayton, H.W., Osipovich, A.B., Stancill, J.S., Schneider, J.D., Vianna, P.G., Shanks, C.M., et al., 2016. Pancreatic inflammation redirects acinar to beta cell reprogramming. *Cell Reports* 17(8):2028–2041. <http://dx.doi.org/10.1016/j.celrep.2016.10.068>.
- [131] Cavelti-Weder, C., Li, W.D., Zumsteg, A., Stemann-Andersen, M., Zhang, Y.M., Yamada, T., et al., 2016. Hyperglycaemia attenuates in vivo reprogramming of pancreatic exocrine cells to beta cells in mice. *Diabetologia* 59(3):522–532. <http://dx.doi.org/10.1007/s00125-015-3838-7>.
- [132] Baeyens, L., Bonne, S., Bos, T., Rooman, I., Peleman, C., Lahoutte, T., et al., 2009. Notch signaling as gatekeeper of rat acinar-to-beta-cell conversion in vitro. *Gastroenterology* 136(5):1750–1760. <http://dx.doi.org/10.1053/j.gastro.2009.01.047>.
- [133] Dios De, I., 2010. Inflammatory role of the acinar cells during acute pancreatitis. *World Journal of Gastrointestinal Pharmacology and Therapeutics* 1(1):15. <http://dx.doi.org/10.4292/wjgpt.v1.i1.15>.
- [134] Roy, N., Hebrok, M., 2015. Regulation of cellular identity in cancer. *Developmental Cell* 35(6):674–684. <http://dx.doi.org/10.1016/j.devcel.2015.12.001>.
- [135] Wei, D., Wang, L., Yan, Y., Jia, Z., Gagea, M., Li, Z., et al., 2016. KLF4 is essential for induction of cellular identity change and acinar-to-ductal

- reprogramming during early pancreatic carcinogenesis. *Cancer Cell* 29(3): 324–338. <http://dx.doi.org/10.1016/j.ccell.2016.02.005>.
- [136] Pallagi, P., Hegyi, P., Rakonczay, Z., 2015. The physiology and Pathophysiology of pancreatic ductal secretion. *Pancreas* 44(8):1211–1233. <http://dx.doi.org/10.1097/MPA.0000000000000421>.
- [137] Al-Hasani, K., Pfeifer, A., Courtney, M., Ben-Othman, N., Gjernes, E., Vieira, A., et al., 2013. Adult duct-lining cells can reprogram into  $\beta$ -like cells able to counter repeated cycles of toxin-induced diabetes. *Developmental Cell* 26(1):86–100. <http://dx.doi.org/10.1016/j.devcel.2013.05.018>.
- [138] Xu, X., D'Hoker, J., Stangé, G., Bonn e, S., De Leu, N., Xiao, X., et al., 2008.  $\beta$  cells can Be generated from endogenous progenitors in injured adult mouse pancreas. *Cell* 132(2):197–207. <http://dx.doi.org/10.1016/j.cell.2007.12.015>.
- [139] Yamada, T., Cavelti-Weder, C., Caballero, F., Lysy, P.A., Guo, L., Sharma, A., et al., 2015. Reprogramming mouse cells with a pancreatic duct phenotype to insulin-producing  $\beta$ -like cells. *Endocrinology* 156(6):2029–2038. <http://dx.doi.org/10.1210/en.2014-1987>.
- [140] Grun, D., Muraro, M.J., Boisset, J.C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., et al., 2016. De Novo Prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* 19(2):266–277. <http://dx.doi.org/10.1016/j.stem.2016.05.010>.
- [141] Wagner, A., Regev, A., Yosef, N., 2016. Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology* 34(11):1145–1160. <http://dx.doi.org/10.1038/nbt.3711>.
- [142] Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., et al., 2016. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167(7):1853–1866. <http://dx.doi.org/10.1016/j.cell.2016.11.038>, e17.
- [143] Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nu ez, J.K., Chen, Y., et al., 2016. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167(7):1867–1882. <http://dx.doi.org/10.1016/j.cell.2016.11.048>, e21.
- [144] Jaitin, D.A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., et al., 2016. Dissecting immune circuits by linking CRISPR-Pooled screens with single-cell RNA-seq. *Cell* 167(7):1883–1896. <http://dx.doi.org/10.1016/j.cell.2016.11.039>, e15.
- [145] Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., et al., 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* 14(3):297–301. <http://dx.doi.org/10.1038/nmeth.4177>.
- [146] Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., et al., 2011. Single-molecule transcript counting of stem-cell markers in the mouse intestine. *Nature Cell Biology* 14(1):106–114. <http://dx.doi.org/10.1038/ncb2384>.
- [147] Bahar Halpern, K., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., et al., 2017. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 542(7641):352–356. <http://dx.doi.org/10.1038/nature21065>.
- [148] Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., et al., 2014. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods* 11(4):417–422. <http://dx.doi.org/10.1038/nmeth.2869>.
- [149] Rezanian, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., et al., 2014. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nature Biotechnology* 32(11):1121–1133. <http://dx.doi.org/10.1038/nbt.3033>. <http://www.nature.com/nbt/journal/v32/n11/abs/nbt.3033.html#supplementary-information>.