Wissenschaftszentrum Weihenstephan Fakultät für Ernährung, Landnutzung und Umwelt

Molecular and cellular mechanisms underlying the GRB14/COBLL1 diabetes risk locus

Viktoria Ocvirk (geb. Glunk)

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzende/-r: Prof. Dr. med. H. Witt

Prüfende/-r der Dissertation:

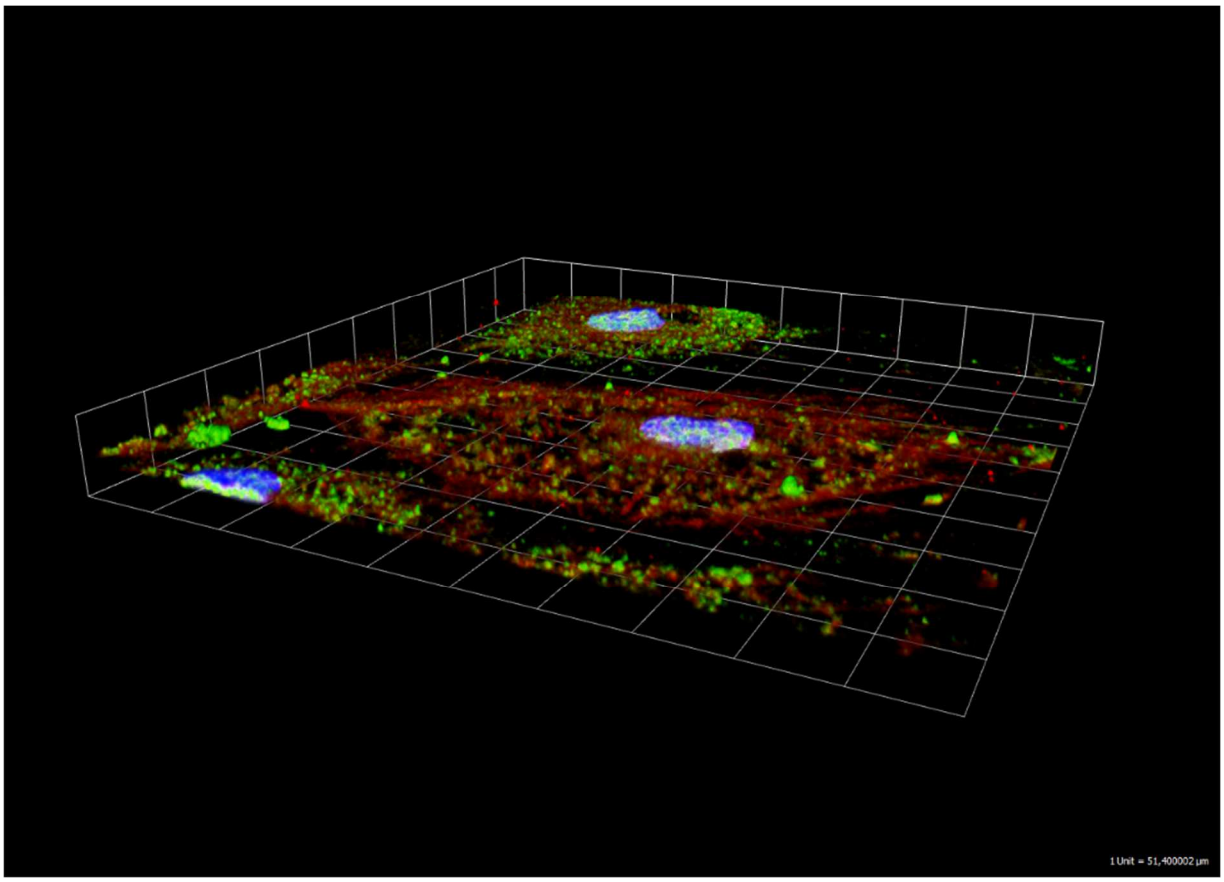
1. Prof. Dr. Johann J. Hauner

2. Prof. Dr. Michael W. Pfaffl

Die Dissertation wurde am 09.12.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung um Umwelt am 20.05.2020 angenommen.

„Zweifle nie daran, dass eine kleine Gruppe engagierter Menschen die Welt verändern kann - tatsächlich ist dies die einzige Art und Weise, in der die Welt jemals verändert wurde.“

Margaret Mead



3D-image of 120-fold magnified in vitro differentiated human adipocytes, where nuclei (blue), filamentous actin (red) and COBLL1 (green) were stained. Image was produced using the Olympus IX81 inverted confocal microscope and the Volocity 3D Image Analysis software version 5.4.1.

Table of contents

List of figures

List of tables

List of abbreviations

General abbreviations

3T3-L1	Mouse 3T3-L1 preadipocyte cell line
A-431	Human squamous carcinoma cell line
BPS/PIR	Between PH and SH2 domain
CucE	Cucurbitacin E
F-actin	Filamentous actin
G-actin	Globular monomeric actin
GO-term	Gene ontology pathway-term
GWAS	Genome-wide association studies
hAC	Primary human adipocytes
hASC	Primary human adipose-derived stromal cells
HeLa	Human cervical cancer cell line
Huh7	Human liver cell line
IBMX	3-Isobutyl-1-methylxanthine
INS1	Rat INS1 beta-cells
KKRAP motif	Repeated lysine, arginine, and proline-rich region
LatA	Latrunculin A
LD	Linkage disequilibrium
MOI	multiplicity of infection
PH	pleckstrin homology domain
PMCA	Phylogenetic module complexity analysis
RA	Ras-associated domain
SAT	Subcutaneous adipose tissue
SGBS	Simpson–Golabi–Behmel syndrome cell line
SH2	Src homology 2 domain
SVF	Stromal vascular fraction of adipose tissue
T2D	Type 2 diabetes
U-2 OS	Human osteosarcoma cell line
WH2 domain	WASP (Wiskott–Aldrich syndrome protein)-homology 2 domain
WHR	Waist-to-hip-ratio

Gene abbreviations

ACTN1	Actinin, alpha 1
ACTN4	Actinin, alpha 4
ADAM12	ADAM metallopeptidase domain 12
ALPP	Alkaline phosphatase, placental
ALPPL2	Alkaline phosphatase, placental-like 2
Arp2/3	Actin related protein 2/3 complex subunit 2 and 3
ARPC2	Actin related protein 2/3 complex, subunit 2, 34 kDa
ARPC5L	Actin related protein 2/3 complex, subunit 5-like
ATGL/PNPLA2	Patatin-like phospholipase domain containing 2
BCAR1	Breast cancer anti-estrogen resistance protein 1
CAV1	Caveolin 1
COBL	Cordon-bleu protein
COBLL1	Cordon bleu like protein 1
COL15A1	Collagen, type XV, alpha 1
COL16A1	Collagen, type XVI, alpha 1
COL18A1	Collagen, type XVIII, alpha 1
COL3A1	Collagen, type III, alpha 1
COL5A2	Collagen, type V, alpha 2
COL6A1	Collagen, type VI, alpha 1
COL6A2	Collagen, type VI, alpha 2
COL8A1	Collagen, type VIII, alpha 1
CXCR7	Chemokine (C-X-C motif) receptor 7
EGFR	Epidermal growth factor receptor
EPAS1	Endothelial PAS domain protein 1
FGFR	Fibroblast growth factor receptor
FKBP11	FK506 binding protein 11, 19 kDa
FLNA	Filamin A, alpha
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GDF15	Growth differentiation factor 15
GLIS3	GLIS family zinc finger 3
GLUT4	Solute carrier family 2, facilitated glucose transporter member 4
GnRH	Gonadotropin releasing hormone receptor
GPDH	Glycerol-3-phosphate dehydrogenase
GRAP	GRB2-related adaptor protein

GRB10	Growth factor receptor-bound protein 10
GRB14	Growth factor receptor-bound protein 14
GRB7	Growth factor receptor-bound protein 7
HSL/LIPE	Lipase, hormone-sensitive
IL1RL1	Interleukin 1 receptor-like 1
ILK	Integrin-linked kinase
INSL4	Insulin-like 4 (placenta)
IR	Insulin receptor
ITGA1	Integrin, alpha 1
ITGA11	Integrin, alpha 11
ITGA5	Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
ITGB1	Integrin, beta 1 (fibronectin receptor, beta polypeptide)
ITGB5	Integrin, beta 5
ITIH5	Inter-alpha-trypsin inhibitor heavy chain family, member 5
KISS1	KiSS-1 metastasis-suppressor
LAMA4	Laminin, alpha 4
LAMA5	Laminin, alpha 5
LAMB2	Laminin, beta 2
LAMC1	Laminin, gamma 1
LEP	Leptin
LIMS2	LIM and senescent cell antigen-like domains 2
MAPK3	Mitogen-activated protein kinase 3
MORC4	MORC family CW-type zinc finger 4
N-WASP	Neural Wiskott-Aldrich syndrome
PACSIN1	Protein kinase C and casein kinase substrate in neurons protein 1
PACSIN2	Protein kinase C and casein kinase substrate in neurons protein 2
PACSIN3	Protein kinase C and casein kinase substrate in neurons protein 3
PCDH1	Protocadherin 1
PDGFR	Platelet-derived growth factor receptor
PIK3C2B	Phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 beta
PLIN/PLIN1	Perilipin 1
POU2F1	Pou domain, class 2, transcription factor 1
POU2F2	Pou domain, class 2, transcription factor 2
PPARG	Peroxisome proliferator activated receptor gamma
PSG2	Pregnancy specific beta-1-glycoprotein 2

PSG8	Pregnancy specific beta-1-glycoprotein 8
RAD51L1	RAD51 paralog B
RAP2A	RAP2A, member of RAS oncogene family
SIGLEC6	Sialic acid binding Ig-like lectin 6
SLC16A4	Solute carrier family 16, member 4 (monocarboxylic acid transporter 5)
SLC19A3	Solute carrier family 19, member 3
SRC	V-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)
Srv2/CAP	Adenylyl cyclase associated protein
SVEP1	Sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1
TCF7L2	Transcription factor 7 like 2
TekTie2	Tunica endothelial kinase
TINAGL1	Tubulointerstitial nephritis antigen-like 1
TLN1	Talin 1
VGLL3	Vestigial like 3 (Drosophila)
WAVE-1	WASP-family verprolin homologous protein 1

Abstract

Type 2 diabetes (T2D) is a common complex metabolic disease, caused by the interplay of environmental and genetic risk factors. Diet and lifestyle are long acknowledged as the most important environmental risk factors. Twin studies support a strong genetic component, with an estimated narrow-sense heritability of 54–83 %. Additional evidence for the heritability of T2D originates from genome-wide association studies (GWAS). These studies report associations between genomic haplotypes and a disease. Until today, GWAS have revealed more than 186 genetic risk loci for T2D. The majority of these loci are located in the non-coding part of the genome, which challenges the dissection of the signals and the type of action leading to T2D.

One complex and challenging-to-interpret T2D risk locus is *COBLL1/GRB14*. Several intronic variants near the 3'-UTR (2q24.3) of the *cordon bleu like protein 1 (COBLL1)* and intergenic between the *growth factor receptor-bound protein 14 (GRB14)* and *COBLL1*, were associated with insulin resistance and an increased risk for T2D. Although, *GRB14* – the negative regulator of the insulin receptor – was proposed to be causal for the T2D risk association at this locus, the mechanisms underlying this pathomechanism are not yet understood.

Here, we aim to dissect the role of non-coding variants at the T2D risk locus *COBLL1/GRB14*. We establish (1) the relevant cell type of action, (2) the causal variant, (3) the upstream regulator and (4) the downstream target leading to the organismal phenotype. We exploit epigenomics and regulatory genomics data sets, coupled with a systematic experimental validation of regulatory and cellular phenotypes.

We show that the non-coding rs6712203 C-to-T single-nucleotide polymorphism may underlie in parts the T2D association of the *GRB14/COBLL1* locus association, by a disrupted POU class 2 homeobox 2 (POU2F2) activator motif in risk C-allele carriers. Epigenetic maps reveal a cell type-specific enhancer, surrounding rs6712203 in adipose- and mesenchymal-derived cells. Cis- and trans- conditional analysis in CRISPR/Cas9 edited primary human adipocytes reveal its functional core: Conditional on POU2F2, the rs6712203-T non-risk variant enhancer activation increases the expression of the *COBLL1* gene in adipocytes, whereas the expression of the *GRB14* gene appears to be unchanged. Further, we find that a perturbed *COBLL1* gene expression leads to disturbed actin remodeling during adipogenesis followed by a lower insulin-stimulated glucose uptake, triglyceride storage and response to lipolytic stimulation.

Our results explore the mechanistic basis for the *COBLL1/GRB14* association with T2D. We show a connection between the expression of the *COBLL1* gene in adipose tissue and the non-coding T2D signal at 2q24.3, by uncovering the regulatory variant rs6712203 and its transcriptional regulator POU2F2. For *COBLL1*, we reveal a previously unknown role in adipocyte actin remodeling, which provides a link to adipocyte maturation, lipid storage and release - contributors to insulin resistance and T2D.

Kurzbeschreibung

Typ 2 Diabetes (T2D) ist eine Stoffwechselkrankheit mit einer weltweit hohen und stetig steigenden Prävalenz, die durch das Zusammenspiel von Umwelt und Genetik hervorgerufen wird. Ernährung und Lebensweise sind seit langem als Einflussfaktoren für T2D bekannt, die genetische Grundlage von T2D - mit einer Vererbungsrate von 54-83 % - wurde hingegen erst vor einigen Jahren in Zwillingsstudien beschrieben. Genomweiten Assoziationsstudien (GWAS) lieferten weitere Evidenz für den Einfluss der Genetik auf eine Erkrankung mit T2D. GWAS, welche den Zusammenhang zwischen genomischen Haplotypen und einer Krankheit untersuchen, haben mehr als 186 Varianten identifiziert, welche das Risiko für T2D erhöhen. GWAS wurden in der Vergangenheit kritisiert, da die meisten der identifizierten Risikovarianten im nicht kodierenden Teil des Genoms und in komplexen Haplotypen lokalisiert sind, wodurch die Interpretation der zugrundeliegenden biologischen Mechanismen nicht direkt möglich ist.

Eine besonders schwer zu interpretierende T2D-Assoziation liegt am *GRB14/COBLL1* Lokus vor. Hier wurden nicht kodierende Varianten, die in der Nähe der 3'- nicht kodierenden Region (3'-UTR) (2q24.3) des *Cordon bleu like protein 1 (COBLL1)* Gens und zwischen den Genen *growth factor receptor-bound protein 14 (GRB14)* und *COBLL1* liegen, mit Insulinresistenz und T2D assoziiert. Obwohl *GRB14*, welches als negativer Regulator des Insulinrezeptors bekannt ist, als kausaler Kandidat für das Risikosignal vorgeschlagen wurde, sind die zugrundeliegenden Mechanismen für das erhöhte T2D-Risiko bisher noch nicht bekannt.

In dieser Arbeit sollen die Rolle der nicht kodierenden Varianten innerhalb des T2D-Risikolokus *GRB14/COBLL1* untersucht werden: Es sollen (1) der verantwortliche Zelltyp, (2) die kausale Variante (3) der Masterregulator und (4) das regulierte Gen, welche zum erhöhten Risiko für T2D führen, identifiziert werden. Hierzu nutzen wir epigenomische und genetische Techniken und Datensätze, in Verbindung mit einer systematisch experimentellen Validierung von regulatorischen und zellulären Phänotypen.

Unsere Daten zeigen, dass der nicht kodierende rs6712203 C → T Polymorphismus, durch ein konserviertes POU2F2 Aktivator-Motiv in T-Nicht-Risiko Allel-Trägern, dem GWAS T2D Risiko-Signal im *GRB14/COBLL1* Lokus, zum Teil zugrunde liegen könnte. Epigenetische Analysen in Adipozyten und mesenchymalen Zellen zeigen einen Zelltyp spezifischen Enhancer am rs6712203-Lokus. Cis- und trans- abhängige Experimente in CRISPR/Cas9 editierten primären humanen Adipozyten zeigen die Funktionsweise: Abhängig von POU2F2 wird die Expression des Gens *COBLL1* in Adipozyten von rs6712203-T Nicht-Risiko Allel-Trägern erhöht, wobei keine Regulation von *COBLL1* in rs6712203-C Risiko Allel-Trägern stattfindet. Die Expression von *GRB14* wird durch die rs6712203 Editierung nicht beeinflusst. Die Funktion von *COBLL1* in Adipozyten war bisher noch nicht bekannt. Unsere Untersuchungen zeigen, dass eine Perturbation von *COBLL1* zu einer fehlerhaften Aktinfilamentierung während der Adipogenese, gefolgt von einer

niedrigeren Insulin-stimulierten Glukoseaufnahme, Triglyzeridspeicherung und stimulierten Lipolyse, führt.

Unsere Experimente zeigen eine mögliche mechanistische Grundlage für den Zusammenhang zwischen dem *COBLL1/GRB14* Lokus und T2D. Wir zeigen eine Verbindung zwischen der Expression von *COBLL1* und dem nicht kodierenden T2D Signal 2q24.3, indem wir die regulatorische Variante rs6712203 und den am T-Nicht-Risiko Allel bindenden Transkriptionsfaktor POU2F2 identifizieren. Darüber hinaus beschreiben wir zum ersten Mal den Einfluss von *COBLL1* auf die Aktinfilamentierung in Adipozyten, was zu einer verringerten Adipogenese, Lipidspeicherung und stimulierten Lipolyse beitragen kann - Risikofaktoren für Insulinresistenz und T2D.

Introduction

Diabetes prevalence and threats

Diabetes is one of the most common diseases worldwide, with an alarmingly continuous increase in worldwide prevalence. The number of people with diabetes is predicted to rise between 2017 and 2045 from 451 to 693 million and about 49.7 % of them are undiagnosed (Cho et al. 2018). For treatment and prediction purposes diabetic cases are commonly divided into type 1 diabetes (T1D) and type 2 diabetes (T2D), with intermediates classified as latent autoimmune diabetes in adults (LADA), maturity-onset diabetes of the young (MODY) and others (Prasad & Groop 2015). T2D is characterized as a group of disorders, where the inability to assimilate glucose, caused by defects in insulin secretion, the response to insulin or a combination of both, results in a disturbed glucose homeostasis (Anon 2014). It was previously referred to as “non-insulin-dependent” or “adult-onset” diabetes and accounts with > 90 % of diabetic cases, for the fastest growing subgroup of diabetes (American Diabetes Association 2014). The increase in T2D prevalence can be partially attributed to increases in overweight and obesity (McAllister et al. 2009) in the general population, due to a higher caloric intake, accompanied with less physical activity. These lifestyle changes result not only in an increase of T2D prevalence but also an earlier onset (D’Adamo & Caprio 2011) and more severe diabetic complications, including limb-threatening neuropathy, vasculopathy, retinopathy, renal insufficiency, cardiovascular disease, stroke, myocardial infarction and sudden death (Hannon 2005). As such, T2D is one of the major threats to the global health systems (Groop & Pociot 2014).

Although body mass index (BMI), age and sex are major predictors for the risk of T2D, population-based differences can still be found after adjusting for these factors. The differences are attributable to other environmental and genetic factors (Prasad & Groop 2015). The heritability of T2D risk was initially found in family and twin based studies (Meigs et al. 2000; Newman et al. 1987) and the subsequent “revolution” of genome wide association studies (GWAS) (Welter et al. 2014), enabled the identification of genetic variations associated with the risk for T2D and related traits.

The genetic architecture of T2D

The genetic architecture of T2D, defined by the number, frequencies and effect sizes of causal alleles (Prasad & Groop 2015), is still not completely understood, its understanding is challenged by several confounding factors: One is the incoherent and complex sum of phenotypes summarized by the term T2D (Grant et al. 2009). Projects, like ANDIS (All New Diabetics in Scania) and ANDIU (All new diabetics in Uppsala), are using genetic and diabetic trait markers in order to further sub-classify T2D cases (Prasad & Groop 2015; Mohlke & Boehnke 2015) with potentially more coherent disease phenotypes. In order to sustainably improve our understanding

of the genetic architecture of T2D, deciphering and understanding GWAS signals is crucial. In the last years, GWAS have increased their sample sizes and were able to associate more than 186 genetic variants with T2D (Xue et al. 2018). This number is higher when glycemic and lipid traits, like insulin secretion, insulin action, obesity, adverse lipid profiles, high blood pressure and inflammation are separately considered (Prasad & Groop 2015; Mohlke & Boehnke 2015). In 2008, the concept of the "missing heritability" was introduced (Visscher et al. 2017). Some hypothesized that a large number of common variants of small effect sizes with none or a small fraction of additive effects may explain the heritability of complex diseases (Plomin et al. 2009). Others hypothesized that rare alleles might be responsible for the heritability of complex diseases which may have higher effect sizes (Lupski et al. 2011; McClellan & King 2010), based on the idea that severe deleterious variants are usually removed by natural selection before they reach high allele frequencies in the population (Diamond 2003). However, this hypothesis was rejected for T2D, with an average of 54 % mostly common (Visscher et al. 2017) so far identified risk alleles in the general population (Fuchsberger et al. 2016; Prasad & Groop 2015). These high allele frequencies for T2D risk alleles may be explained by a low natural selection during periods with limited food supply (energy-saving thrifty genotypes theory) (Chukwuma & Tuomilehto 1998). Today, it has been shown that the theory of the "missing heritability" was due to the small number of detected T2D risk loci at the time, which could only explain a small proportion of the heritability and confounding factors like Gene x Environment effects and unknown causal variants. It is estimated that the consideration of all variants genotyped in GWAS cohorts, instead of only the tagging variants with the lowest p-values, could already explain over 50 % of the estimated T2D heritability (Prasad & Groop 2015) and the continuous increase in GWAS sample sizes will also increase discovered T2D risk variants, which will cumulatively explain most of the heritability in the next couple of years (Visscher et al. 2017). In addition, with the rise of more advanced computational methods (Lin & Lane 2017) the introduction of complex mathematical models, simultaneously including all available measured omics-data, disease relevant traits (Mohlke & Boehnke 2015) and all GWAS genotyped variants (Prasad & Groop 2015) may give rise to a better understanding of the etiology and the risk for T2D in individuals (Lupski et al. 2011; McClellan & King 2010).

Genome wide association studies (GWAS)

Inherently, GWAS have the great advantage of generating hypothesis free phenotype-genotype associations without the bias of previously collected functional data. The promise of GWAS is to enable novel insight into gene regulatory processes, the function of so far uncharacterized genes, and their mode of action which may promote advances towards precision medicine. However, one disadvantage of GWAS is that the regulatory network underlying a risk signal cannot be directly inferred from the association signal (Visscher et al. 2017). Complex trait mapping by GWAS is based on the premise that a causal variant is located within a haplotype, and therefore a marker

allele in LD with the causal variant can be associated only by proxy with the trait of interest (Supplementary Figure 1) (Stranger et al. 2011). Additionally, more than 80 % of the detected variants lie in non-coding sequences lacking protein-altering variants in LD, which makes them inaccessible to therapeutics. These non-coding sequences cover 99 % of the genome and their functional annotation is still rudimentary and difficult to interpret. This challenges the translation of GWAS-derived associations into mechanistic insights about diseases and traits (Claussnitzer et al. 2015).

Nevertheless, experimental advances in genomics and epigenomics research facilitate in parts the interpretation of GWAS. Deep-sequencing of large cohorts may enable the identification of rare and private variants, the imputation of newly calculated linkage data into GWAS cohorts may enable a better understanding of haplotypes and causal variants and the sequencing of additional ethnic groups may reveal differences in minor genetic variations (Grant et al. 2009; McCarthy 2008), leading to altered LD, allele frequencies and allelic effect sizes when compared to the european reference populations (Morris 2011). Studies leveraging these differences cannot only reproduce already identified disease risk loci but can also lead to the identification of new variants and the prioritization of candidate genes (Morris 2011; Li & Keating 2014; Kooner et al. 2011). Furthermore, epigenomic projects like the ENCODE project, the NIH Epigenomics Roadmap (Roadmap Epigenomics Consortium et al. 2015) and the eGTEx project generated epigenomics reference maps for primary human cells and tissues, analyzing DNA accessibility and methylation, histone modifications and RNA expression. They have been used to identify regulatory elements, modules of coordinated activity, their activators and repressors and the relevant cell types (Roadmap Epigenomics Consortium et al. 2015). In conclusion, each novel or improved experimental genomics and epigenomics approach will add information to understanding the biological mechanisms leading to T2D. Still, completely dissecting the genetic basis of T2D requires an integral framework, comprising epigenomics profiling, comparative genomics and large-scale expression data combined with experimental validations. Claussnitzer and colleagues have developed the comparative genomics approach PMCA (Claussnitzer et al. 2014), which combined with cutting edge experimental validation identified rs4684847, a non-coding functional variant in the *PPARG* locus associated with T2D. Briefly, for the *PPARG* locus the homeobox factor PRRX1 was found to bind to the rs4684847 risk allele, which leads to a repression of the *PPARG* isoform *PPARG2* in adipose cells. *PPARG2* is a master regulator in adipocytes (Rosen & MacDougald 2006) and its repression leads to adverse effects on lipid metabolism and systemic insulin sensitivity. In this project the isoform specific *cis*-regulatory role of a non-coding intronic variant, was shown in an adipocyte specific manner (Claussnitzer et al. 2014). An adopted version of PMCA was established by Claussnitzer and colleagues at the Computer Science and Artificial Intelligence Lab at MIT, which was integrated with large-scale epigenomic profiles across 127 cell types and tissues and has been used to identify the regulatory variant underlying the strongest association with obesity, the *FTO* locus (Claussnitzer et al. 2015). For the reported *FTO* locus, a

long range regulatory super-enhancer specifically active in adipocyte progenitors from risk allele carriers was identified. The stretch enhancer harbors a disrupted ARID5B repressor motif at rs1421085, resulting in an overexpression of *IRX3* and *IRX5*. Changes in mitochondrial activity lead to a shift from energy storage to dissipation (Claussnitzer et al. 2015).

These studies highlight the potential of an in-depth analysis of GWAS statistical signals for complex diseases in humans, uncovering novel molecular, cellular and organismal functions of genes, previously not identified by hypothesis driven approaches. Furthermore, these studies help to overcome certain challenges deciphering GWAS, including the identification of causal variants, long genomic regulations, relevant cell types, regulators, target genes and their biological role. GWAS have already increased our knowledge about the causes and genetic architecture of T2D (Groop & Pociot 2014). The interpretation of their signals will help to increase their predictive power.

The T2D risk locus *GRB14/COBLL1*

GWAS reported several non-coding T2D risk variants at the *GRB14/COBLL1* locus, but the regulatory circuitry leading to T2D still has to be uncovered. The variant rs3923113, was first associated with T2D (OR 1.09 (1.06–1.13), p-value = 1.0×10^{-8}) and insulin sensitivity (p-value = 5.0×10^{-4}) in a population of South Asian ancestry (Kooner et al. 2011) and was later verified in a trans-ethnic meta-analysis (p-value = 1.5×10^{-6}) (Europeans: RAF = 0.61, OR = 1.04 (1.00 - 1.09), p-value = 4.0×10^{-2}; East Asians: RAF = 0.77, OR = 1.03 (0.95 - 1.12), p-value = 4.8×10^{-1}; South Asians: RAF=0.75, OR 1.15 (1.09-1.21), p-value = 8.9×10^{-7}; Mexican (American): RAF = 0.76, 1.22 (1.03 - 1.45), p- value = 2.3×10^{-2}) (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. 2014). Another tagging variant rs13389219, in the same haplotype (in LD $r^2 = 0.79$ with the tagging variant rs3923113), was associated in a European Metabochip-study with T2D (OR = 1.07 (1.05–1.10), p-value = 1.0×10^{-8}) and insulin resistance (HOMA-IR, p-value = 2.2×10^{-3}) (Morris et al. 2012). The haplotype is located in an intergenic region between *COBLL1* and *GRB14*. The association region spans from the fifth 3'-intron of *COBLL1* to the intergenic region between *GRB14* and *COBLL1*, clustering around the 3'-UTR of *COBLL1*. Since the nearest gene to the tagging variant rs3923113 is *GRB14*, which can act as negative regulator of the insulin receptor, the association signal was previously referred to as *GRB14* locus (Schleinitz et al. 2014; Scott et al. 2014). However, up to now *GRB14* has not been convincingly shown to be the target gene of the T2D association.

Notably, the exonic *COBLL1* variant rs7607980 (MAF 12.5 %) (Albrechtsen et al. 2013), a missense mutation in the third *COBLL1* 3'-exon, has been associated with T2D (OR 0.88, $p = 1.2 \times 10^{-11}$) (Albrechtsen et al. 2013), lower fasting insulin and HOMA-IR levels (Mancina et al. 2013). This exonic variant, is unlikely to completely underlie the non-coding *GRB14/COBLL1* T2D

association signal, since it is in partial LD with rs3923113 (CEU r2 = 0.27; ASN/AFR r2 < 0.2, haploreg v3) and rs13389219 (r2 = 0.24, CEU; ASN/AFR r2 < 0.2, haploreg v3).

In addition to T2D and insulin resistance, GWAS also reported non-coding variants at the *GRB14/COBLL1* locus with waist-hip ratio (WHR) (Heid et al. 2010), HDL cholesterol (Teslovich et al. 2010; Randall et al. 2013) and triglyceride levels (Desmarchelier et al. 2014; Teslovich et al. 2010). Pleiotropic studies associated *GRB14/COBLL1* with metabolic syndrome and inflammation (HDLC, TG, PAI-1 and ADIP) (Kraja et al. 2014), with endometriosis and WHR (adjusted for BMI) (Rahmioglu et al. 2015) and with body fat percentage and cardiovascular events (Lu et al. 2016). GWAS also found a sexual dimorphism at the GRB14/COBLL1 locus for WHR and BMI, consistently conveying stronger effects in women (Heid et al. 2010; Randall et al. 2013; Morris 2011). Interestingly, both *GRB14* and *COBLL1* have been mentioned in breast cancer studies, investigating the effects of estrogen. *GRB14* has been shown to be overexpressed in estrogen receptor positive breast cancer (Scharf et al. 2004) and *COBLL1* expression is repressed after the stimulation with estradiol in MCF-7 breast cancer cells (Ovaska et al. 2013). Both *GRB14* and *COBLL1* are compelling candidate target genes, for the underlying mechanisms of the *GRB14/COBLL1* risk T2D signal. However, results from expression analyses implicating *GRB14* or *COBLL1* with the disease signal have not been comprehensive and could not elucidate the regulatory circuitry (Heid et al. 2010; Morris 2011; Albrechtsen et al. 2013; Schleinitz et al. 2014).

Structure and function of *GRB14*

The cytoplasmic adaptor protein GRB14 belongs to the Grb family, which consists of 14 members. They were originally identified as *epidermal growth factor receptor (EGFR)* interacting proteins. The Grb7 subfamily (Kasus-Jacobi et al. 1998), comprises three closely related proteins GRB7, GRB10 and GRB14 (Scharf et al. 2004). All of them have been reported to bind to a number of activated receptor tyrosine kinases. For *GRB14*, these include the *fibroblast growth factor receptor (FGFR)*, the *platelet-derived growth factor receptor (PDGFR)*, the *tunica endothelial kinase (Tek/Tie2)* receptor and most importantly the *insulin receptor (IR)*. *GRB14* consists of the ras-association (RA), the pleckstrin homology (PH), the Src homology 2 (SH2) and the BPS/PIR (between PH and SH2) domain (Depetris et al. 2005). While SH2 is the most important domain for the interaction of GRB14 with most receptor tyrosine kinases, the primary binding interaction with the IR is believed to be mediated by the BPS domain (Scharf et al. 2004). Upon binding of GRB14 to the activated IR, the cellular insulin response is down-regulated, by limiting the IR-tyrosine kinase activity (Supplementary Figure 2). In rats *rGrb14* has been shown to be specifically expressed in the brain and insulin responsive tissues (Kasus-Jacobi et al. 1998). In fasting mice insulin and the insulin sensitizing thiazolidinediones, have been shown to increase the expression of *Grb14* in adipose tissue (Cariou et al. 2004). On the other hand, metformin has been shown to decrease the *Grb14* gene expression (Cariou et al. 2004). Adult male *Grb14* knock-out mice have

an increased body weight but at the same time an improved glucose tolerance, lower insulin levels, a higher glycogen production in the liver and skeletal muscle accompanied with an increased glucose uptake in muscle. In these mice, no effect has been observed on insulin signaling in white adipose tissue (Cooney et al. 2004). When both *Grb10* and *Grb14* are ablated in mice, they are additionally protected from the impaired glucose tolerance, usually resulting from high-fat feeding (Holt et al. 2009). In human visceral adipose tissue, the expression of *GRB14* has been shown to be increased in obese and type 2 diabetic subjects but no difference was found between men and women (Schleinitz et al. 2014).

Structure and function of *COBLL1*

The *COBLL1* gene encodes for the COBLL1 protein, which comprises two repeated lysine, arginine, and proline-rich regions (KKRAP motifs) several phosphorylation sites, an N-terminal COBL -ubiquitin like (PDB 2DAJ) and C-terminal WASP (Wiskott–Aldrich syndrome protein)-homology 2 (WH2) domain (RCSB Protein Data Bank ID Q53SF7 (Zhao et al. 2006)). The assembly of the *COBLL1* gene can be traced back to the occurrence of the first bilateria and was presumably a result of the combination of an already existing ubiquitin fold, which became the COBL-domain, and a C-terminal WH2 domain. In contrast to *GRB14*, the functional role of *COBLL1* in metabolic diseases is poorly understood in humans and other mammals. The actin-interacting WH2 domain (Schultz & Terhoeven 2013) of *COBLL1* has been shown to enable actin filamentation in prostate cancer cells (Takayama et al. 2018). Interestingly orthologs of *COBLL1* have been identified in many nematodes (Schultz & Terhoeven 2013). In humans the paralog *COBL* exists, which is believed to have originated from *COBLL1* gene duplication, with an aggregation of three adjacent WH2 domains (Schultz & Terhoeven 2013). *COBL* functions as potent actin dynamizer with an ability to nucleate and sever actin filaments (Wayt & Bretscher 2014; Jiao et al. 2014).

COBL/GRB10 and *COBLL1/GRB14* risk loci for T1D and T2D

GRB14 (GRCh38/hg38; chr2:164492816-164621848) and *COBLL1* (GRCh38/hg38; chr2:164684816-164842025) are adjacent genes with a distance of ~ 63 kb, they are oriented in the same direction. Their paralogs *GRB10* (GRCh38/hg38; chr7:50590068-50792969) and *COBL* (GRCh38/hg38; chr7:51016212-51316799) are also adjacent to each other with a distance of about 223 kb and also oriented in the same direction, giving evidence that *GRB10/COBL* may have emerged from a combined duplication of both *GRB14* and *COBLL1*. Interestingly, GWAS associated the *GRB10/COBL* variant rs4948088 with T1D (Barrett et al. 2009). The haplotype of rs4948088 is located near the 3'-UTR of *COBL* and between *COBL* and *GRB10*, which is similar to the intergenic rs3923113 T2D risk signal between *GRB14* and *COBLL1*.

Independent diseases usually have a random distribution of risk alleles between each other. This is not true for T1D and T2D, where more alleles - which can be risk or protective alleles - are shared than expected by chance. For example, some T1D risk loci may be protective for T2D, like *BCAR1*, *GLIS3* and *RAD51L1* (Morris et al. 2012). The risk locus *TCF7L2* - with the so far highest odds ratio for T2D may be protective for T1D (Prasad & Groop 2015) and some loci have been associated with the risk for both T1D and T2D, like *COBL* (Morris et al. 2012) and the adipocyte master regulator (Rosen & MacDougald 2006) *PPARG* (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. 2014; Eftychi et al. 2004; Raj et al. 2009). Taken together, the T2D tagging variant rs3923113 at *GRB14/COBLL1* and the T1D tagging variant rs4948088 at *GRB10/COBL* are not only located at a similar location between highly conserved paralogs but may also have a constraint regulatory circuitry leading to the increased risks for T2D and T1D.

Aim of the thesis

In this study we aimed to mechanistically dissect the *GRB14/COBLL1* locus. The *GRB14/COBLL1* locus shows one of the strongest associations with T2D (Morris et al. 2012; Kooner et al. 2011) and its underlying cell type of action, downstream target gene(s), upstream regulators and causal variant(s) are so far not known. This disease association signal is both, interesting and challenging, because there has been an inconsistency of reported target genes, the biological function, of one of these genes - *COBLL1* - is still not completely characterized and the T2D risk locus *GRB14/COBLL1* has been found to be highly conserved during evolution.

Here, we combine techniques of epigenomics, comparative genomics, human genetics, directed perturbations and genome editing to dissect the regulatory circuitry and the mechanistic basis of the *GRB14/COBLL1* type 2 diabetes risk locus (Claussnitzer et al. 2015; Claussnitzer et al. 2014).

Experimental procedures

Targeted *in-vitro* experiments

Immortalized cell lines

Unless otherwise indicated, cell culture consumables were received from Falcon (BD Bioscience, NJ, USA), Corning (Corning, NY, USA) or TPP (TPP Techno Plastic Products, Trasadingen, Schweiz). Rat INS1 β-cells were grown in advanced RPMI 1640 medium (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 1 % penicillin/streptomycin (Thermo Fisher Scientific, Waltham, MA, USA), 10 % FCS (Thermo Fisher Scientific, Waltham, MA, USA) and 50 µM β-mercaptoethanol (Sigma-Aldrich, Steinheim, Germany). Human Huh7 hepatoma and mouse 3T3-L1 cell lines were cultured in DMEM (Thermo Fisher Scientific, Waltham, MA, USA) medium supplemented with 1 % penicillin/streptomycin and 10 % FCS. For induction of differentiation 3T3-L1 cells were grown to 90 % confluence, before the medium was supplemented with 860 nM human insulin (Sigma-Aldrich, Steinheim, Germany), 250 nM dexamethasone (Sigma-Aldrich, Steinheim, Germany) and 500 µM 3-Isobutyl-1-methylxanthine (IBMX, Serva, Heidelberg, Germany). After 72 hours, cells were incubated in medium without dexamethasone and IBMX. The human preadipocyte SGBS (Simpson–Golabi–Behmel Syndrome) cell line was cultured as previously described (Fischer-Posovszky et al. 2008) in DMEM/Ham's F12 (1:1) medium supplemented with 1 % penicillin/streptomycin, 10 % FCS, 16 µM biotin (Roth, Karlsruhe, Germany) and 36 µM pantothenic acid (proliferation medium). For adipose differentiation, cells were grown to confluence and induced with serum free MCDB-131/DMEM/Ham's F12 (1:2) medium supplemented with 1 % penicillin/streptomycin, 11 µM biotin, 24 µM pantothenic acid, 10 µg/ml human transferrin (Sigma-Aldrich, Steinheim, Germany), 66 nM human insulin, 100 nM cortisol (Sigma-Aldrich, Steinheim, Germany), 1 nM trijodthyronin (T3) (Sigma-Aldrich, Steinheim, Germany), 20 nM dexamethasone, 500 µM IBMX (Serva Electrophoresis, Uetersen, Germany) and 2 µM rosiglitazone (Alexis Biochemicals, Grünberg, Germany), 96 hours post induction of differentiation dexamethasone, IBMX and rosiglitazone were removed from the medium. All cell culture experiments were conducted at 37°C and 5 % CO2.

Primary human cell culture

Informed, written consent was given by each human participant before adipose tissue samples were collected. The studies were approved by the local ethics committee of the Faculty of Medicine of the Technical University of Munich, Germany, or the Regional Committee for Medical Research Ethics (REK) of Haukeland University Hospital, Bergen, Norway (Claussnitzer et al. 2014). Primary human adipose-derived stromal cells (hASCs) were obtained by liposuction or surgical excision of subcutaneous tissue. They were isolated as previously described (Hauner et al. 2001; Claussnitzer et al. 2014). Cells were cultured in DMEM/F12 medium supplemented with 2.5 % FCS, 1 %

penicillin/streptomycin, 16 µM biotin, 36 µM pantothenic acid, 132 nM human insulin, 10 ng/ml EGF (ImmunoTools, Friesoythe, Germany), and 1 ng/ml FGF (ImmunoTools) until they were confluent. For differentiation, cells were cultured in serum free DMEM/F12 medium supplemented with 1 % penicillin/streptomycin, 16 µM biotin, 36 µM pantothenic acid, 860 nM insulin, 100 nM Cortisol, 10 µg/ml transferrin and 1 nM triiodothyronine. For initiation of differentiation, the medium was additionally supplemented with 500 nM rosiglitazone, 25 nM dexamethasone and 250 µM IBMX for the first 72 hours. Mature human adipocytes were isolated as previously described (Claussnitzer et al. 2015). Briefly, adipose tissue was dissected and digested in Krebs-Ringer-Phosphate (KRP) buffer, pH 7.4 with 127 mM NaCl (Roth, Karlsruhe, Germany), 12.3 mM NaH2PO4, 5 mM KCl, 1.27 mM MgSO4, 1.36 mM CaCl2, 4 % BSA (Sigma-Aldrich, Steinheim, Germany) and 100 U/ml collagenase (Serva Electrophoresis, Uetersen, Germany). The cells were incubated for 60 min in a shaking water bath, they were filtered through a 250 µm nylon mesh (VWR, Darmstadt, Germany) and the floating adipocytes were washed three times with KRP containing 0.1 % BSA (Sigma-Aldrich, Steinheim, Germany). The cells were incubated in DMEM/F12 containing 1 % penicillin/streptomycin.

To obtain hAC-fractions, isolated hAC were separated into fractions by flotation as previously described (Skurk et al. 2007).

Cellular hAC phenotypes were obtained from Prof. Peter Arner, Karolinska Institutet (Sweden). The donors were genotyped for variant rs6712203 as earlier described (Strawbridge et al. 2016), rs6712203 was tested for HWE and alleles were converted into a dominant-recessive model. From 505 women and 176 men (body mass index (BMI) range 17.7-62.3 kg/m^2 and mean BMI 32.2 kg/m^2), 9 had to be excluded due to missing rs6712203 sequencing results (rs6712203-C risk allele carriers 262, rs6712203-CT heterozygous allele carriers 320rs, 6712203-T non-risk allele carriers 92). Phenotypes were log transformed and NULL were substituted with minimum values in each phenotype to avoid NaN due to the transformation. Linear models for each phenotype against rs6712203, adjusted for age and BMI were computed. They were corrected for multiple testing using the QVALUE package in R.

CRISPR/Cas9 homology directed repair genome editing

To edit the rs6712203 risk (CC) and non-risk (TT) alleles we applied the CRISPR/Cas9 homology directed repair genome editing approach as previously described (Claussnitzer et al. 2015). Briefly, the hCas9 and the gRNA cloning vector were purchased from Addgene (Plasmid ID #41815 and #41824, respectively). The guide RNAs (gRNAs) were selected using the CRISPR design online tool from the Zhang lab (Zhang Lab, MIT 2013, http://crispr.mit.edu/). For amplification of the 2009 bp homology region with the risk or non-risk allele of rs6712203 at mid position, genomic DNA from SGBS cells was amplified with primers 5'-GGTGGTCCCATTAAAAAGAAAGAAGCTTGG-3' and 5'-

CTTCTCTTTTACCCTGCTGGCTACTGGTTG-3' using the High-Fidelity Q5 DNA polymerase (New England Biolabs, Frankfurt, Germany). The gel purified PCR product was cloned into the blunt end pJet1.2 vector using the CloneJET PCR Cloning kit (Thermo Fisher Scientific, Waltham, MA, USA). A clone with the rs6712203 C allele was selected and the corresponding T allele vector was generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, Frankfurt, Germany) with the primers 5'-TCATTCATCATATGCAATTCTGG -3' and 5'-GGCAAATTAATATTTAGGATTATATC-3'. To avoid Cas9 reactivity after genome editing, the NGG guide target sequence was mutated to NCG in both homology vectors with the primers 5'-CCATTGCCAACGGCTGAGTCAG-3' and 5'-TAGTGGAGAGTTCTCACAAAAC-3'. Cells were co-transfected with the GFP (Lonza, Basel, Switzerland), the hCas9, the respective gRNA, and the pMACS 4.1 (Miltenyi, Bergisch Galdbach, Germany) plasmids using the Amaxa-Nucleofector device (program U-033) (Lonza, Basel, Switzerland). Sorted cells were cultured for 3–5 days and clones were propagated from single cell. Nucleotide exchange was confirmed by SangerDNA sequencing and lack of random indels was confirmed by sequencing the 1000 bp genomic DNA flanking the rs6712203 targeted site. Further, we sequenced the top four predicted off-target sites, as computationally predicted by the CRISPR design tool (crispr.mit.edu).

Lentiviral cell transduction

To produce lentiviral particles, the MISSON® Lentiviral Packaging Mix (Sigma Aldrich, Steinheim, Germany) was used according to the manufacturer's instructions. Briefly, packaging cells HEK293T were grown in low antibiotic growth medium (DMEM, 10 % FCS, 0.1 % penicillin/streptomycin). When cells were about 70 % confluent, they were co-transfected, using X-treme GENE HP (Roche, Penzberg, Germany), with the packaging plasmid pCMVdeltaR8.91, the envelope plasmid pMD2.G and the pLKO-based plasmid containing shRNA against the human target gene (Supplementary Table 1), the empty-vector control or the non-target control MISSION® TRC2 pLKO.5-puro plasmid (SHC202; Sigma Aldrich, Steinheim, Germany). The cells were incubated for 24 hours, the medium was discarded and replaced with a serum rich medium (30 % FCS). The supernatant containing the viable virus particles was collected 48- and 72-hours post transfection, centrifuged to remove cellular debris, and stored at -80°C.

SGBS cells were seeded at a concentration of 2.6 x 10^4 cells per 6-well plate and grown in normal growth medium. After 24 hours the medium was replaced and supplemented with 8 µg/ml Polybrene (Sigma-Aldrich, Steinheim, Germany) and virus supernatant with a multiplicity of infection (MOI) of 2. On the consecutive 2 days cells were washed with PBS and medium was replaced to remove the virus. The medium was supplemented with 0.5 µg/ml puromycin, 96 hours after infection, to select stable clones. When cells were grown confluent, puromycin was removed from the medium and the cells were differentiated until day 16. Target gene silencing was confirmed after selection and during each experiment by qRT-PCR.

Oil-red O staining

For intracytoplasmic lipid staining, SBGS cells differentiated until day 16 in 6 well plates were incubated for one hour in 3.7 % formaldehyde at room temperature. The fixed cells were washed with PBS and incubated for one hour with a filtered 0.3 % Oil-red O (Sigma-Aldrich, Steinheim, Germany) H_2O:Isopropanol (2:3) solution. They were washed twice with PBS, photographed (10x/0.22; Leica DC 300FX camera; Leica, Bensheim, Germany) and scanned (EPSON EXPRESSION 1680 Pro, Epson, Klosterneuburg, Germany).

Glycerol-3-phosphate dehydrogenase (GPDH) activity measurement

Cells were grown to confluence and differentiated until day 16 in 6 well plates. Cells were collected in GPDH buffer with 0.05 M Tris/HCl (pH 7.4), 1 mM EDTA and 1 mM 2-mercaptoethanol, before they were stored at - 80°C until further use. Samples were gently defrosted on 4°C, and were sonified for 7 sec at 29 % and centrifuged for 10 min at 10.000 g on 4°C. GPDH activity was measured as previously described (Pairault & Green 1979). Briefly, GPDH activity was assessed, measuring the conversion of dihydroxyacetone phosphate (DHAP) (Sigma, St. Louis, USA), in the presence of the coenzyme nicotinamide adenine dinucleotide (NADH) (Omnilab-Applichem, Bremen, Germany) at a wavelength of 340 nm, using the Tecan Infinite 200 (Tecan, Crailsheim, Germany). Protein concentrations were assessed using the BCA-RAC protein assay kit (Thermo Fisher Scientific, Waltham, MA, USA), with BSA standard samples in GPDH buffer for quantification. The value for each condition was calculated using the ratio between GPDH activity and protein concentration.

Quantitative polymerase chain reaction (qRT-PCR)

Primer pairs (Supplementary Table 2) were designed using published nucleotide sequences from the human genome GenBank NCBI/UCSC and ensemble, the online tool “primer3input” (http://frodo.wi.mit.edu/primer3/) was used for primer design. They were optimized to reduce secondary structures, primer dimer and self-priming formation with the online design tool “net primer” (http://www.premierbiosoft.com/netprimer/index.html). The specificity of the primers to the gene of interest was verified using “primer blast” (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) NCBI GenBank. The optimal annealing temperature for each synthesized (Eurofins, Ebersberg, Germany) primer pair was determined in a gradient PCR using the Mastercycler® ep realplex (Eppendorf, Hamburg, Germany).

Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) and 0.5 µg was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems - Thermo Fisher Scientific, Waltham, MA, USA). qRT-PCR was performed using 96 well plates (black frame, white wells; 4titude, Surrey, UK) with heat sealing films (4titude, Surrey, UK), which

were fixed using the 4s2 Automated Heat Sealer (4titude, Surrey, UK). The Maxima SYBR Green Mix (Thermo Fisher Scientific, Waltham, MA, USA) was used for the amplification in a qRT-PCR Mastercycler® ep realplex (Eppendorf, Hamburg, Germany), with an initial denaturation of 95°C, 10min step and 40 cycles of 95°C, 15 sec, primer specific annealing temperature (Supplementary Table 2), 40 sec, followed by a melting curve. Relative gene expression was calculated by the delta delta Ct method (Pfaffl 2001).

Electrophoretic mobility shift assay (EMSA)

EMSA experiments were performed using double stranded Cy5-labelled probes with the risk or non-risk allele of each variant at mid-position. The forward Cy5-labelled strand (for rs6712203 5'-TTAATTTGCCTCATTCATCA[C/T]ATGCAATTCTGGCAAGGAA-3' and for rs10195252 5'-CCCCACTTCCCTCTAGGGAA[T/C]GGGAAAGAACATTTAACCT-3') and respective unlabeled reverse complementary strands were synthesized (Eurofins, Ebersberg, Germany), annealed and purified from single stranded remains by excision from a 12 % polyacrylamide gel. Nuclear protein extracts from undifferentiated hASCs were prepared with an adapted protocol based on the method described by Schreiber and colleagues (Schreiber et al. 1989) according to Claussnitzer and colleagues (Claussnitzer et al. 2014). Nuclear protein extracts from primary mature human adipocytes were extracted according to the protocol described by Dugail and colleagues (Dugail 2001).

For EMSA experiments, 1-2 µl buffer containing 3-5 µg protein was added to 10 mM TrisHCl (pH 7.5), 1 mM MgCl2, 50 mM NaCl, 0.5 mM EDTA, 4 % (v/v) glycerol, 0.5 mM DTT and 30 ng/µl poly(dI-dC). For competition experiments unlabeled oligonucleotides containing the POU2F2 binding motif 5'-GAAGAATTC**ATGCAAAT**GAATTCGAAGAAG-3' were used. The forward strand was annealed with the respective reverse complementary sequence and purified as described above. After 10 minutes incubation on 4°C for all EMSA conditions, 1 ng of the respective Cy-5 labelled probe was added, and the samples were incubated for 20 min at 4°C. After addition of loading buffer with 25 mM TrisHCl pH 7.5, 0.02 % OrangeG, 4 % glycerol, the samples were subjected onto a nondenaturing 5.3 % polyacrylamide gel. After gel separation, Cy-5 fluorescence was detected using the Typhoon TRIO+ imager (GE Healthcare, München, Germany).

Glucose uptake, lipolysis and western blot analysis

For glucose, glycerol and Western Blot analysis shRNA *COBLL1* and shRNA empty-vector SGBS cells were differentiated until day 16 in 6 well plates.

The insulin-stimulated 2-desoxy-D-glucose uptake experiment was performed as previously described (Claussnitzer et al. 2011). Briefly cells were incubated in glucose- free DMEM and F12

(1:1) containing 1 % penicillin/streptomycin, 16 µM biotin, 36 µM pantothenic acid, 14.3 mM NaHCO3 and 0.5 mM Na-pyruvate (Sigma-Aldrich, Steinheim, Germany) for 12 hours. The medium was replaced with 118 mM NaCl, 1.2 mM KH2PO4, 4.8 mM KCl, 1.2 mM MgSO4, 2.5 mM CaCl2, 10 mM HEPES, 2.5 mM Na-pyruvate (Sigma-Aldrich, Steinheim, Germany), 0.5 % BSA (Sigma-Aldrich, Steinheim, Germany) (pH 7,35). After 1.5 hours the same buffer was added fresh either without supplement or with 1 µM insulin for 30 min. The radioactive uptake was started by addition of KRH [^{3}H]-2-desoxy-D-glucose at an activity of 1 µCi/ml and 50 µM 2-desoxy-D-glucose. Cells were incubated for 30 min and then washed with PBS. The cells were scraped off, after addition of 200 µL IGEPAL and 150 µM phloretin. The radioactivity was measured using liquid scintillation counting with an external standard.

For the measurement of glycerol release, cells were washed with PBS and incubated for 3 hours in phenol red free DMEM containing 2 % FFA (free fatty acid)-free BSA (Roth, Karlsruhe, Germany). The medium was changed, and the cells were incubated for 1 hour without supplement for basal lipolysis or addition of 10 µM Isoproterenol (Sigma-Aldrich, Steinheim, Germany) and 0.5 mM IBMX for lipolysis stimulation. The supernatant was collected for spectrophotometric glycerol measurement in a Sirius tube luminometer (Berthold Technologies, Bad Wildbad, Germany), using the glycerokinase (Sigma-Aldrich, Steinheim, Germany) and the ATP Kit SL (BioThema, Handen, Sweden). Remaining cells were collected for protein quantification and Western Blot analysis in RIPA buffer containing 50 mM Tris-HCl (pH 8), 150 mM NaCl, 0.2 % SDS, 1 % NP-40, 0.5 % deoxycholate, 1 mM PMSF, phosphatase and protease inhibitors. Western Blot analysis was performed using a mouse anti-human *GAPDH* IgG (Ambion - Thermo Fisher Scientific, Waltham, MA, USA) and the Lipolysis Activation Antibody Sampler Kit #8334 (Cell Signaling, Danvers, MA, USA) according to the manufacturer's protocol. Secondary IRDye IgG (LI-COR, Bad Homburg, Germany) were used to generate the fluorescence, detected by the Odyssey scanner (LI-COR, Bad Homburg, Germany).

Staining and confocal microscopy

Primary human adipose-derived stromal cells (hASC) were isolated and differentiated for 6-7 days in 8 well µ-slides ibiTreat (ibidi, Martinsried, Germany). Cells were cultured for 12 hours in full control medium or glucose deprived medium (1:1 DMEM w/o glucose: F12, 1 % v/v B/P, 1 % v/v P/S, 55 mg/l Na-Pyruvate (Sigma-Aldrich, Steinheim, Germany), 1,2 g/l $NaHCO_3$). They were incubated with siRNA against *COBLL1*, non-target siRNA (ON-TARGETplus SMARTpool - Human, Dharmacon, Lafayette, CO, USA), 0.1 µM Latrunculin A or 0.8 µM Cucurbitacin E (both Sigma-Aldrich, Steinheim, Germany) for 7 hours. For stimulation of deprived cells, 0.2 µM insulin and 50 µM glucose was added, 30 minutes before the cells were fixed with 3.7 % paraformaldehyde for 15 min. Cells were washed with PBS, incubated for 10 min in dibasic buffer (PBS supplemented with 1 % BSA, 0.05 % Triton X-100 (all from Sigma-Aldrich, Steinheim, Germany)) and incubated for 16

hours with the *COBLL1* anti-rabbit antibody (ATLAS ANTIBODIES, Stockholm, Sweden). Cells were washed and incubated with the secondary fluorescent antibody 484-anti-rabbit (Invitrogen-Thermo Fisher Scientific, Waltham, MA, USA) and 2-(4-Amidinophenyl)-6-indolecarbamidine dihydrochloride DAPI (Sigma-Aldrich, Steinheim, Germany) for nucleic acid staining in dibasic buffer for 1 hour. They were washed and incubated for 20 minutes with 500 nM ATTO565-Phalloidin (Sigma-Aldrich, Steinheim, Germany). Before imaging, the cells were washed and covered with PBS. For imaging, the Olympus IX81 inverted confocal microscope (Olympus, Hamburg, Germany) was used with 40x magnification. Fluorescent signal intensities were analyzed using the Volocity 3D Image Analysis software version 5.4.1 (PerkinElmer, Waltham, MA, USA).

Large scale data analysis and human cohort data

Chromatin state map visualization

Chromatin state information of 127 reference epigenomes was obtained from the Roadmap-(Roadmap Epigenomics Consortium et al. 2015) and the ENCyclopedia Of DNA Elements-(Encode) (ENCODE Project Consortium 2012; ENCODE Project Consortium 2004) Epigenomics Projects. The genomic segment 165491546 - 165712339 intergenic between *GRB14* and *COBLL1* on chromosome 2 was visualized in the WashU Epigenome Browser (Zhou et al. 2011). Chromatin states were classified, using a 25-state model predicted by ChromHMM (Ernst & Kellis 2012), with combinations of 12 histone modification marks, using ChromImpute (Claussnitzer et al. 2015; Ernst & Kellis 2015).

Intragenomic replicates (IGR) method

The Intragenomic Replicates (IGR) method was used for POU2F2 affinity modeling using POU2F2 ChIP-seq data as previously described (Cowper-Sal lari et al. 2012). In order to correct for systematic bias in the sequencing depth around particular k-mers, all scores were offset by a "baseline" value, defined as the average signal between the forward and reverse complement instances of the k-mer between -200 and -195 and between 195 and 199 bases away from the k-mer center. Thus, if the value across the whole -200 to +199 context was approximately equal, then the overall score is approximately zero, and positive estimated affinities are only possible in cases where the score in the middle of the context is significantly higher than the outside. To further include only large effect binding differences, the "prominence" was defined as the maximum score across any point in the context for either the forward or reverse complement version of the k-mer for both alleles and the "maximum difference" as the maximum absolute difference in scores between the two alleles at any point in the window. The "baseline ratio" was defined as the ratio of

the maximum difference to the prominence, which varies between 0 (if the two alleles are equal at all points) and 2 (if they are perfectly complementary at their highest absolute point).

In order to find only high-quality putative disrupted binding sites, the k-mer sequence that gave the highest affinity under the germline was recorded as "reference" and the k-mer sequence which gives the highest affinity under the somatic variant as "alternate." The "quality" of a given kmer was defined as the correlation between the average context plot forward and the reverse of the average context plot of the reverse complement, and the "symmetry" of a given k-mer as the correlation between the average context plot forward and the average context plot reverse. Quality is high when the antiparallel binding is preserved, and symmetry is high when the peak signal is centered with respect to the variant. The results were included as "passed" when the Bonferroni corrected p-value for the comparison is less than 0.05, the baseline ratio is greater than 0.5, the quality and symmetry are both greater than 0.85 for one of the alleles, and the quality and symmetry are both greater than 0.5 for the other allele.

Chromosome conformation capture with next-generation sequencing (Hi-C)

Hi-C data were obtained from human IMR90 myofibroblasts and NHEK primary normal human epidermal keratinocytes as described earlier (Dixon et al. 2012). The data were mapped to the human reference genome assembly hg19, and visualized according to Imakaev and colleagues (Imakaev et al. 2012) using the Python matplotlib package.

GRB14/COBLL1 haplotypes

For *GRB14/COBLL1* haplotypes from Africans (AFR), Europeans (EUR), East Asians (ASN) and Ad Mixed Americans (AMR), LD information from the 1000 Genomes Phase 1 project (Ensembl release 74) was exported from haploreg v2. In the analysis, we used the GWAS tagging variant rs3923113 and included all variants with $r^2 > 0.7$ (CEU). The data was visualized using the LDheatmap package in R.

Microarray expression data

A global gene expression measurement using Illumina HumanRef-8 v.3 BeadChip microarrays from whole abdominal subcutaneous adipose tissue from 13 lean and 17 obese subjects was performed and analyzed by Simon N. Dankel (University of Bergen). Signal intensities were quantile normalized, before the expression analysis of *COBLL1* in hAC and hASC and correlation analysis using PANTHER.

EuroBATS cohort

In the EuroBATS (No. 259749) project, samples from TwinsUK (Moayyeri et al. 2013) were used, which consist in our analysis of 720 females with a mean age of 59.4, ranging from 39 to 85 and a mean BMI of 26.68, ranging from 16 to 47. To determine whether expression of *POU2F2* or *COBLL1* exons is associated with BMI, each rank normalized exon expression level was treated as a quantitative trait in a linear mixed effects model using the lme4 package. BMI was tested for approximate normality using a Shapiro-wilk test and visualized using a qqnorm plot, implemented in R. BMI was treated as continuous independent variable. Depending on the distribution, outliers were removed, and normalization performed. Diabetic and non-fasting subjects were removed, and BMI was adjusted for age and age squared. RNA-Seq and population based technical covariates, such as GC mean, Primer index, insert size mode and family structure were adjusted. The full fitted model (with BMI) was compared against a null model using ANOVA to assess significance of each trait affecting exon-expression levels. The FDR 5 % significance threshold was computed using the QVALUE package in R.

Relative quantification of splicing events for the calculation of fractions between *COBLL1* exon-links was performed using the Altrans method (Ongen & Dermitzakis 2015). The exon counts were used in a linear mixed effects model, to assess the rs6712203 dependent effect.

Each linear mixed effect model was adjusted for RNA sequencing and population-based covariates, such as GC mean, Primer index, insert size mode and family structure, BMI (for all but BMI association) and age. The full fitted model, including the test variable, was compared against a null model using ANOVA to assess significance. The FDR 5 % significance threshold or the qvalue was computed using the QVALUE package.

Dietary intervention study

The dietary intervention study was performed and analysed by Beate Ott, while performing her PhD-Thesis at the Lehrstuhl für Ernährungsmedizin (Prof. Dr. Johann J. Hauner), TUM. The study was conducted as earlier described (Ott et al. 2017; Cordes et al. 2015). Briefly, twenty obese women received a hypocaloric diet with 800 kcal per day and ad libitum intake of vegetables over a period of 4 weeks. Before and after the intervention, phenotypes were assessed using anthropometric and laboratory measures relevant in inflammation, lipid and glucose metabolism. RNA was extracted from needle biopsies of SAT and the expression of *COBLL1* mRNA was measured by qRT-PCR as described above. Three participants were disregarded during analysis due to incomplete or outlier measurements. All variables were log transformed in order to receive comparable scales. Phenotypes were adjusted for age, BMI, WHR and amount of fat in a linear mixed effects model using the lmer package in R. The full fitted model was compared against the

null model using ANOVA and p-values were corrected for multiple testing using the QVALUE package in R.

Statistical analysis

Unless otherwise indicated, statistical analysis was performed in R version 3.2.2. Normal distribution was tested and statistical tests were chosen accordingly as described in each figure description. When linear (mixed effects) models were used, their adjustments were separately described for each experiment in the methods section above. For multiple comparisons p-values were corrected using the Q-VALUE package in R.

Results

Adipose and mesenchymal enhancer marks at the T2D risk locus *GRB14/COBLL1*.

Several GWAS reported an association between the *GRB14/COBLL1* locus and the risk for T2D. The tagging variant rs3923113 (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. 2014; Kooner et al. 2011) is in high linkage disequilibrium (LD) with 28 other non-coding variants in Europeans (EUR $r^2 > 0.7$; HaploReg v2; Broad Institute) (Figure 1B), located in a region of about 55.5 kb (chr2: 164645339-164700808; hg38), reaching from the intergenic region between *GRB14* and *COBLL1* to an intron of *COBLL1* (Figure 1A). Another tagging variant rs13389219 is also robustly associated with T2D and located in the same haplotype (Morris et al. 2012) (Figure 1B). The association of at least two tagging variants within *GRB14/COBLL1* with T2D highlights the complex haplotype and the challenge to identify the disease causal variant on a genome wide level, without the integration of data from several levels of computational and experimental approaches.

To narrow down the association region to the causal variant and to identify the mechanism leading to T2D, we first determined the relevant cell type of action at the *GRB14/COBLL1* locus. We leveraged chromatin state information, from the Roadmap project, with 127 human cell or tissue types (Figure 1C). In order to identify putative regulatory elements, we primarily focused on the enhancer states 9-18 inferred by ChromHMM (Supplementary Figure 5). In primary human cells and tissues from the Roadmap project (Upper panel Figure 1C), we found a specific overlap at the *GRB14/COBLL1* locus with adipose-derived and mesenchymal cell types and only few enhancer marks across other cell types. In data from the ENCODE project (Lower panel Figure 1C), we additionally found enhancer marks across other cell types, which may be caused by the immortalization of the cells. We concluded that adipose tissue cultured, or derived cells may be relevant for T2D causing mechanisms at the *GRB14/COBLL1* locus.

Strikingly, across all 127 human cell types and tissues from the Roadmap Consortium, the potentially functional variant rs6712203 (chromosome 2: 164700808) identified by PMCA (Claussnitzer et al. 2014), showed enhancer states specifically in adipose tissue cultured or derived cells, and not in any other cell type. More precisely, enhancer marks were only observed in (1) mesenchymal stem cell derived adipocyte cultured cells, (2) adipose derived mesenchymal stem cell cultured cells, and (3) adipose nuclei (Figure 1C).

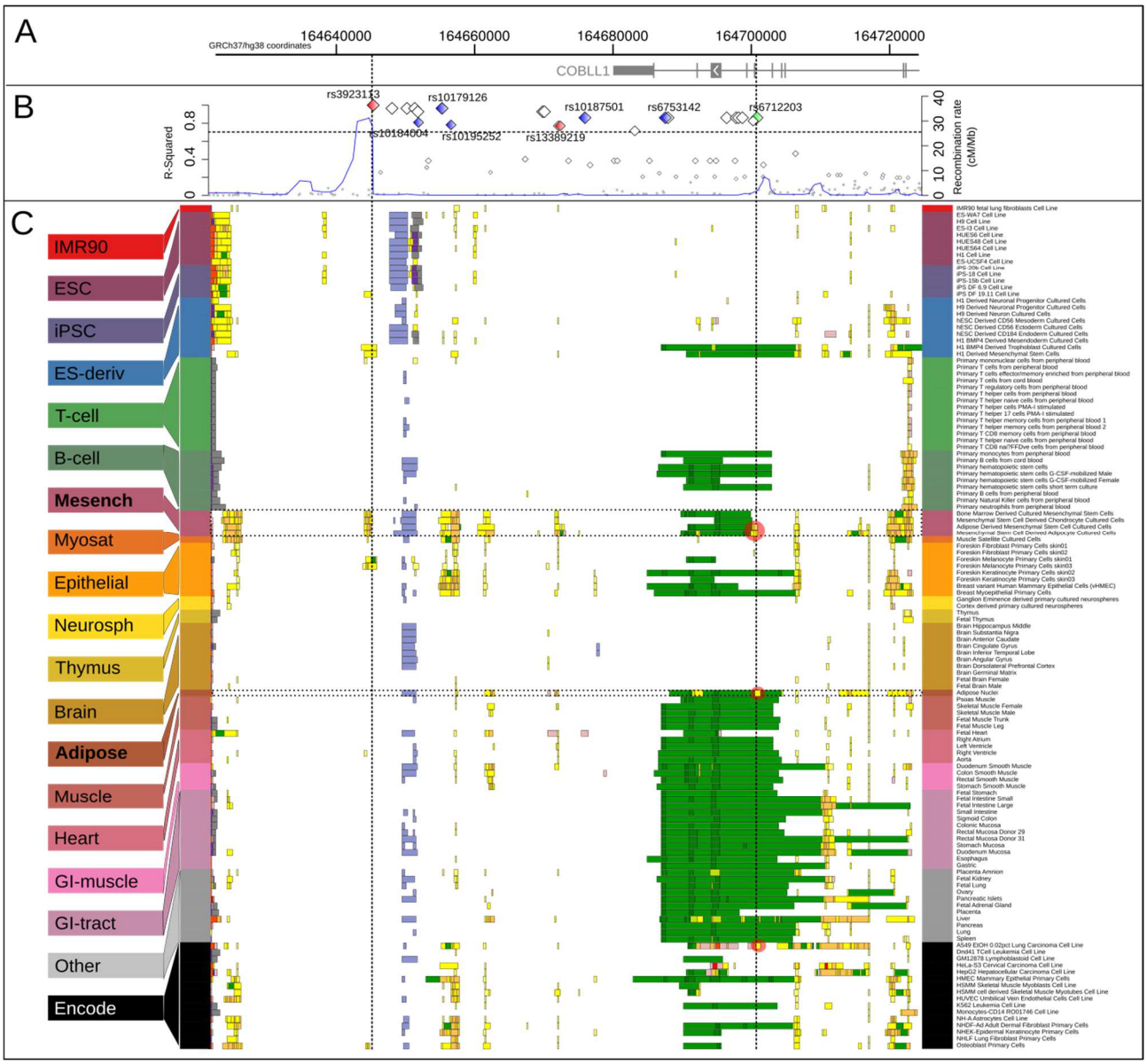


Figure 1: Cell type specific enhancer marks at the T2D risk locus GRB14/COBLL1. Panel A shows annotated COBLL1 3'-exons located in the range between 163768829 and 163864298 (GRCh37/hg38) on chromosome 2. In Panel B, variants in LD with the tagging variant rs3923113 are shown (SNAP proxy server, Broad Institute; CEU, 1000 Genomes pilot 1, variants in a range of 500kb were tested for LD). The horizontal dotted line marks $r^2 \geq 0.7$. The vertical dotted lines delineate the GRB14/COBLL1 haplotype, which is shown in a window of ± 20kb. Variants are represented by diamonds; their size and color correspond to r^2. Filled diamonds are annotated with the respective rs-numbers. Red diamonds represent the main GWAS tagging variants. The green diamond represents the top-scoring PMCA variant rs6712203, which may contribute to the T2D association signal at the GRB14/COBLL1 locus. Blue diamonds represent further PMCA predicted candidate regulatory variants (Claussnitzer et al. 2014). The recombination rate is shown by the blue line. R plot was used to plot the Figure Panels A and B. In Panel C, epigenetic states of 111 Roadmap and 16 immortalized Encode cell lines are shown. They were illustrated in the WashU Epigenome browser. Light red bubbles highlight enhancer marks overlapping with rs6712203 in Adipose Derived Mesenchymal Stem Cells Cultured Cells, Mesenchymal Stem Cell Derived Adipocyte Cultured Cells, Adipose Nuclei and the immortalized ENCODE A549 EtOH 0.02pct Lung Carcinoma Cell line.

rs6712203-T non-risk POU2F2 binding increases *COBLL1* mRNA expression.

We next aimed to identify the likely regulatory variant(s) responsible for the T2D association. We leveraged the cross-species conservation analysis PMCA, a computational approach calculation regulatory probability scores (range 1-9) using conserved transcription factor binding site (TFBS) patterns within cis-regulatory modules (CRMs) (Supplementary Figure 8) (Claussnitzer et al. 2014). PMCA identified rs6712203 as highest scoring variant with a striking clustering of 312 TFBS ($p<0.0001$), 32402 TFBS modules ($p<0.0001$), 763 TFBS in these modules ($p<0.0001$), resulting in the maximal combined overall score of 9 (Claussnitzer et al. 2014). Notably, rs6712203 localizes within an adipose specific enhancer region (Figure 1).

To evaluate the predicted regulatory effect of the variant rs6712203, we first aimed to identify the regulator. We used nuclear protein extracts from mature hAC in EMSA experiments (Figure 2A). Supporting the PMCA analysis, we found a distinct protein binding pattern at rs6712203 alleles, with an increased signal at the rs6712203-T non-risk allele. To identify the differential regulator, we examined disrupted rs6712203 motifs (Kheradpour & Kellis 2014) using PMCA inferences and Haploreg v3 predictions (Ward & Kellis 2012). We predicted differences in the binding affinity of the transcriptional regulator POU2F2 and confirmed the change in affinity using the IGR k-mer model conditioned on POU2F2 binding assayed by ChIP-Seq (Cowper-Sal lari et al. 2012). Using the IGR algorithm, we compared the frequency of k-mers matching the alternative alleles at rs6712203; the baseline scaled difference between the C risk and T non-risk allele was -0.38 (p-value < 10^{-10}) with a higher frequency in POU2F2 peaks across the genome at the rs6712203-T non-risk allele (Figure 2B and Supplementary Figure 10). The rs6712203-T non-risk allele specific POU2F2 binding was supported by competition EMSA experiments (Supplementary Figure 12), where we added an unlabeled POU2F2 oligomer in an increasing concentration to the rs6712203-C risk and T non-risk probes with nuclear protein extracts from mature hAC. With increasing competitor concentration, we found a decreased signal of the differential protein band. These data indicate an increased POU2F2 regulator binding to the rs6712203-T non-risk allele resulting in an increased regulatory activity in adipocytes from non-risk allele carriers.

To predict putative target genes of the rs6712203 variant we examined chromatin interactions between *GRB14/COBLL1* variants and relevant target genes using Hi-C analyses in IMR90 myofibroblasts and NHEK primary normal human epidermal keratinocytes (Supplementary Figure 9) (Claussnitzer et al. 2015; Dixon et al. 2012; Smemo et al. 2014).

We found the topological associated domains (TAD) at the *GRB14/COBLL1* T2D association locus exclusively spanning nearby located genes, including *GRB14* and *COBLL1*. We found no evidence for genomic long-range interactions. Importantly, we observed that rs6712203 localizes within a boundary of two topological domains (Supplementary Figure 9). TAD boundaries tend to be conserved during evolution and to be enriched for active enhancer associated chromatin marks and other important regulatory factors (Dixon et al. 2012). This strengthens our finding that

rs6712203 may be regulatory active and may contribute to the T2D association signal at the *GRB14/COBLL1* locus. Since we could not identify any genomic long-range interaction at the *GRB14/COBLL1* locus, we hypothesized that rs6712203 may affect the expression of the host gene *COBLL1*, which is supported, by the identification of the T2D risk *COBLL1* coding variant Asn939Asp (MAF = 0.12 4.7 x 10^{-11}) (Albrechtsen et al. 2013).

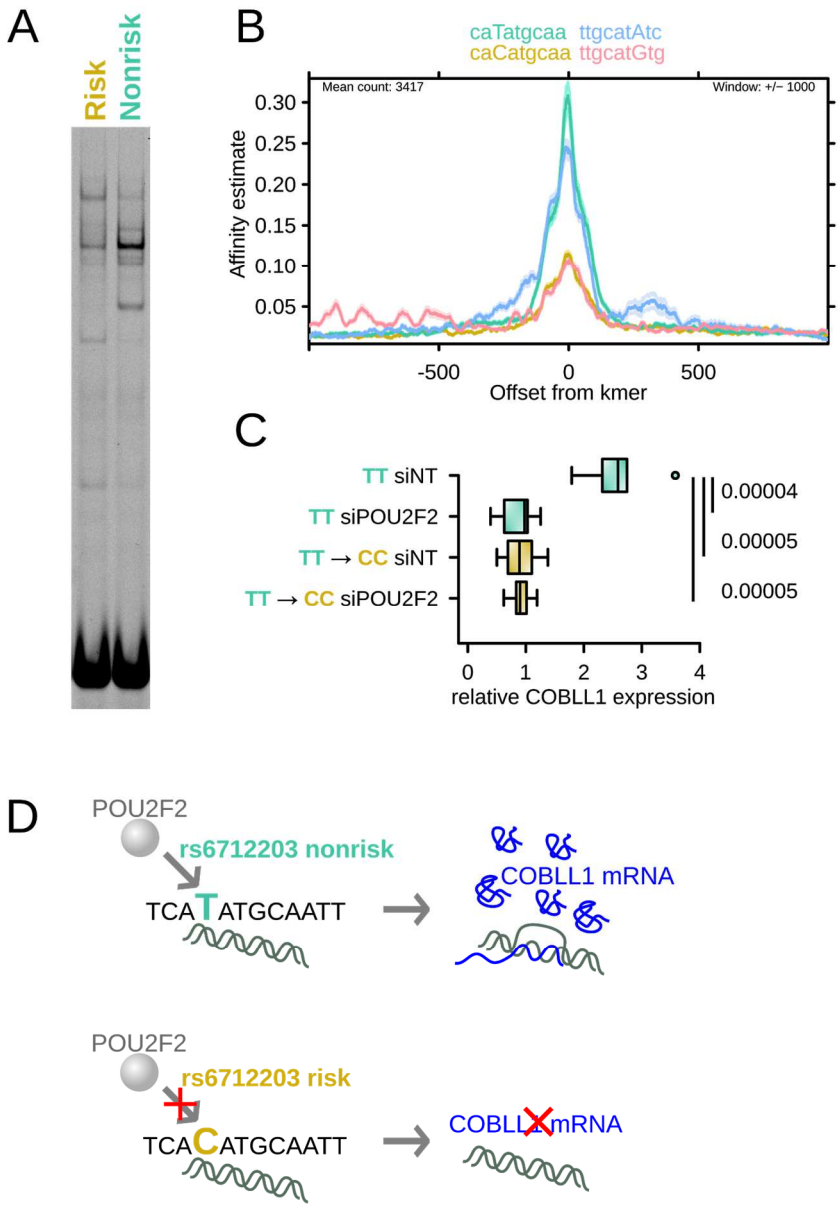


Figure 2: POU2F2 dependent regulation of COBLL1 mRNA expression in rs6712203-T non-risk adipocytes. Panel A shows an EMSA experiment with nuclear extract of primary hAC, with rs6712203- C risk allele in lane 1 and rs6712203-T non-risk allele in lane 2. Panel B shows the IGR 9-mer model estimating the affinity of POU2F2 to the rs6712203- C risk and rs6712203-T non-risk allele. The baseline scaled difference between the alleles is -0.38 (Z test p-value < 10^{-10}) and the context permutation p-value is 0.0034. Panel C shows the expression of COBLL1 mRNA (COBLL1 pair 1) relative to HPRT and dependent on T-to-C editing of rs6712203 in SGBS (n = 5), using siRNA mediated knock-down of the regulator POU2F2 with an average knockdown efficiency of 52 %. Panel D shows the mechanistic model for a POU2F2 dependent regulation of COBLL1 mRNA expression in rs6712203- T non-risk allele carriers, which is disturbed in rs6712203-C risk allele carriers (for an extended mechanistic model see Figure 4).

CRISPR/Cas9 targeted genome editing enables investigating the phenotypic effect of altering a risk variant in its endogenous genomic context independent of other variants in the same haplotype (Hsu et al. 2014). We used CRISPR/Cas9 technology to test the predicted rs6712203 regulatory activity in adipocytes, the differential POU2F2 binding at the rs6712203-T non-risk allele and its effect on downstream target gene expression. We conducted the rs6712203 editing experiment in the human adipose cell line SGBS (Figure 2C) and in primary hASC (Supplementary Figure 11). Starting from the T non-risk haplotype, T-to-C editing reduced the *COBLL1* gene expression 3.2-fold (p-value = 0.00005), to levels observed in C risk allele cells. This difference between T non-risk and C-risk adipose cells was completely abolished when the predicted regulator POU2F2 was reduced by siRNA mediated knockdown, establishing a missing POU2F2 activation as the mechanistic basis of the *COBLL1* expression ablation in rs6712203-risk allele carriers. The experiment revealed a disrupted regulation of *COBLL1* mRNA expression in rs6712203-C risk allele carriers, and a POU2F2 dependent up-regulation of *COBLL1* mRNA expression in rs6712203-T non-risk allele carriers (Figure 2D). These data indicate that the POU2F2 dependent upregulation of *COBLL1* expression in rs6712203-T non-risk allele carriers but not rs6712203-C risk allele carriers may contribute to the T2D signal at the *GRB14/COBLL1* locus.

COBLL1 perturbation decreases actin remodeling during adipogenesis

COBLL1 expression in adipocytes can be linked to actin filamentation.

We next aimed to identify the *COBLL1* downstream mechanisms contributing to T2D, but the molecular functions of *COBLL1* in adipocytes have not yet been conclusively characterized. Therefore, we conducted a PANTHER co-expression analysis from human adipocyte microarray data. We found an enrichment of genes from the integrin (p-value 6.32 x 10^{-8}) and inflammation signaling pathways (Table 2) positively correlated with *COBLL1* mRNA expression (Figure 3A) and genes from the gonadotropin releasing hormone receptor and p38 MAPK pathways negatively correlated (Table 1). The gene with the highest correlation coefficient with *COBLL1* was β1 integrin (β = 0.83; p-value = 0.00004). Interestingly, in data from the gene enrichment profiler (Benita et al. 2010) - a database which uses publicly available microarray data from 142 (16 cancer, 126 normal) different tissues and cells for gene expression analysis - we also found *ADAM12* co-regulated with *COBLL1* (Supplementary Figure 15). *ADAM12* was previously reported to induce actin remodeling during adipogenesis by regulating β1 integrin function (Kawaguchi et al. 2003). Integrins enable outside-in signaling from the extracellular matrix (ECM) into the cell. As such, they are involved in cellular processes like adipocyte proliferation and differentiation (Morandi et al. 2016). It is not yet completely understood how ECM signals are transmitted by integrins into the cells, but F-actin has been shown to play a role in the process (Morandi et al. 2016) and in prostate cancer cells *COBLL1* has been shown to enable actin filamentation (Takayama et al. 2018) which suggests a comparable function in adipocytes. Furthermore, we found *actinin alpha 4 (ACTN4)* and *ARPC2*

co-expressed with *COBLL1. ACTN4* links filamentous actin to membrane proteins and is involved in insulin-stimulated GLUT4 trafficking in muscle cells (Talior-Volodarsky et al. 2008). *ARPC2* is one subunit of the *actin related protein 2/3 complex subunit 2 (Arp2/3)*, which is responsible for the nucleation of branched actin filaments (Weaver et al. 2003) (Figure 3A). Large scale Affinity Capture-MS experiments found an interaction of *COBLL1* with *PACSIN1*, *PACSIN2* and *PACSIN3 (Stark et al. 2006)*. *PACSINs* are highly conserved Src-homology 3 (SH3)-domain-containing proteins, which are involved in cytoskeletal mediated membrane trafficking, through interaction with WH2 containing proteins like the *Arp2/3* complex activator *N-WASP (Kessels & Qualmann 2004)*.

The actin cytoskeleton is involved in a multitude of cellular processes, besides providing structural support and vesicle trafficking. A coordinated regulation of the cytoskeleton is essential during adipogenesis (Yang, Thein, Wang, et al. 2014; Kawaguchi et al. 2003; Kanzaki & Pessin 2001), GLUT4 translocation to the plasma membrane (Kanzaki & Pessin 2001), the maintenance of lipid droplets and lipid droplet dynamics during lipolysis (Orlicky et al. 2013). Taking into account, that the *COBLL1* homolog *COBL* is a known F-actin dynamizer (Husson et al. 2011), and GO-term pathways list *COBLL1* as actin interacting protein (Ashburner et al. 2000), we hypothesized that *COBLL1* may play a role in the regulation of the actin cytoskeleton in adipocytes, thereby affecting GLUT4 vesicle trafficking, lipid metabolism and adipocyte differentiation.

Table 1: Integrin pathway genes correlate with COBLL1. Pearson's correlation coefficient between each measured gene and COBLL1 was calculated and the overrepresentation of pathways was analyzed using PANTHER version 9.0 (released 2014-01-24). The Table shows the number of genes listed per PANTHER Pathway (REFLIST), the number of genes found to be significantly positively correlated with COBLL1 (positive correlation), the expected number of positively correlated genes per pathway (positive correlation (expected)), the over (+)- or under (-)-representation of the number of genes per pathway, the fold enrichment and the resulting p-value. The Integrin pathway reaches the lowest p-value.

PANTHER Pathway	**REFLIST (# of genes 21804)**	**Positive correlation (# of genes 1116)**	**Positive correlation (expected)**	**Over/under representation**	**Fold enrichment**	**p-value**
Unclassified	19446	924	995,31	-	0,93	0
GnRH	228	20	11,67	+	1,71	1
p38 MAPK	41	2	2,1	-	0,95	1
Integrin signaling	175	33	8,96	+	3,68	$6{,}32 \times 10^{-8}$
Inflammation signaling	233	27	11,93	+	2,26	$1{,}68 \times 10^{-2}$

Table 2: Gonadotropin releasing hormone receptor and p38 MAPK pathway genes correlate negatively with COBLL1. Pearson's correlation coefficient between each measured gene and COBLL1 was calculated and the overrepresentation of pathways was analyzed using PANTHER version 9.0 (released 2014-01-24). The Table shows the number of genes listed per PANTHER Pathway (REFLIST), the number of genes found to be significantly negatively correlated with COBLL1 (negative correlation), the expected number of negatively correlated genes per pathway (negative correlation (expected)), the over (+)- or under (-)-representation of the number of genes per pathway, the fold enrichment and the resulting p-value.

PANTHER Pathway	**REFLIST (# of genes 21804)**	**Negative correlation (# of genes 1033)**	**Negative correlation (expected)**	**Over/under representation**	**Fold enrichment**	**p-value**
Unclassified	19446	889	921,29	-	0,96	0
GnRH	228	28	10,8	+	2,59	$1{,}22 \times 10^{-3}$
p38 MAPK	41	10	1,94	+	> 5	$5{,}47 \times 10^{-3}$
Integrin signaling	175	4	8,29	-	0,48	1
Inflammation signaling	233	11	11,04	-	1	1

COBLL1 influences actin remodeling during adipogenesis.

To understand the role of *COBLL1* in actin filamentation, we analyzed the subcellular localization in *in vitro* differentiated hASC using a fluorescent imaging assay. We found that COBLL1 proteins localize in the nuclei and the cytoplasm of adipocytes (Figure 3B). This is consistent with published results of *COBLL1* in U-2 OS, HeLa and A-431 cells (Uhlén et al. 2015; Uhlen et al. 2010), which report COBLL1 localization most abundantly in the nucleus and cytoplasm and to a lesser extend in the plasma membrane and cell junctions (Uhlén et al. 2015). We found that COBLL1 localizes in the cytoplasm in proximity to actin filaments, suggesting a potential role in actin filamentation. Furthermore, the localization of COBLL1 in nuclei is intriguing, since there is increasing evidence that actin and actin nucleation factors are present in the nucleus, where they may be involved in chromatin remodeling and the regulation of gene transcription (Weston et al. 2012).

Cortical actin filaments are found in mature adipocytes, while preadipocytes have stress fibers, which are remodeled into cortical actin during adipogenesis (Yang, Thein, Lim, et al. 2014). An intact and dynamic actin remodeling is also important for the insulin-stimulated glucose uptake into the cells, since both cortical actin disruption as well as stabilization inhibit GLUT4 translocation (Kanzaki & Pessin 2001). The regulation of this process is triggered by insulin action. In adipocytes, insulin is important for the break-down of F-actin stress fibers (Martin, Rose, et al. 1996) and insulin stimulation of adipocytes initially results in cortical actin remodeling followed by an increase in polymerized actin to enable GLUT4 translocation to the plasma membrane (Kanzaki & Pessin 2001).

To investigate the effect of *COBLL1* on F-actin, we conducted fluorescent imaging assays in *in vitro* differentiated hASC with and without a transient ablation of *COBLL1*. We quantified F-actin by measuring Phalloidin (which binds to F-actin) areas and found higher amounts of F-actin in siCOBLL1 cells compared to siNT cells (Supplementary Figure 16B). Furthermore, we found profound changes in the structure and localization of F-actin (Figure 3B and Supplementary Figure 16A), reflected in the missing cortical actin but presence of actin stress fibers in mature adipocytes upon *COBLL1* ablation.

To further understand the effect of *COBLL1* ablation in adipocytes on F-actin remodeling from stress fibers to cortical actin, we performed stimulation experiments with insulin in *in vitro* differentiated hASC. Generally, in cells with siRNA mediated perturbation of *COBLL1*, we found a decreased sensitivity of the cells to insulin with a lower capacity to remodel F-actin. First, we starved the cells for 12 hours from insulin and found a significantly increased amount of F-actin in starved control cells (siNT, p-value = 1×10^{-7}) (Supplementary Figure 16A and B). Next, we added insulin to starved cells and found that the increased amount of F-actin in control cells, was fully remodeled to cortical actin (siNT, p-value = 1×10^{-4}). In contrast, *COBLL1* ablation stalled the response to insulin stimulation. There was no significant difference of F-actin structures or amount in siCOBLL1 cells with or without insulin. In all conditions mainly F-actin stress fibers were visible.

Our data show that actin dynamics during *in vitro* 'starving-refeeding' conditions are perturbed, or at least attenuated upon COBLL1 ablation in adipocytes. Since, dynamic regulations of the actin cytoskeleton are a prerequisite for the cellular translocation of GLUT4 (Kanzaki & Pessin 2001), a decreased regulation of *COBLL1* expression in rs6712203-C risk allele carriers, could result in an inflexibility to remodel F-actin strands, diminish the cellular glucose uptake and lead to cellular insulin resistance. Taken together, our results establish a functional role of *COBLL1* in adipocyte actin remodeling, which may result in a disturbed adipogenesis and GLUT4 vesicle trafficking.

COBLL1 perturbation decreases insulin-stimulated glucose uptake in adipocytes.

Dynamic regulations of the actin cytoskeleton are a prerequisite for the cellular translocation of GLUT4 (Kanzaki & Pessin 2001). We next asked if a decreased regulation of *COBLL1* expression in rs6712203-C risk allele carriers would result in an impaired actin-mediated GLUT4 vesicle trafficking to the plasma membrane (Kanzaki & Pessin 2001). We performed insulin-stimulated glucose uptake assays using radioactive 2-deoxyglucose combined with a stable lentiviral knock down of *COBLL1* in the human adipocyte cell line SGBS (Figure 3C). Control (shEV) cells showed the expected 2.1-fold increased glucose uptake after insulin stimulation (p-value = 5×10^{-6}). Adipocytes with a stable reduction of *COBLL1* (shCOBLL1) had a lower glucose uptake than control cells under basal conditions (p-value = 1×10^{-3}) and there was no significantly increased glucose uptake after insulin stimulation. Our data show that shCOBLL1 adipocytes lose their response to insulin almost entirely as measured by insulin-stimulated glucose uptake. Reduced glucose uptake was accompanied by a decrease in *GLUT4* mRNA between basal and insulin conditions in shCOBLL1 cells (Figure 3I). Furthermore, we show preliminary GLUT4 staining experiments, which suggest less GLUT4 vesicles in siCOBLL1 cells (Supplementary Figure 17). Further experiments will be required to evaluate the link between GLUT4 vesicle trafficking and the effect of *COBLL1* on actin filamentation. In conclusion, we here show that *COBLL1* perturbation in mature adipocytes lead to a reduced glucose uptake and a loss of insulin sensitivity.

COBLL1 perturbation decreases adipogenesis and lipid accumulation.

To assess the relevance of *COBLL1* in adipose tissue, we used the publicly available Gene Enrichment Profiler (Benita et al. 2010) to analyze the relative abundance of *COBLL1* mRNA in different cell and tissue types (Supplementary Figure 15). In comparison to all other tested tissues and cell types, *COBLL1* was most enriched in adipocytes, the placenta and mast cells (203641_s_at). Next, we aimed to determine the expression of *COBLL1* in undifferentiated preadipocytes and mature adipocytes to assess the regulation of *COBLL1* expression during adipocyte differentiation. We measured *COBLL1* mRNA expression in SVF and hAC of subcutaneous adipose tissue samples and found a more than 3-fold decreased expression of

COBLL1 in SVF compared to hAC (p-value < 0.001) (Supplementary Figure 18A). Further, we investigated the expression of *COBLL1* during *in vitro* differentiation of primary human SVF cells and found a significant (p-value = 0.011) increase of *COBLL1* mRNA expression during the time-course experiment between preadipocytes and adipocytes (Supplementary Figure 18B). We can conclude that *COBLL1* is highly expressed in adipose tissue and increases during adipogenesis.

To further investigate the effect of *COBLL1* on adipogenesis and lipid storage, we used stable repression of *COBLL1* in SGBS cells. We differentiated the cells *in vitro* and stained the lipid droplets for quantification. In shCOBLL1 cells we found a lower amount of lipids compared to shEV cells (Figure 3D) and less (p-value = 3×10^{-3}) glycerol-3-phosphate dehydrogenase (GPDH) activity (Figure 3E) - a common marker for adipocyte *in vitro* differentiation. These results provide a link between *COBLL1* expression and adipogenesis with an effect on lipid accumulation. In rs6712203-C risk allele carriers, this could translate in a lower adipogenesis due to a lower ability to up-regulate *COBLL1* mRNA expression. Mature adipocytes have an increased insulin-stimulated glucose uptake when compared to undifferentiated cells (Hauner et al. 1998). Since, not only the maturation of adipocytes, but also their size has been shown to be important for the response of the cells to insulin (Salans et al. 1968), we evaluated the relationship between *COBLL1* mRNA expression and adipocyte size. We used fractions of isolated mature primary human adipocytes (Supplementary Figure 19). In the large fraction we found some evidence for a positive correlation between cell size and *COBLL1* mRNA expression (p-value = 0.01). This finding may be explained by the increased capability of the cells to store lipids, when *COBLL1* is not perturbed. However, to substantiate this result the experiment should be reproduced in a larger cohort.

Leptin is an important factor in the regulation of body fat storage and is primarily produced in adipose tissue (Harris 2014). To assess whether the observed differences in lipid accumulation and adipogenesis between shCOBLL1 and shNT cells affect the expression of this hormone, we measured leptin mRNA levels in control (basal), and insulin/glucose stimulated cells (Figure 3I). As expected from previously published results about fasting humans (Boden 1996), we found a 4.6-fold significantly (p-value = 2×10^{-2}) lower leptin expression level in the control condition, when compared to insulin/glucose stimulated cells. In both conditions, the mRNA expression of leptin was slightly lower in shCOBLL1 cells, which is in concordance with the assumption that leptin levels correspond to the size of the body lipid storage (Harris 2014). To assess whether this observation was unique to our experimental setup with stably reduced *COBLL1* levels, we also assessed the correlation between *COBLL1* and leptin in primary isolated subcutaneous hAC. We found a significant (p-value = 0.00005) positive correlation (r = 0.74) (Figure 3H), confirming the previous result in shCOBLL1 cells.

In conclusion, we found evidence for higher *COBLL1* levels in mature and large adipocytes, confirming the previous observation of a diminished maturation capacity and lipid storage when *COBLL1* is reduced. The finding that leptin mRNA expression levels, an adipokine produced in

proportion to the size of fat depots (Harris 2014), correlate with *COBLL1* mRNA expression further supports the role of *COBLL1* in fat storage.

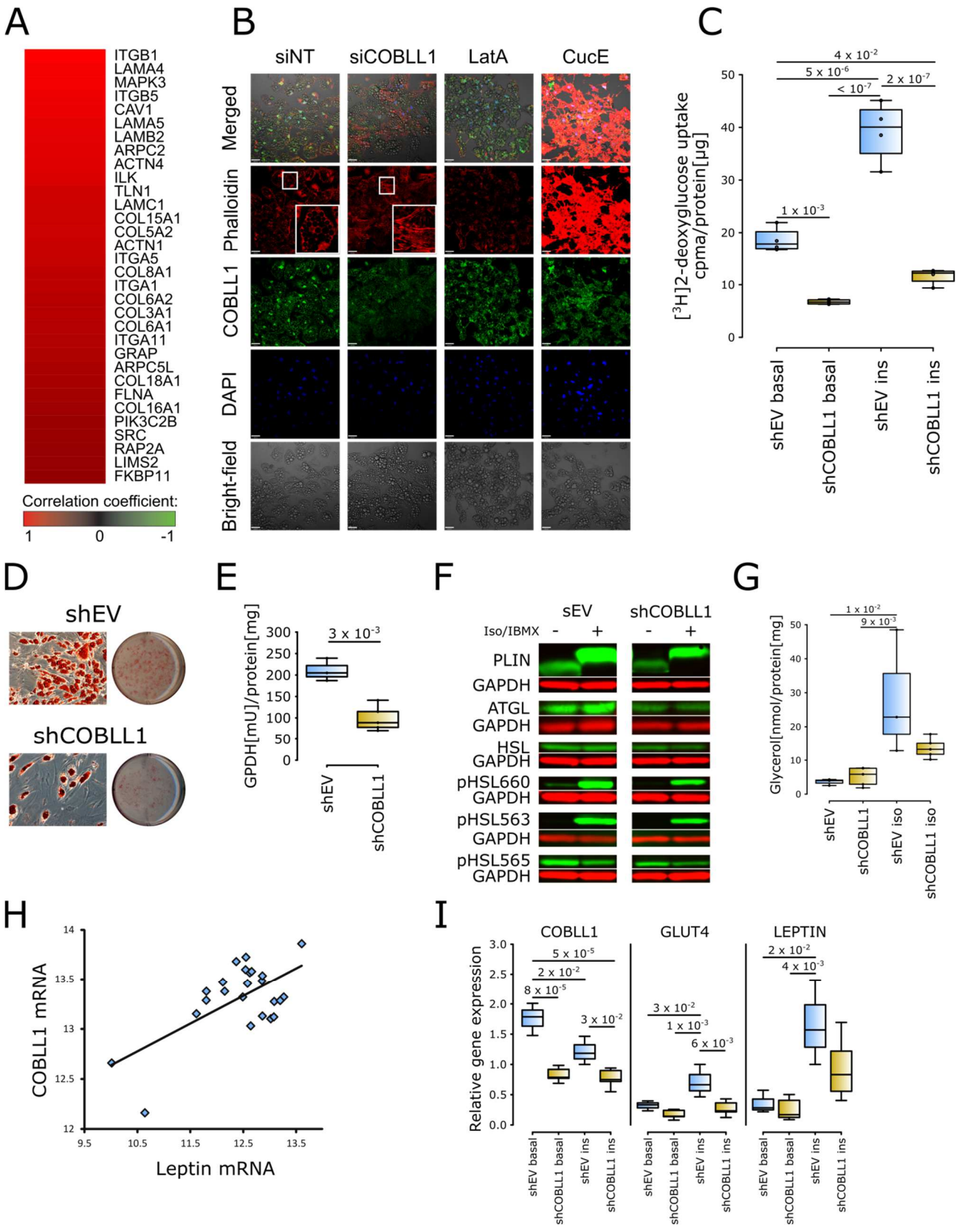


Figure 3: COBLL1 can be linked to actin remodeling during adipogenesis, insulin-stimulated glucose uptake, lipid accumulation and lipolysis in adipocytes. Panel A shows the positive correlation of COBLL1 mRNA expression (COBLL1 probe ILMN_1761260) with the expression of integrin pathway genes. Pearson's correlation coefficient between COBLL1 and each gene was calculated using microarray data from primary hAC. Gene enrichment was tested using the PANTHER pathway analysis. The top-scoring pathway was the positively correlated integrin

pathway (p-value = 6.32 x 10^{-8}). All genes listed in this pathway and significantly correlated with the expression of COBLL1 are shown in the heatmap, where red indicates a positive correlation and black no correlation. Panel B shows fully differentiated primary hAC incubated with COBLL1 specific siRNA (siCOBLL1) or non-targeting siRNA (siNT). The knock down efficiency of COBLL1 was in average 37 %. The chemical compounds Latrunculin A (LatA) and Cucurbitacin E (CucE) were used as controls for actin filamentation. Actin filaments were stained red with fluorescent Phalloidin against filamentous actin, COBLL1 was stained with specific first antibody in combination with green fluorescent secondary antibody and nuclei were stained blue using DAPI. Merged pictures include bright field, filamentous actin, COBLL1 and nuclei. Panel C shows insulin-stimulated glucose uptake in shCOBLL1 and shEV SGBS adipocytes, given as ratio between radioactive cpma counts and amount of protein in µg. The protein concentrations were obtained from simultaneously cultivated cells. Panel D, shows Oil-red O stained fully differentiated shCOBLL1 and shNT SGBS adipocytes. Panel E shows Glycerol-3-phosphate dehydrogenase (GPDH) enzyme activity measures, normalized to the protein content of each well. P-values of 3 independent experiments were calculated using Welch t-test. Panel F shows western blot experiments with shEV and shCOBLL1 SGBS adipocytes, with or without Isoproterenol/IBMX (iso) incubation. Equal protein loading was verified in each condition using antibodies against GAPDH. Panel K shows the amount of glycerol per protein with or without Isoproterenol/IBMX (iso) incubated shEV and shCOBLL1 SGBS adipocytes. Panel H shows correlation between the expression of COBLL1 ILMN_1761260 (Illumina microarray probe) and leptin (LEP) ILMN_2207504 expression (n = 24). Panel I shows relative mRNA expression of COBLL1 (COBLL1 pair 1), GLUT4 and Leptin in fully differentiated SGBS cells in cells with a stable repression of COBLL1 (shCOBLL1) or control empty vector (shEV) cells, using a lentiviral system. P-values for experiments with more than two groups were calculated in R using Anova with TukeyHSD.

COBLL1 perturbation decreases stimulated lipolysis.

We have established a role of *COBLL1* in actin remodeling, glucose uptake, adipocyte maturation and lipid storage. Because it has been suggested, that the cytoskeleton may play a role during lipolysis dynamics (Orlicky et al. 2013), we hypothesized that the break-down of triglycerides and release of FA and glycerol could be affected by *COBLL1* perturbations. To test our hypothesis, we conducted a glycerol release experiment in cells with a stable reduction of *COBLL1* (shCOBLL1). In control cells (shEV) with a basal lipolysis rate, we investigated the expected low levels of released glycerol (Figure 3G). After stimulation of lipolysis in shEV cells, using the β-AR agonist isoproterenol and phosphodiesterase (PDE) Inhibitor IBMX (iso/IBMX), the glycerol levels were 7.8-fold increased (p-value = 1×10^{-2}). In shCOBLL1 the β-adrenergic lipolysis stimulation was only 2.6-fold increased (p-value = not significant) (Figure 3G), suggesting that *COBLL1* expression is important for an adequate response to lipolysis stimulation. Lipolysis stimulation has previously been linked to obesity and related complications like insulin resistance. It has been shown, that weight reduction results in a decreased basal lipolysis but increased sensitivity to stimulation by catecholamines (Reynisdottir et al. 1995) and depot-specific differences between subcutaneous and visceral adipose tissue have been shown as possible contributor for an unfavorable fat distribution (Spalding et al. 2017). Our data suggest that COBLL1 ablation - potentially through a dysregulation of the cytoskeleton – affects the ability of the cells to respond to β-adrenergic stimulation, reflected by a perturbation of lipolytic activity. We next asked if the effect described above manifests at the protein level of enzymes and proteins involved in lipolytic activity using western blot experiments. We focused on lipolytic enzymes ATGL, hormone sensitive lipase (HSL), phosphorylated HSL (pHSL660, pHSL563, pHSL565) and on the lipid droplet associated protein perilipin (PLIN). We analyzed three different HSL phosphorylation sites: the PKA phosphorylation sites Ser-660 and Ser-563, which are phosphorylated upon isoproterenol or noradrenaline stimulation (Anthonsen et al. 1998) and Ser-565 which is phosphorylated under basal conditions and has an antilipolytic role due to its negative interaction with pHSL563 (Garton et al. 1989). Western blot experiments (Figure 3F) and their semi-quantification (Supplementary Table 3) showed an overall decreased protein level for PLIN, ATGL and HSL under basal and stimulated conditions in shCOBLL1 adipocytes and decreased pHSL563 and pHSL660 upon stimulation in shCOBLL1 cells which confirms the decreased response of shCOBLL1 cells to lipolytic stimulation.

Given the impact of rs6712203 on *COBLL1* gene expression (3.2-fold decrease, see Figure 2) we asked if the variant rs6712203 affects the response of adipocytes to lipolytic stimulation. We were interested to see, whether the perturbed adaptation to lipolytic stimulation would be limited to the effects of *COBLL1* or whether they would also be observed on rs6712203 genotype level. Therefore, we leveraged data from isolated primary human adipocytes genotyped for rs6712203 where several phenotypes including basal and stimulated lipolysis in response to the β-AR agonist's isoproterenol, noradrenaline and the adenylate cyclase - PKA pathway activator forskolin, were assessed. While we did not observe an effect on the basal lipolysis rate, we found a

decreased stimulated lipolysis rate in adipocytes from female rs6712203-C risk compared to rs6712203-T non-risk allele carriers. More specifically, in adipocytes from female rs6712203-C allele carriers, we found a consistently decreased lower glycerol release following the treatment with isoproterenol (β = -2.2, p-value = 0.014), noradrenalin (β = -1.3, p-value = 0.047) or forskolin (β = -2.9, q-value = 0.002) (Table 3). After adjusting for multiple testing, only forskolin was significantly different. Interestingly, we found no significant difference in males. These results substantiate the effect of *GRB14/COBLL1* variants - possibly rs6712203 - likely via *COBLL1* gene regulation on the response of adipocytes to lipolytic stimulation. Furthermore, we here identified a sexual dimorphism for stimulated lipolysis between rs6712203-C risk and T non-risk allele carriers.

Table 3: Stimulated lipolysis reveals sexual dimorphism between rs6712203-C risk and T non-risk allele carriers. Associations were performed using a linear model adjusted for BMI and age in 505 women and 176 men. Effects are modeled on the rs6712203 risk-C allele and lipolysis measures are given per cell.

Lipolysis/Cell	**β ± se**	**p-value**	**q-value**
Isoproterenol female	- 2.2 ± 0.9	0.014	0.08
Noradrenalin female	- 1.3 ± 0.7	0.047	0.20
Forskolin female	- 2.9 ± 0.9	0.002	0.02 *
Basal female	- 0.5 ± 0.4	0.2	0.3
Isoproterenol male	0.7 ± 1.5	0.6	0.9
Noradrenalin male	- 0.5 ± 1.2	0.7	0.9
Forskolin male	0.3 ± 1.6	0.8	0.9
Basal male	- 0.2 ± 0.6	0.7	0.9

rs6712203-C risk alleles coincide with a decreased hip circumference in females.

The *GRB14/COBLL1* locus is a pleiotropic locus associated with a multitude of metabolic phenotypes including T2D and WHR (Heid et al. 2010). We tested the relationship between rs6712203 and hip- and waist circumference in a cohort of 681 volunteers. After adjusting for age and BMI, we found a decreased hip circumference in female rs6712203-C risk allele carriers β = -1.1 (q value = 0.04) (Table 4). We did not observe any effect on waist circumference. We also did not observe an association with WHR, waist or hip circumference in men carriers (Table 4). Publicly available data from the GIANT data (Randall et al. 2013) confirmed these observations. They showed a significant association for the variant rs6712203 with BMI adjusted hip circumference (p-value = 3.51 x 10^{-6}) (Supplementary Figure 23A). In summary, we find some

evidence for a female specifically decreased BMI adjusted hip circumference in rs6712203-C risk allele carriers.

Table 4: Lower BMI adjusted hip circumference in female rs6712203-C risk allele carriers compared to female rs6712203-T non-risk allele carriers. Associations were performed using a linear model adjusted for BMI and age in 505 women and 176 men. Effects are modeled on the rs6712203-C risk allele.

Phenotypes	**β ± se**	**p-value**	**q value**
Hip circumference women	- 1.1 ± 0.4	0.006	0.04 *
Waist circumference women	- 0.5 ± 0.5	0.4	0.3
Waist-Hip ratio women	0.005 ± 0.005	0.3	0.3
Hip circumference men	0.4 ± 0.6	0.5	0.9
Waist circumference men	0.5 ± 0.7	0.5	0.9
Waist-Hip ratio men	- 0.0006 ± 0.006	0.9	0.9

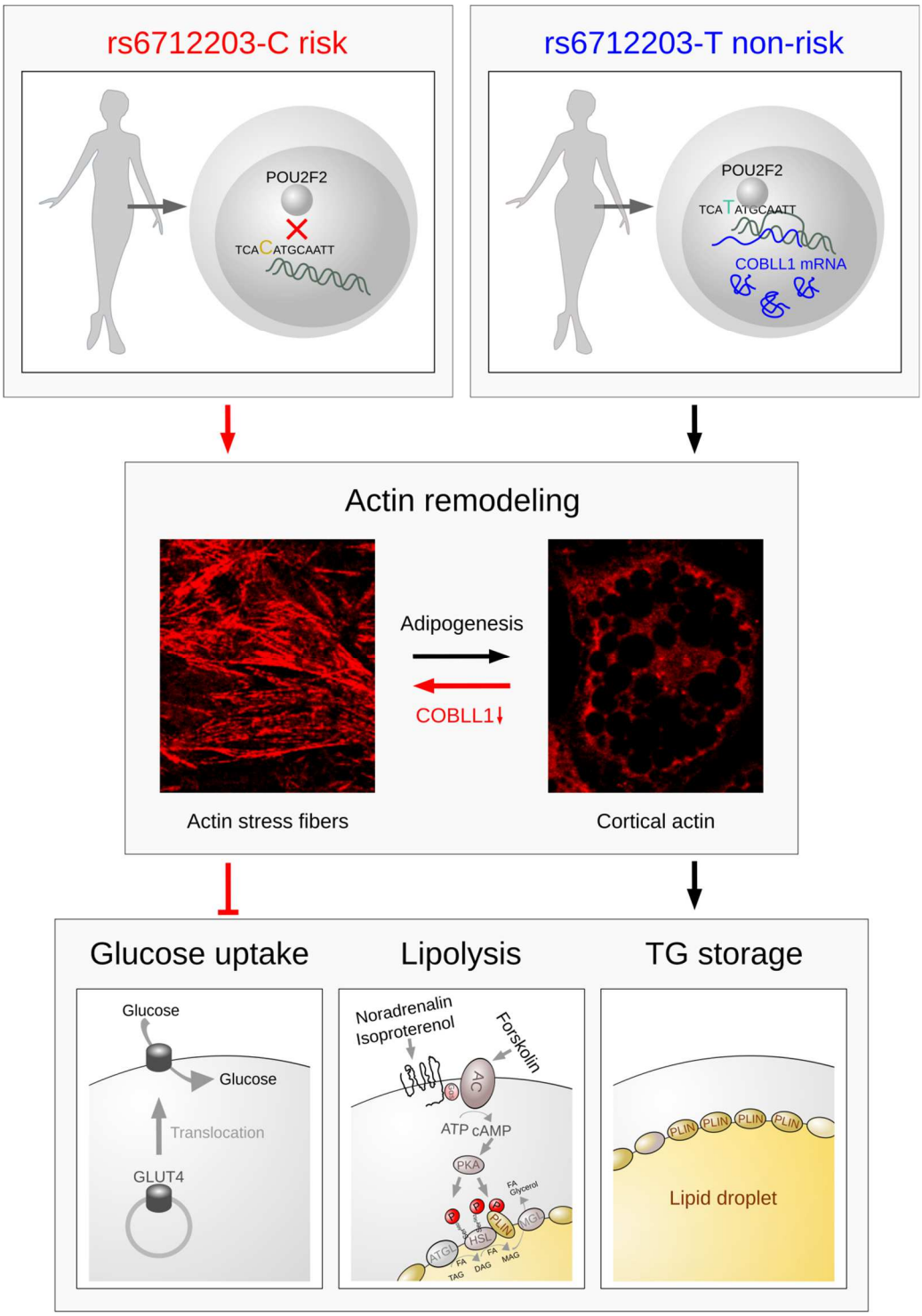


Figure 4: Mechanistic model detailing the POU2F2 dependent up-regulation of COBLL1 expression in rs6712203-T non-risk allele carriers. In rs6712203-T non-risk human adipocytes, the expression of COBLL1 is POU2F2 dependent upregulated. In rs6712203-C risk adipocytes, the POU2F2 motif is partially disrupted, which prevents the up-regulation of COBLL1. COBLL1 perturbation in human adipocytes, leads to a disturbed remodeling of F-actin fibers from stress fibers to cortical actin, which is essential for adipogenesis and for the adaptation to changes in the microenvironment of the cell. We propose that a disrupted regulation of COBLL1 in rs6712203-risk adipocytes may lead to lower cortical actin structures, which can cause disturbances during adipogenesis and result in a decreased insulin-stimulated glucose uptake, lipolysis and TG storage. A disturbed energy regulation in adipocytes and resulting challenges in energy storage in the human body are known risk factors for T2D and could contribute to the GWAS risk signal for T2D at the GRB14/COBLL1 locus.

Discussion

In this work we use our computational and experimental model, detailing the POU2F2 dependent up-regulation of *COBLL1* mRNA in rs6712203-T non-risk adipocytes (Figure 4) possibly contributing to the *GRB14/COBLL1* GWAS risk association with insulin resistance and T2D (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. 2014; Kooner et al. 2011; Morris et al. 2012). Our work provides evidence that *COBLL1* is involved in adipocyte actin remodeling, maturation and metabolism, which has not been reported before. Although GRB14, an adaptor protein for the insulin receptor (Scharf et al. 2004) has been suggested as potentially causal for the T2D risk association at *GRB14/COBLL1* (Schleinitz et al. 2014), the regulatory circuitry underlying the genome-wide disease association has not previously been identified. To unravel the regulatory variant, its regulator and the affected gene at the *GRB14/COBLL1* locus, we integrated publicly available data with specifically designed targeted perturbations in human adipocytes.

We identified an adipocyte-specific enhancer region surrounding the intronic *COBLL1* variant rs6712203, its upstream regulator POU2F2 and the up-regulation of the target gene *COBLL1* in rs6712203-T non-risk allele carriers. Until now, the function of the COBLL1 protein in adipocytes has not been described. Following *COBLL1* perturbation, we found a lower ability to remodel F-actin stress fibers into cortical actin, a decreased adipogenesis, insulin-stimulated glucose uptake, triglyceride storage and a decreased stimulated lipolysis. The observed cellular phenotypes are relevant to type 2 diabetes and consistent with the association of the *GRB14/COBLL1* locus with type 2 diabetes across populations that could result from a peripheral insulin resistance with a dysfunctional energy regulation in adipose tissue.

The non-coding *GRB14/COBLL1* locus

Non-coding variants at the *GRB14/COBLL1* locus were repeatedly reported by GWAS, to be associated with the risk for T2D. Across several populations, different tagging variants, including rs3923113 (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. 2014; Kooner et al. 2011) and rs13389219 (Morris et al. 2012) were published. However, the causal variant(s) have not yet been identified. It has been shown, that association signals for complex traits tend to be located across the genome, often near genes without an obvious connection to the disease itself (Boyle et al. 2017). In order to identify the cause for GWAS signals, it is therefore necessary to identify the regulatory circuitry, starting with the causal variant in the relevant tissue (Claussnitzer et al. 2015). Tissue specificity has been described to arise from gene regulatory networks which control common and tissue specific processes in human cells, by regulating the expression of genes (Sonawane et al. 2017). The identification of the relevant tissues for GWAS

signals are detrimental because it is not only necessary to understand the regulatory circuitry leading to a disease but can also influence the effectivity of a drug.

To identify the relevant tissue for the *GRB14/COBLL1* GWAS signal, we leveraged chromatin state information, from the NIH Roadmap Epigenomics Consortium, which is up-to date the largest collection of human epigenomes for primary cells and tissues (Roadmap Epigenomics Consortium et al. 2015). This data has been shown to provide information about the enrichment of disease-associated genetic variants in tissue-specific epigenomic marks and about biologically relevant cell types (Roadmap Epigenomics Consortium et al. 2015). For the *GRB14/COBLL1* haplotype and specifically for the intronic *COBLL1* variant rs6712203, we found enhancer states in adipose tissue cultured or derived cells. Since variants at the *GRB14/COBLL1* have also been associated with increased adiposity, body fat percentage (Lu et al. 2016) and WHR (Heid et al. 2010), regional differences in fat storage may lead to insulin resistance and ultimately contribute to T2D. It is therefore likely that adipose tissue is relevant for the T2D association at the *GRB14/COBLL1* locus.

The variant rs6712203 was found by the cross-species conservation analysis PMCA (Claussnitzer et al. 2014) to be located in a region with conserved transcription factor binding site (TFBS) patterns within cis-regulatory modules (CRMs), having the highest likelihood in the *GRB14/COBLL1* locus to be a regulatory variant contributing to the GWAS association with T2D. We also found the variant located in a boundary region between two topological domains separating the exons of the *COBLL1* gene. Human chromosomes have been shown to be segmented into megabase-sized topological domains, which are ubiquitous within the genome, stable between cell types and conserved throughout the mammalian evolution (Dixon et al. 2012). Within these topological domains loci tend to interact with each other, enabling promotor-enhancer interactions and other gene-regulatory features (Morris et al. 2012; Lajoie et al. 2015). The boundaries between topological regions have been suggested, to be of importance for the organization of heterochromatin, and as such may link topological domains with their transcriptional control (Dixon et al. 2012). We therefore focused in this work on characterizing the regulatory activity of rs6712203 and the regulatory circuitry potentially contributing to the T2D association signal at the *GRB14/COBLL1* locus.

The pou-domain homeobox transcription factor POU2F2

To identify the regulatory circuitry at rs6712203, we performed a motif analysis and identified a rs6712203-T non-risk allele specific POU2F2 binding site. We conducted CRISPR/Cas9 experiments to validate our regulatory model and found an increased level of *COBLL1* in rs6712203-T non-risk allele carriers, which was completely abolished in rs6712203-C risk allele carriers and in rs6712203-T non-risk allele carriers with a perturbed *POU2F2* expression. The

transcriptional regulator *POU2F2 (Oct2)* is encoded in humans on chromosome 19 and belongs to the pou-domain homeobox transcription factors (Clerc et al. 1988). These transcription factors possess two domains with the ability to bind to DNA, the amino-terminal POU-domain, and the carboxy-terminal POU specific homeodomain (Sivaraja et al. 1994). Both contain a helix-turn-helix motif, common in DNA binding domains. They are joined by a variable linker, which can vary between 15 and 56 amino acids. The linker allows the functional domains different arrangements on the DNA and the interaction with a plethora of regulatory proteins (Herr & Cleary 1995).

POU specific homeodomains are distantly related to the classical homeodomains (Zhao 2013), which are 180 bp long (Phillips & Luisi 2000), were first described in 1984 from Drosophila (McGinnis et al. 1984) and can be found in a wide range of DNA-binding proteins (Herr & Cleary 1995). Homeodomain containing transcription factors are known to coordinate the development of organisms (Banerjee-Basu 2001), by regulating the axial patterning, the identity of segments, the differentiation (Cowherd et al. 1997) and the proliferation of cells (Gehring 1994). Interestingly, an overrepresentation of conserved homeodomain transcription factor binding sites has previously been identified at genetic risk loci for T2D (Claussnitzer et al. 2014), suggesting an important impact of these developmental factors in the onset of the disease.

POU2F2 was initially believed to be only expressed in B-cells (Clerc et al. 1988) and neuronal cells (Latchman 1996). Later it was found to be present in multiple tissues (Hatzopoulos et al. 1990), including subcutaneous adipocytes (Carithers et al. 2015) and it may even be ubiquitous (Zhao 2013). Nevertheless, the expression of *POU2F2* may not be the most relevant parameter in determining the regulatory circuitry at the *GRB14/COBLL1* locus. The recent work of Sonawane and colleagues (Sonawane et al. 2017), focused on tissue-specific gene regulations. They suggest that not the expression pattern of transcription factors but rather their targeting patterns and resulting regulatory networks are responsible for the regulation of tissue-specific biological processes (Sonawane et al. 2017). *POU2F2* has one activation domain in the C-terminus, a second activation and one repressor domain in the N-terminus (Zhao 2013). Multiple alternatively spliced isoforms have been identified for *POU2F2*. Three of them are termed Oct2.1, Oct2.4 and Oct2.5. The isoform Oct2.1 is predominantly expressed in B lymphocytes and the mammary gland. It includes the C-terminal activation domain and has an overall stimulatory effect on the transcription of its target genes. The isoforms Oct2.4 and Oct2.5 are expressed in neuronal cells. They lack the C-terminal activation domain and repressive transcription (Zhao 2013). Besides the octamer sequence ATGCAAAT, POU2F2 is able to recognize other related promotor and enhancer associated DNA sequences (Zhao 2013), including the rs6712203-T non-risk allele sequence. POU2F2 and the closely related family member POU2F1 can bind to the DNA as monomers, homo- or heterodimers and both transcription factors confer a high versatility in their interaction with heterologous non-POU proteins. Among others, POU2F2 is able to associate with the TATA box-binding protein TBP (Herr & Cleary 1995) through its POU domain which can activate gene

transcription (Zhao 2013). The genomic region surrounding the variant rs6712203 does also confer a T non-risk allele specific TATA motif (Ward & Kellis 2016), which could contribute to the effect we observe on *COBLL1* expression. Interestingly, progesterone or a combination of lactogenic hormones (insulin, hydrocortisone and prolactin) (Dong & Zhao 2007) with estradiol (Zhao et al. 2002) have been observed to induce POU2F2 octamer binding in the mammary gland, which can activate the expression of milk genes (Zhao 2013). This highlights not only a responsiveness of the transcription factor to insulin but also to progesterone and estradiol. If these hormones also influence the regulation of POU2F2 in adipocytes, it could contribute to the sexual dimorphism observed at the *GRB14/COBLL1* locus (Morris et al. 2012; Heid et al. 2010). Taken together, the POU family has several members with a complex effect on gene regulation affecting disease. We here found a role of POU2F2 on *COBLL1* expression in rs6712203-T non-risk allele carriers, potentially contributing to T2D.

COBLL1 and actin filamentation in adipocytes

Since we have identified a POU2F2 dependent rs6712203-T non-risk allele specific up-regulation of *COBLL1* in adipocytes, we investigated the effects of the target gene *COBLL1* in adipose tissue. *COBLL1* has previously been identified to modulate cell morphology and enable actin filamentation in prostate cancer cells (Takayama et al. 2018) but the functional role of *COBLL1* in T2D has not been revealed yet. We describe for the first time a role for *COBLL1* in adipocyte filamentation, potentially influencing adipose tissue function. The role of *COBL* - the most important paralog of *COBLL1* - in actin nucleation has been identified much earlier (Husson et al. 2011; Jiao et al. 2014). *COBLL1* and *COBL* are paralogs (Husson et al. 2011) which share closely related DNA sequences. *COBL* possesses two repeated lysine, arginine, and proline-rich regions (KKRRAP motifs) (Carroll et al. 2003; UniProt Consortium 2015), a K region and three consecutive WH2 domains. *COBLL1* has also a KKRRAP motif but only one WH2 domain.

KKRAP motifs have been shown to enable protein–protein interactions (Carroll et al. 2003). *COBL* (Schwintzer et al. 2011) and *COBLL1* (Sowa et al. 2009) have been found in complexes with PACSIN's, which may be mediated by the KKRAP motifs. PACSIN's are involved in actin filamentation by interaction with N-WASP, an activator of Arp2/3 (Kessels & Qualmann 2004; Schwintzer et al. 2011b). Arp2/3 is a protein complex responsible for the nucleation of branched actin filaments (Weaver et al. 2003).

The WH2 domain is known to interact with actin; it is present in many proteins able to modify the cytoskeleton (Husson et al. 2011). *COBL* is able to nucleate and sever actin filaments, enabling the dynamic regulation of the actin cytoskeleton (Husson et al. 2011; Jiao et al. 2014). Studies investigating the function of *COBL*, found that one or two WH2 domains, in combination with the K region are sufficient for the nucleation function of *COBL* (Husson et al. 2011), while at least two

WH2 domains are required for the severing function. When the K region is not present, the *COBL* protein is not able to nucleate or sever actin filaments efficiently. But, when at least one WH2 domain is available it still inhibits the F- actin pointed-end growth (Husson et al. 2011), due to the ability of the *COBL* WH2 domain, to form a complex with monomeric actin (Husson et al. 2011). In *COBLL1*, only one WH2 domain and no domain like the K-region are known. This implies that *COBLL1* has no F-actin nucleation function but may influence actin filamentation by interacting with monomeric actin (G-actin). Actin dynamics have been shown to be essential for many processes in the development and metabolism of adipocytes (Schwintzer et al. 2011; Smas & Sul 1995; Yang, Thein, Wang, et al. 2014), including adipogenesis (Nobusue et al. 2014; Yang, Thein, Wang, et al. 2014), intracellular vesicle trafficking (Mundy et al. 2002; Kanzaki & Pessin 2001) like GLUT4 translocation (Kanzaki & Pessin 2001), lipid storage (Orlicky et al. 2013; Smas & Sul 1995) and lipolytic processes (Pellegrinelli et al. 2014). Therefore, we characterized the function of *COBLL1* on the development and metabolism of adipocytes and we found an overall diminished sensitivity of *COBLL1* perturbed cells to extracellular influences.

COBLL1 and adipocyte maturation

In *in vitro* differentiated hASC, we found an increased amount of actin stress fibers, when *COBLL1* was transiently repressed, implicating *COBLL1* in the remodeling of adipocyte actin structures. Generally, changes in the morphology of flat fibroblast-like cells to increasingly round shaped cells are taking place during adipogenesis (Kawaguchi et al. 2003). The remodeling of F-actin from stress fibers to cortical actin structures is essential for this process (Yang, Thein, Wang, et al. 2014; Kanzaki & Pessin 2001). Latrunculin A, a chemical compound decreasing actin filamentation and increasing cellular G-actin levels, has previously been shown to be sufficient to induce adipocyte differentiation even when adipogenic stimuli, like IBMX or dexamethasone are absent (Nobusue et al. 2014). G-actin may promote adipogenesis by forming a complex with the transcriptional coactivator and PPARG antagonist MKL1 (megakaryoblastic leukemia 1). When MKL1 is bound to G-actin, its translocation to the nucleus is prevented and PPARG expression is increased (Nobusue et al. 2014). But, not only the availability of G-actin has been found to be necessary to promote adipogenesis, also the reorganization of F-actin stress fibers into cortical actin structures need to take place (Yang, Thein, Lim, et al. 2014). When the nucleation complex Arp2/3 is dysfunctional, then the maturation of adipocytes is impaired and increased amounts of actin stress fibers are visible (Yang, Thein, Lim, et al. 2014). Correspondingly, we identified a decreased adipogenesis rate in *in vitro* differentiated *COBLL1* perturbed preadipocytes with an increased amount of actin stress fibers. These results highlight the importance of *COBLL1* during preadipocyte actin stress fiber remodeling to cortical actin, which may influence adipocyte maturation. In summary, the mechanism by which *COBLL1* facilitates F-actin remodeling remains

to be investigated, but our results imply that *COBLL1* may be able to interact with G-actin through its WH2 domain which may promote actin filamentation.

COBLL1 and insulin response in adipocytes

We established a role of *COBLL1* gene expression in adipocyte actin remodeling and the regulation of adipogenesis. Since an essential part of the human body's energy homeostasis is maintained by adipocytes and tightly regulated by chemical, enzymatic and mechanical influences (Kanzaki 2006), we investigated the effect of *COBLL1* on adipose tissue energy uptake.

One of the most important regulators of the human body's energy homeostasis is the insulin receptor, the key trigger of several intracellular pathways, which together orchestrate the insulin response (Kanzaki 2006). They include the phosphatidylinositol 3-kinase pathway, and the Rho family small GTP binding protein TC10 pathway. Both have been proposed to be involved in the regulation of the actin cytoskeleton (Kanzaki et al. 2002; Martin, Haruta, et al. 1996). In Rat-1 fibroblasts overexpressing the insulin receptor, insulin has been shown to stimulate actin rearrangements, with an initial breakdown of stress fibers (Martin, Rose, et al. 1996). Concordantly, we found an increased amount of F-actin in insulin depleted hASC, which was reversible to control F-actin levels when they were incubated with insulin, showing that the effect of insulin starvation on actin filamentation was reversible. In contrast to these results, we found a diminished sensitivity to insulin stimulated actin filamentation, in transiently *COBLL1* repressed hASC, which implies that the effect of *COBLL1* on actin filamentation may be involved in the response of adipocytes to insulin.

To test, whether perturbed *COBLL1* adipocytes have a decreased glucose uptake, we used stable *COBLL1* repressed SGBS adipocytes. Confirming our previous results, we observed a decreased insulin-stimulated glucose uptake in *COBLL1* repressed SGBS adipocytes. The formation of cortical actin by the Arp2/3 complex is essential for an efficient GLUT4 translocation (Yang, Thein, Lim, et al. 2014). Since *COBLL1* is involved in the adipocyte actin remodeling, we aimed to investigate whether a decreased GLUT4 translocation may be causal for the observed decreased glucose uptake. In adipocytes, GLUT4 is primarily stored in intracellular compartments, comprising the trans-Golgi-network and recycling endosomes. When the cells are not stimulated with insulin only a very small amount of about 2 – 5 % GLUT4 can be found at the plasma membrane (Bogan et al. 2003). Efficient endocytosis processes maintain this distribution pattern in the basal state. When the cells are stimulated with insulin, the amount of GLUT4 in the plasma membrane increases, by a more efficient vesicle trafficking. This results in a higher exocytosis rate, exceeding the basal endocytosis rate (Kanzaki 2006). Chemical inhibition as well as stabilization of F-actin fibers dose dependently inhibit the translocation of GLUT4 after insulin stimulation in adipocytes (Kanzaki 2006; Kanzaki & Pessin 2001). It is assumed, that the translocation of GLUT4 requires an

active cortical actin remodeling. Time lapse experiments, have revealed that the remodeling of cortical actin in adipocytes initially result in increased peri-nuclear F-actin (Kanzaki & Pessin 2001) and thereby potentially facilitate the (Kanzaki & Pessin 2001) insulin mediated exocytosis from the intracellular storage to the plasma membrane. When the effect of insulin decreases, GLUT4 is recycled by endocytosis and again stored in the intracellular storage compartments (Kanzaki 2006). We measured the expression of *GLUT4* in SGBS with a stable repression of *COBLL1* and found a slightly decreased *GLUT4* expression and a lower adaptation to insulin when compared to control cells. *GLUT4* expression has been shown in mouse models to be an important factor for insulin sensitivity. Mice overexpressing Glut4 in adipocytes, have a slightly increased body weight but more importantly an improved insulin sensitivity (Gnudi et al. 1995), whereas mice with an adipocyte specific Glut4 knock-out show higher insulin resistance (Abel et al. 2001).

In addition, we stained GLUT4 in *in vitro* differentiated human adipocytes with a transient *COBLL1* repression and found a decreased amount of GLUT4 protein and some evidence for a lower GLUT4 translocation rate to the plasma membrane in transient *COBLL1* repressed cells. However, we were not able to sufficiently quantify the translocation of GLUT4. Moreover, we have previously shown that adipocytes with a stable repression of *COBLL1* show a lower maturation rate, which has been shown to have an effect on insulin sensitivity and respectively also glucose uptake (Hauner et al. 1998). We are therefore not able to conclude from these results whether the decreased insulin-stimulated glucose uptake in *COBLL1* repressed adipocytes is caused by lower adipogenesis, *GLUT4* expression or translocation.

In conclusion, we here identified a decreased sensitivity of *COBLL1* perturbed cells to insulin which leads to a decreased glucose uptake, potentially caused by the effect of *COBLL1* on actin filamentation.

COBLL1 and adipogenic capacity

We established a role of *COBLL1* in adipocyte actin remodeling potentially influencing adipogenesis and the amount of lipids in the cells. The actin cytoskeleton and the ability of cells to remodel actin filaments have been implicated in the storage of lipids (Orlicky et al. 2013; Smas & Sul 1995) which implies that the role of *COBLL1* in actin remodeling could also be important for lipid storage in differentiated adipocytes. In mature unilocular white adipocytes energy is mainly stored in one lipid droplet (Greenberg et al. 2011), which occupies most of the adipocyte, controls the volume (Karpe et al. 2011) and leads to a rounded cell shape (Orlicky et al. 2013; Smas & Sul 1995). Adipose tissue is the primary storage organ of the human body, when surplus energy needs to be stored, adipocytes are able to massively expand in number – hyperplasia – or size – hypertrophy – to prevent ectopic fat accumulation (Karpe et al. 2011). Hypertrophy as well as hyperplasia are dependent on structural changes (Karpe et al. 2011), which are mainly controlled

by actin dynamics (Schwintzer et al. 2011) and accommodate the emergence and growth of lipid droplets (Karpe et al. 2011). Body mass index (BMI), a measure for obesity is a major predictor for T2D, but some genetically predisposed individuals develop T2D with a lower BMI than others, which highlights the importance of genetic factors (Prasad & Groop 2015), influencing the ability to safely store excess energy (Perry et al. 2012). A combined defective and overloaded "healthy" subcutaneous adipose tissue (Henninger et al. 2014) expansion, may lead to an unfavorable lipid storage in visceral adipose tissue or non-adipocyte cells and may - accompanied with fibrosis - lead to insulin resistance (Wernstedt Asterholm et al. 2014). Generally, our results provide evidence for an increased expression of *COBLL1* in "healthy" mature adipocytes. A disturbed *COBLL1* expression or a lower ability to regulate *COBLL1* expression may be associated with disturbances in subcutaneous lipid storage. These observations were supported by the correlation between *leptin* and *COBLL1*. Leptin is produced in proportion to the size of fat depots in the human body (Harris 2014). When *COBLL1* was perturbed or decreased we found a lower amount of *leptin*, substantiating the observation that *COBLL1* may be important during adipogenesis and lipid storage enabling a "healthy" expansion of human adipose tissue. We here provide evidence that the (up-) regulation of *COBLL1* during adipogenesis could be important to remodel actin from stress fibers to cortical actin, influencing the adipogenic capacity.

COBLL1 and lipolysis

Our results show that lower *COBLL1* expression levels coincide with a decreased β-adrenergic lipolytic response. Lipolysis is the process, which provides energy for the body in the catabolic state by facilitating the decay of triglycerides into glycerol and free fatty acids (Langin 2006). It is mainly activated by the binding of catecholamines to adrenergic G protein-coupled receptors (Jaworski et al. 2007), that increases cellular cAMP levels (Jaworski et al. 2007) and promotes the phosphorylation of the cAMP-dependent protein kinase A, which phosphorylates the hormone-sensitive lipase and the lipid droplet coating protein perilipin (Jaworski et al. 2007). In recent years, some evidence emerged for the involvement of the cytoskeleton in the regulation of lipid droplets and lipolysis. A study using *in vitro* 3D adipocyte cultures, found the involvement of the ß-Integrin/cytoskeleton signaling pathway in the lipolytic response (Pellegrinelli et al. 2014) and for perilipins a central role in lipolysis and more specifically in the fusion of lipid droplets (Orlicky et al. 2013), in concerted action with elements of the cytoskeleton has been described (Takahashi et al. 2016). Taking the effect of *COBLL1* on adipocyte actin remodeling into consideration, it may be possible that not only the storage of lipids but also their release may be disturbed in cells with *COBLL1* perturbation. We found a decreased stimulated lipolysis rate and reduced perilipin levels, when *COBLL1* was reduced. Strikingly, perilipin null mice show a similar phenotype as we have observed in *COBLL1* perturbed human adipocytes. The mice have a resistance to high fat diet induced obesity but at the same time a higher tendency to develop glucose intolerance, with an

elevated basal but reduced stimulated lipolytic response (Tansey et al. 2001). Humans with frameshift mutations in the perilipin gene have been shown to develop partial lipodystrophy, severe dyslipidemia, and insulin-resistant diabetes, with small subcutaneous adipocytes, macrophage infiltration and fibrosis (Gandotra et al. 2011). In conclusion, we found a decreased β-adrenergic lipolytic response in *COBLL1* perturbed adipocytes. In combination with a decreased adipogenic capacity, this may lead to the inability to safely store excess energy (Perry et al. 2012).

rs6712203 and body fat distribution in women

GWAS have reported a sexual dimorphism at the *GRB14/COBLL1* locus (Morris et al. 2012; Heid et al. 2010), with a stronger association with Waist-to-hip ratio (WHR) in women (Morris et al. 2012; Heid et al. 2010). Interestingly, a study focusing on the distribution of body fat found 49 loci; their putative regulatory elements were enriched in adipose tissue linking adipogenesis and insulin resistance to the regulation of body fat distribution (Shungin et al. 2015). The association with body fat distribution at the *GRB14/COBLL1* locus, may also be mediated by adipose insulin response (Kan et al. 2016). We identified a sexual dimorphism in our experiments, indicating an effect of rs6712203 on stimulated lipolysis and potentially fat distribution only in women. It is intriguing to speculate, that a decreased adipogenic capacity due to *COBLL1* perturbation in subcutaneous adipocytes may decrease the adipocyte insulin response and the ability to safely store excess energy in women, but more work will be required to investigate the effects of *COBLL1* on female fat distribution. Insight may be gained from the comparison between visceral and subcutaneous fat depots. Taken together, we here provide some evidence for a female specific effect of *COBLL1* on fat distribution, which may contribute to the risk for insulin resistance and T2D.

Summary and outlook

In the last 15 years, GWAS have facilitated a remarkable range of robust discoveries in population and complex-trait genetics, but SNP-trait associations do not directly inform about regulatory mechanisms by which a genomic variant influences phenotype and leads to an increased risk for a common complex disease like T2D (Visscher et al. 2017). In order to translate GWAS signals into biological information to predict the risk for diabetes, diabetes-related complications and the response to treatment and pharmacogenomics (Grant et al. 2009), it is of relevance to identify the regulatory circuitries at GWAS loci. For the *GRB14/COBLL1* locus, we have used the cutting-edge experimental and computational framework, previously established and continuously developed by Claussnitzer and colleagues (Claussnitzer et al. 2015; Claussnitzer et al. 2014), to identify biologically relevant mechanisms tagged by GWAS signals. We identified the POU2F2 dependent up-regulation of *COBLL1* in rs6712203-T non-risk allele carriers, which may link the *GRB14/COBLL1* locus to the risk for T2D. In the here used framework, we additionally introduced actin staining experiments to identify cellular phenotypes of the previously sparsely characterized gene *COBLL1*. We uncover a role of *COBLL1* in adipocyte actin remodeling, which may influence adipogenesis and lead to a decreased adipocyte insulin sensitivity and ability to store and release lipids. Furthermore, we find evidence that the here identified mechanisms may lead to a sexual dimorphism, with an effect on fat distribution specifically in women.

While we here show results about the successful identification of regulatory mechanisms at the *GRB14/COBLL1* locus, leading to the identification of the previously unknown role of *COBLL1* in adipocyte actin remodeling, our experimental and computational framework should (and will) in the future be extended to perform analysis on whole genome level, where effects of all loci of a specific trait can be investigated in an integrative model. Advances in computational approaches like machine learning in combination with the integration of larger and new data-sets will be detrimental to achieve this goal (Lin & Lane 2017). Projects like ENCODE, Epigenome RoadMap, GTEx (Visscher et al. 2017) and UK biobank (Bycroft et al. 2018) have been crucial to generate data in relevant tissues for the interpretation of non-coding variants. However, to fully understand the impact of genomic variants on complex diseases, it will be important to collect data more balanced between different populations. It is still estimated that only about 20 % of participants in GWAS are not of European descent. Most of these 20 % are from Asia or populations of Asian ancestry, while African and Latin American, Hispanic and indigenous peoples are rarely represented in GWAS (Popejoy & Fullerton 2016).

In the past, GWAS have been criticized because most of the common variants detected for complex diseases have only small effect sizes which may not be relevant to a disease. While this may be true in populations, it has also been shown that effect sizes of individual variants on molecular phenotypes can be large (Claussnitzer et al. 2015) and the effect of drugs targeting them, their regulators or downstream targets may be magnified (Visscher et al. 2017). As such,

also our findings about the regulatory mechanisms at the *GRB14/COBLL1* GWAS T2D risk locus may have therapeutic relevance in tackling obesity and insulin resistance related issues in the future. While we lay the groundwork in the identification of the function of *COBLL1* in adipose tissue, there were follow-up questions, raised during our investigations:

1) **Causal non-coding and coding variants:** In this thesis, we focused primarily on the top-scoring non-coding PMCA variant rs6712203, which showed consistent effects across our *in silico* and *in vitro* experiments. The variant may contribute to the regulation of *COBLL1* and may therefore affect the risk for T2D. However, there is evidence that additional regulatory and/or coding variants are contributing to the regulatory circuitry at the *GRB14/COBLL1* locus. Particularly, the exonic *COBLL1* variant rs7607980 (MAF 12.5 %), which causes a missense mutation in the third *COBLL1* 3'-exon has been shown to be associated with the risk for T2D (Albrechtsen et al. 2013). We also found some evidence for the variant rs10195252 to be involved in the non-coding GWAS signal at the *GRB14/COBLL1* locus. We found distinctly different binding patterns between the rs101095252-C non-risk and rs101095252-T risk allele (Supplementary Figure 24), which represents some evidence for a regulatory role of rs10195252. In hASC, we found evidence for CTCF binding at rs101095252 (Supplementary Figure 6). Further experiments evaluating an insulator role at rs10195252 could be interesting in the future. An integrative computational model with large scale/genome wide experimental data investigating the effects of all variants will be helpful to further prioritize the variants and their effect sizes.

2) **Lipid trait associations:** Besides insulin resistance and T2D, variants at the *GRB14/COBLL1* locus have also been associated with HDL cholesterol (Randall et al. 2013; Teslovich et al. 2010) and triglyceride levels (Desmarchelier et al. 2014; Teslovich et al. 2010). The variant rs6712203 has been associated by GC/MS metabolite measurement using the KORA and TwinsUK cohorts with glycerol and n-Butyl Oleate release (Suhre et al. 2011). Our results substantiate the association with HDL cholesterol (Supplementary Figure 21B), but it remains to be investigated how the role of *COBLL1* in adipocytes is involved in these associations.

3) **Pleiotropy:** It has been reported, that pleiotropy maybe widespread across complex traits, since many are associated with hundreds-to-thousands of genomic variants, implying that underlying causal variants are sometimes shared (Voight et al. 2010). Accordingly, studies investigating pleiotropic signals have identified cross-phenotype risk associations at the *GRB14/COBLL1* locus. They found associations shared between the metabolic syndrome and inflammation (HDLC, TG, PAI-1 and ADIP) (Kraja et al. 2014), endometriosis and WHR (adjusted for BMI) (Rahmioglu et al. 2015) and body fat percentage and cardiovascular events (Lu et al. 2016). Deciphering these pleiotropic signals will be of interest in the future.

4) **Age, BMI and gender:** At the *GRB14/COBLL1* locus, GWAS have reported sex (Morris et al. 2012; Heid et al. 2010) and BMI specific (Manning et al. 2012) associations with insulin resistance and T2D (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al. 2014;

Kooner et al. 2011; Morris et al. 2012). The female specific effect on fat distribution was further substantiated by our results. Nevertheless, the influence of age, BMI and gender at the *GRB14/COBLL1* locus should be further investigated in different datasets to gain an in-depth analysis of the influence of rs6712203 (and other variants) on *COBLL1* and their effects in adipose tissue.

5) **Actin remodeling:** We have found an interesting effect of *COBLL1* on actin remodeling during adipogenesis, potentially through interaction with actin. Nevertheless, the exact mechanism by which *COBLL1* influences the dynamics of actin filamentation and whether the localization of *COBLL1* in the nucleus plays a role remains to be investigated.

6) **Fat distribution:** Our results provide evidence to the previously observed sexual dimorphism at the *GRB14/COBLL1* locus on female specific fat distribution. In particular, effects of *COBLL1* on leptin may be of interest in this regard. So far, we were not able to further investigate the exact effects of *COBLL1* on body fat distribution, but the involvement of adipocyte actin filamentation in body fat distribution and the effect on insulin resistance will be interesting to further investigate.

Literature

Abel, E.D. et al., 2001. Adipose-selective targeting of the GLUT4 gene impairs insulin action in muscle and liver. *Nature*, 409(6821), pp.729–733.

Albrechtsen, A. et al., 2013. Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia*, 56(2), pp.298–310.

American Diabetes Association, 2014. 2. Classification and Diagnosis of Diabetes. *Diabetes care*, 38(Supplement_1), pp.S8–S16.

Anon, 2014. Diagnosis and Classification of diabetes mellitus. In *Diapedia*.

Anthonsen, M.W. et al., 1998. Identification of novel phosphorylation sites in hormone-sensitive lipase that are phosphorylated in response to isoproterenol and govern activation properties in vitro. *The Journal of biological chemistry*, 273(1), pp.215–221.

Ashburner, M. et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), pp.25–29.

Banerjee-Basu, S., 2001. Molecular evolution of the homeodomain family of transcription factors. *Nucleic acids research*, 29(15), pp.3258–3269.

Barrett, J.C. et al., 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics*, 41(6), pp.703–707.

Benita, Y. et al., 2010. Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor. *Blood*, 115(26), pp.5376–5384.

Boden, G., 1996. Effect of fasting on serum leptin in normal human subjects. *The Journal of clinical endocrinology and metabolism*, 81(9), pp.3419–3423.

Bogan, J.S. et al., 2003. Functional cloning of TUG as a regulator of GLUT4 glucose transporter trafficking. *Nature*, 425(6959), pp.727–733.

Boyle, E.A., Li, Y.I. & Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7), pp.1177–1186.

Bycroft, C. et al., 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726), pp.203–209.

Cariou, B. et al., 2004. Increased adipose tissue expression of Grb14 in several models of insulin resistance. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 18(9), pp.965–967.

Carithers, L.J. et al., 2015. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking*, 13(5), pp.311–319. Available at: http://dx.doi.org/10.1089/bio.2015.0032.

Carlsson, S. et al., 2013. Shared genetic influence of BMI, physical activity and type 2 diabetes: a twin study. *Diabetologia*, 56(5), pp.1031–1035.

Carroll, E.A. et al., 2003. Cordon-bleu is a conserved gene involved in neural tube formation. *Developmental biology*, 262(1), pp.16–31.

Ceccarelli, D.F.J. & Sicheri, F., 2009. Grb-ing hold of insulin signaling. *Nature structural & molecular biology*, 16(8), pp.803–804.

Cho, N.H. et al., 2018. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes research and clinical practice*, 138, pp.271–281.

Chukwuma, C., Sr & Tuomilehto, J., 1998. The “thrifty” hypotheses: clinical and epidemiological significance for non-insulin-dependent diabetes mellitus and cardiovascular disease risk factors. *Journal of cardiovascular risk*, 5(1), pp.11–23.

Claussnitzer, M. et al., 2011. Effect of flavonoids on basal and insulin-stimulated 2-deoxyglucose uptake in adipocytes. *Molecular nutrition & food research*, 55 Suppl 1, pp.S26–34.

Claussnitzer, M. et al., 2015. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *The New England journal of medicine*, 373(10), pp.895–907.

Claussnitzer, M. et al., 2014. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell*, 156(1-2), pp.343–358.

Clerc, R.G. et al., 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes & development*, 2(12A), pp.1570–1581.

Cooney, G.J. et al., 2004. Improved glucose homeostasis and enhanced insulin signalling in Grb14-deficient mice. *The EMBO journal*, 23(3), pp.582–593.

Cordes, C. et al., 2015. MR-detected changes in liver fat, abdominal fat, and vertebral bone marrow fat after a four-week calorie restriction in obese women. *Journal of magnetic resonance imaging: JMRI*, 42(5), pp.1272–1280.

Cowherd, R.M. et al., 1997. Developmental profile of homeobox gene expression during 3T3-L1 adipogenesis. *Biochemical and biophysical research communications*, 237(2), pp.470–475.

Cowper-Sal lari, R. et al., 2012. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nature genetics*, 44(11), pp.1191–1198.

D'Adamo, E. & Caprio, S., 2011. Type 2 Diabetes in Youth: Epidemiology and Pathophysiology. *Diabetes care*, 34(Supplement_2), pp.S161–S165.

Depetris, R.S. et al., 2005. Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. *Molecular cell*, 20(2), pp.325–333.

Desmarchelier, C. et al., 2014. The postprandial chylomicron triacylglycerol response to dietary fat in healthy male adults is significantly explained by a combination of single nucleotide polymorphisms in genes involved in triacylglycerol metabolism. *The Journal of clinical endocrinology and metabolism*, 99(3), pp.E484–8.

DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium et al., 2014. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*, 46(3), pp.234–244.

Diamond, J., 2003. The double puzzle of diabetes. *Nature*, 423(6940), pp.599–602.

Dixon, J.R. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376–380.

Dong, B. & Zhao, F.-Q., 2007. Involvement of the ubiquitous Oct-1 transcription factor in hormonal induction of beta-casein gene expression. *Biochemical Journal*, 401(1), pp.57–64.

Dugail, I., 2001. Transfection of adipocytes and preparation of nuclear extracts. *Methods in molecular biology* , 155, pp.141–146.

Eftychi, C. et al., 2004. Analysis of the type 2 diabetes-associated single nucleotide polymorphisms in the genes IRS1, KCNJ11, and PPARG2 in type 1 diabetes. *Diabetes*, 53(3), pp.870–873.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

ENCODE Project Consortium, 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 306(5696), pp.636–640.

Ernst, J. & Kellis, M., 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*, 9(3), pp.215–216.

Ernst, J. & Kellis, M., 2015. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology*, 33(4), pp.364–376.

Fischer-Posovszky, P. et al., 2008. Human SGBS Cells – a Unique Tool for Studies of Human Fat Cell Biology. *Obesity facts*, 1(4), pp.184–189.

Fuchsberger, C. et al., 2016. The genetic architecture of type 2 diabetes. *Nature*, 536(7614), pp.41–47.

Gandotra, S. et al., 2011. Perilipin deficiency and autosomal dominant partial lipodystrophy. *The New England journal of medicine*, 364(8), pp.740–748.

Garton, A.J. et al., 1989. Phosphorylation of bovine hormone-sensitive lipase by the AMP-activated protein kinase. A possible antilipolytic mechanism. *European journal of biochemistry / FEBS*, 179(1), pp.249–254.

Gehring, W., 1994. Homeodomain Proteins. *Annual review of biochemistry*, 63(1), pp.487–526.

Gnudi, L. et al., 1995. High level overexpression of glucose transporter-4 driven by an adipose-specific promoter is maintained in transgenic mice on a high fat diet, but does not prevent impaired glucose tolerance. *Endocrinology*, 136(3), pp.995–1002.

Grant, R.W., Moore, A.F. & Florez, J.C., 2009. Genetic architecture of type 2 diabetes: recent progress and clinical implications. *Diabetes care*, 32(6), pp.1107–1114.

Greenberg, A.S. et al., 2011. The role of lipid droplets in metabolic disease in rodents and humans. *The Journal of clinical investigation*, 121(6), pp.2102–2110.

Groop, L. & Pociot, F., 2014. Genetics of diabetes – Are we missing the genes or the disease? *Molecular and cellular endocrinology*, 382(1), pp.726–739.

Hannon, T.S., 2005. Childhood Obesity and Type 2 Diabetes Mellitus. *Pediatrics*, 116(2), pp.473–480.

Harris, R.B.S., 2014. Direct and indirect effects of leptin on adipocyte metabolism. *Biochimica et biophysica acta*, 1842(3), pp.414–423.

Hatzopoulos, A.K. et al., 1990. Structure and expression of the mouse Oct2a and Oct2b, two differentially spliced products of the same gene. *Development* , 109(2), pp.349–362.

Hauner, H. et al., 1998. Development of insulin-responsive glucose uptake and GLUT4 expression in differentiating human adipocyte precursor cells. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, 22(5), pp.448–453.

Hauner, H., Skurk, T. & Wabitsch, M., 2001. Cultures of Human Adipose Precursor Cells. In *Adipose Tissue Protocols*. pp. 239–247.

Heid, I.M. et al., 2010. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics*, 42(11), pp.949–960.

Henninger, A.M.J. et al., 2014. Adipocyte hypertrophy, inflammation and fibrosis characterize subcutaneous adipose tissue of healthy, non-obese subjects predisposed to type 2 diabetes. *PloS one*, 9(8), p.e105262.

Herr, W. & Cleary, M.A., 1995. The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes & development*, 9(14), pp.1679–1693.

Holt, L.J. et al., 2009. Dual ablation of Grb10 and Grb14 in mice reveals their combined role in regulation of insulin signaling and glucose homeostasis. *Molecular endocrinology* , 23(9), pp.1406–1414.

Hsu, P.D., Lander, E.S. & Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), pp.1262–1278.

Husson, C. et al., 2011. Cordon-Bleu uses WH2 domains as multifunctional dynamizers of actin filament assembly. *Molecular cell*, 43(3), pp.464–477.

Imakaev, M. et al., 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods*, 9(10), pp.999–1003.

Jaworski, K. et al., 2007. Regulation of triglyceride metabolism. IV. Hormonal regulation of lipolysis in adipose tissue. *American journal of physiology. Gastrointestinal and liver physiology*, 293(1), pp.G1–4.

Jiao, Y. et al., 2014. Mutagenetic and electron microscopy analysis of actin filament severing by Cordon-Bleu, a WH2 domain protein. *Cytoskeleton* , 71(3), pp.170–183.

Kan, M. et al., 2016. Rare variant associations with waist-to-hip ratio in European-American and African-American women from the NHLBI-Exome Sequencing Project. *European journal of human genetics: EJHG*, 24(8), pp.1181–1187.

Kanzaki, M., 2006. Insulin receptor signals regulating GLUT4 translocation and actin dynamics. *Endocrine journal*, 53(3), pp.267–293.

Kanzaki, M. et al., 2002. Small GTP-binding protein TC10 differentially regulates two distinct populations of filamentous actin in 3T3L1 adipocytes. *Molecular biology of the cell*, 13(7), pp.2334–2346.

Kanzaki, M. & Pessin, J.E., 2001. Insulin-stimulated GLUT4 translocation in adipocytes is dependent upon cortical actin remodeling. *The Journal of biological chemistry*, 276(45), pp.42436–42444.

Karpe, F., Dickmann, J.R. & Frayn, K.N., 2011. Fatty acids, obesity, and insulin resistance: time for a reevaluation. *Diabetes*, 60(10), pp.2441–2449.

Kasus-Jacobi, A. et al., 1998. Identification of the rat adapter Grb14 as an inhibitor of insulin actions. *The Journal of biological chemistry*, 273(40), pp.26026–26035.

Kawaguchi, N. et al., 2003. ADAM12 induces actin cytoskeleton and extracellular matrix reorganization during early adipocyte differentiation by regulating beta1 integrin function. *Journal of cell science*, 116(Pt 19), pp.3893–3904.

Kessels, M.M. & Qualmann, B., 2004. The syndapin protein family: linking membrane trafficking with the cytoskeleton. *Journal of cell science*, 117(Pt 15), pp.3077–3086.

Kheradpour, P. & Kellis, M., 2014. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic acids research*, 42(5), pp.2976–2987.

Kooner, J.S. et al., 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature genetics*, 43(10), pp.984–989.

Kraja, A.T. et al., 2014. Pleiotropic genes for metabolic syndrome and inflammation. *Molecular genetics and metabolism*, 112(4), pp.317–338.

Lajoie, B.R., Dekker, J. & Kaplan, N., 2015. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods* , 72, pp.65–75.

Langin, D., 2006. Adipose tissue lipolysis as a metabolic pathway to define pharmacological strategies against obesity and the metabolic syndrome. *Pharmacological research: the official journal of the Italian Pharmacological Society*, 53(6), pp.482–491.

Latchman, D.S., 1996. The Oct-2 transcription factor. *The international journal of biochemistry & cell biology*, 28(10), pp.1081–1083.

Lehtovirta, M. et al., 2010. Evidence that BMI and type 2 diabetes share only a minor fraction of genetic variance: a follow-up study of 23,585 monozygotic and dizygotic twins from the Finnish Twin Cohort Study. *Diabetologia*, 53(7), pp.1314–1321.

Lin, E. & Lane, H.-Y., 2017. Machine learning and systems genomics approaches for multi-omics data. *Biomarker research*, 5, p.2.

Li, Y.R. & Keating, B.J., 2014. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome medicine*, 6(10), p.91.

Lupski, J.R. et al., 2011. Clan genomics and the complex architecture of human disease. *Cell*, 147(1), pp.32–43.

Lu, Y. et al., 2016. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nature communications*, 7, p.10495.

Mancina, R.M. et al., 2013. The COBLL1 C allele is associated with lower serum insulin levels and lower insulin resistance in overweight and obese children. *Diabetes/metabolism research and reviews*, 29(5), pp.413–416.

Manning, A.K. et al., 2012. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature genetics*, 44(6), pp.659–669.

Martin, S.S., Haruta, T., et al., 1996. Activated phosphatidylinositol 3-kinase is sufficient to mediate actin rearrangement and GLUT4 translocation in 3T3-L1 adipocytes. *The Journal of biological chemistry*, 271(30), pp.17605–17608.

Martin, S.S., Rose, D.W., et al., 1996. Phosphatidylinositol 3-kinase is necessary and sufficient for insulin-stimulated stress fiber breakdown. *Endocrinology*, 137(11), pp.5045–5054.

McAllister, E.J. et al., 2009. Ten putative contributors to the obesity epidemic. *Critical reviews in food science and nutrition*, 49(10), pp.868–913.

McCarthy, M.I., 2008. Casting a wider net for diabetes susceptibility genes. *Nature genetics*, 40(9), pp.1039–1040.

McClellan, J. & King, M.-C., 2010. Genetic heterogeneity in human disease. *Cell*, 141(2), pp.210–217.

McGinnis, W. et al., 1984. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature*, 308(5958), pp.428–433.

Meigs, J.B., Cupples, L.A. & Wilson, P.W., 2000. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes*, 49(12), pp.2201–2207.

Mikkelsen, T.S. et al., 2010. Comparative epigenomic analysis of murine and human adipogenesis. *Cell*, 143(1), pp.156–169.

Moayyeri, Alireza, Christopher J. Hammond, Deborah J. Hart, and Timothy D. Spector. 2013. “The UK Adult Twin Registry (TwinsUK Resource).” *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies* 16 (1): 144–49.

Mohlke, K.L. & Boehnke, M., 2015. Recent advances in understanding the genetic architecture of type 2 diabetes. *Human molecular genetics*, 24(R1), pp.R85–92.

Morandi, E.M. et al., 2016. ITGAV and ITGA5 diversely regulate proliferation and adipogenic differentiation of human adipose derived stem cells. *Scientific reports*, 6, p.28889.

Morris, A.P. et al., 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*, 44(9), pp.981–990.

Morris, A.P., 2011. Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology*, 35(8), pp.809–822.

Mundy, D.I. et al., 2002. Dual control of caveolar membrane traffic by microtubules and the actin cytoskeleton. *Journal of cell science*, 115(Pt 22), pp.4327–4339.

Newman, B. et al., 1987. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia*, 30(10), pp.763–768.

Nobusue, H. et al., 2014. Regulation of MKL1 via actin cytoskeleton dynamics drives adipocyte differentiation. *Nature communications*, 5, p.3368.

Ongen, H. & Dermitzakis, E.T., 2015. Alternative splicing QTLs in European and African populations using Altrans, a novel method for splice junction quantification. Available at: http://dx.doi.org/10.1101/014126.

Orlicky, D.J. et al., 2013. Dynamics and molecular determinants of cytoplasmic lipid droplet clustering and dispersion. *PloS one*, 8(6), p.e66837.

Ott, B. et al., 2017. Effect of caloric restriction on gut permeability, inflammation markers, and fecal microbiota in obese women. *Scientific reports*, 7(1), p.11955.

Ovaska, K. et al., 2013. Integrative analysis of deep sequencing data identifies estrogen receptor early response genes and links ATAD3B to poor survival in breast cancer. *PLoS computational biology*, 9(6), p.e1003100.

Pairault, J. & Green, H., 1979. A study of the adipose conversion of suspended 3T3 cells by using glycerophosphate dehydrogenase as differentiation marker. *Proceedings of the National Academy of Sciences of the United States of America*, 76(10), pp.5138–5142.

Pellegrinelli, V. et al., 2014. Human adipocyte function is impacted by mechanical cues. *The Journal of pathology*, 233(2), pp.183–195.

Perry, J.R.B. et al., 2012. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS genetics*, 8(5), p.e1002741.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, 29(9), p.e45.

Phillips, K. & Luisi, B., 2000. The virtuoso of versatility: POU proteins that flex to fit. *Journal of molecular biology*, 302(5), pp.1023–1039.

Plomin, R., Haworth, C.M.A. & Davis, O.S.P., 2009. Common disorders are quantitative traits. *Nature reviews. Genetics*, 10(12), pp.872–878.

Popejoy, A.B. & Fullerton, S.M., 2016. Genomics is failing on diversity. *Nature*, 538(7624), pp.161–164.

Prasad, R.B. & Groop, L., 2015. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes*, 6(1), pp.87–123.

Rahmioglu, N. et al., 2015. Genome-wide enrichment analysis between endometriosis and obesity-related traits reveals novel susceptibility loci. *Human molecular genetics*, 24(4), pp.1185–1199.

Raj, S.M. et al., 2009. No association of multiple type 2 diabetes loci with type 1 diabetes. *Diabetologia*, 52(10), pp.2109–2116.

Randall, J.C. et al., 2013. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS genetics*, 9(6), p.e1003500.

Reynisdottir, S. et al., 1995. Effects of weight reduction on the regulation of lipolysis in adipocytes of women with upper-body obesity. *Clinical science* , 89(4), pp.421–429.

Roadmap Epigenomics Consortium et al., 2015. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539), pp.317–330.

Rosen, E.D. & MacDougald, O.A., 2006. Adipocyte differentiation from the inside out. *Nature reviews. Molecular cell biology*, 7(12), pp.885–896.

Rueden, C.T. et al., 2017. ImageJ2: ImageJ for the next generation of scientific image data. *BMC bioinformatics*, 18(1), p.529.

Salans, L.B., Knittle, J.L. & Hirsch, J., 1968. The role of adipose cell size and adipose tissue insulin sensitivity in the carbohydrate intolerance of human obesity. *The Journal of clinical investigation*, 47(1), pp.153–165.

Scharf, P.J. et al., 2004. Solution structure of the human Grb14-SH2 domain and comparison with the structures of the human Grb7-SH2/erbB2 peptide complex and human Grb10-SH2 domain. *Protein science: a publication of the Protein Society*, 13(9), pp.2541–2546.

Schleinitz, D. et al., 2014. Fat depot-specific mRNA expression of novel loci associated with waist-hip ratio. *International journal of obesity* , 38(1), pp.120–125.

Schreiber, E. et al., 1989. Rapid detection of octamer binding proteins with “mini-extracts”, prepared from a small number of cells. *Nucleic acids research*, 17(15), p.6419.

Schultz, J. & Terhoeven, N., 2013. The bilaterian roots of cordon-bleu. *BMC research notes*, 6, p.393.

Schwintzer, L. et al., 2011. The functions of the actin nucleator Cobl in cellular morphogenesis critically depend on syndapin I. *The EMBO journal*, 30(15), pp.3147–3159.

Scott, R.A. et al., 2014. Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes*, 63(12), pp.4378–4387.

Shungin, D. et al., 2015. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538), pp.187–196.

Sivaraja, M. et al., 1994. Solution structure of a POU-specific homeodomain: 3D-NMR studies of human B-cell transcription factor Oct-2. *Biochemistry*, 33(33), pp.9845–9855.

Skurk, T. et al., 2007. Relationship between adipocyte size and adipokine expression and secretion. *The Journal of clinical endocrinology and metabolism*, 92(3), pp.1023–1033.

Smas, C.M. & Sul, H.S., 1995. Control of adipocyte differentiation. *Biochemical Journal*, 309 (Pt 3), pp.697–710.

Smemo, S. et al., 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, 507(7492), pp.371–375.

Sonawane, A.R. et al., 2017. Understanding Tissue-Specific Gene Regulation. *Cell reports*, 21(4), pp.1077–1088.

Sowa, M.E. et al., 2009. Defining the human deubiquitinating enzyme interaction landscape. *Cell*, 138(2), pp.389–403.

Spalding, K.L. et al., 2017. Impact of fat mass and distribution on lipid turnover in human adipose tissue. *Nature communications*, 8, p.15253.

Stark, C. et al., 2006. BioGRID: a general repository for interaction datasets. *Nucleic acids research*, 34(Database issue), pp.D535–9.

Stranger, B.E., Stahl, E.A. & Raj, T., 2011. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 187(2), pp.367–383.

Strawbridge, R.J. et al., 2016. Effects of Genetic Loci Associated with Central Obesity on Adipocyte Lipolysis. *PloS one*, 11(4), p.e0153990.

Suhre, K. et al., 2011. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 477(7362), pp.54–60.

Takahashi, Y. et al., 2016. Perilipin2 plays a positive role in adipocytes during lipolysis by escaping proteasomal degradation. *Scientific reports*, 6(1). Available at: http://dx.doi.org/10.1038/srep20975.

Takayama, K.-I. et al., 2018. COBLL1 modulates cell morphology and facilitates androgen receptor genomic binding in advanced prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), pp.4975–4980.

Talior-Volodarsky, I. et al., 2008. Alpha-actinin-4 is selectively required for insulin-induced GLUT4 translocation. *The Journal of biological chemistry*, 283(37), pp.25115–25123.

Tansey, J.T. et al., 2001. Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proceedings of the National Academy of Sciences*, 98(11), pp.6494–6499.

Teslovich, T.M. et al., 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307), pp.707–713.

Uhlén, M. et al., 2015. Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220), p.1260419.

Uhlen, M. et al., 2010. Towards a knowledge-based Human Protein Atlas. *Nature biotechnology*, 28(12), pp.1248–1250.

UniProt Consortium, 2015. UniProt: a hub for protein information. *Nucleic acids research*, 43(Database issue), pp.D204–12.

Visscher, P.M. et al., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics*, 101(1), pp.5–22.

Voight, B.F. et al., 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics*, 42(7), pp.579–589.

Ward, L.D. & Kellis, M., 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic acids research*, 40(Database issue), pp.D930–4.

Ward, L.D. & Kellis, M., 2016. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic acids research*, 44(D1), pp.D877–81.

Wayt, J. & Bretscher, A., 2014. Cordon Bleu serves as a platform at the basal region of microvilli, where it regulates microvillar length through its WH2 domains. *Molecular biology of the cell*, 25(18), pp.2817–2827.

Weaver, A.M. et al., 2003. Integration of signals to the Arp2/3 complex. *Current opinion in cell biology*, 15(1), pp.23–30.

Welter, D. et al., 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research*, 42(Database issue), pp.D1001–6.

Wernstedt Asterholm, I. et al., 2014. Adipocyte inflammation is essential for healthy adipose tissue expansion and remodeling. *Cell metabolism*, 20(1), pp.103–118.

Weston, L., Coutts, A.S. & La Thangue, N.B., 2012. Actin nucleators in the nucleus: an emerging theme. *Journal of cell science*, 125(Pt 15), pp.3519–3527.

Xue, A. et al., 2018. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature communications*, 9(1), p.2941.

Yang, W., Thein, S., Lim, C.-Y., et al., 2014. Arp2/3 complex regulates adipogenesis by controlling cortical actin remodelling. *Biochemical Journal*, 464(2), pp.179–192.

Yang, W., Thein, S., Wang, X., et al., 2014. BSCL2/seipin regulates adipogenesis through actin cytoskeleton remodelling. *Human molecular genetics*, 23(2), pp.502–513.

Zhao, C. et al., 2006. Solution Structure of the Novel Identified Ubiquitin-like Domain in the Human COBL-like 1 Protein. Available at: http://dx.doi.org/10.2210/pdb2daj/pdb.

Zhao, F.-Q., 2013. Octamer-binding transcription factors: genomics and functions. *Frontiers in bioscience* , 18, pp.1051–1071.

Zhao, F.-Q., Adachi, K. & Oka, T., 2002. Involvement of Oct-1 in transcriptional regulation of beta-casein gene expression in mouse mammary gland. *Biochimica et biophysica acta*, 1577(1), pp.27–37.

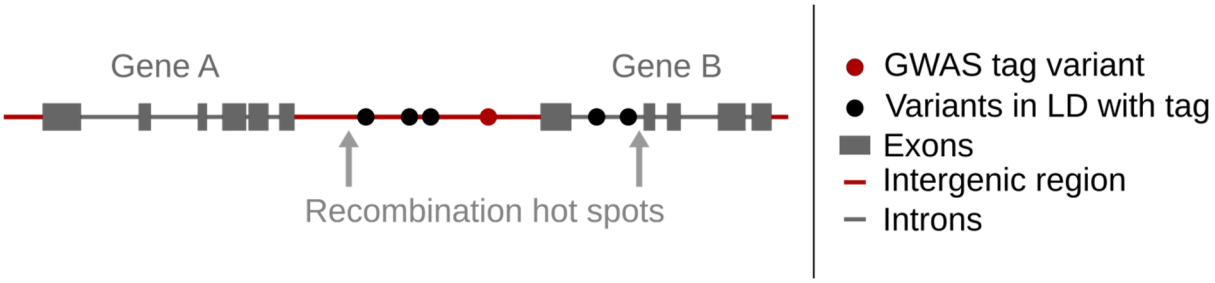
Zhou, X. et al., 2011. The Human Epigenome Browser at Washington University. *Nature methods*, 8(12), pp.989–990.

Appendix

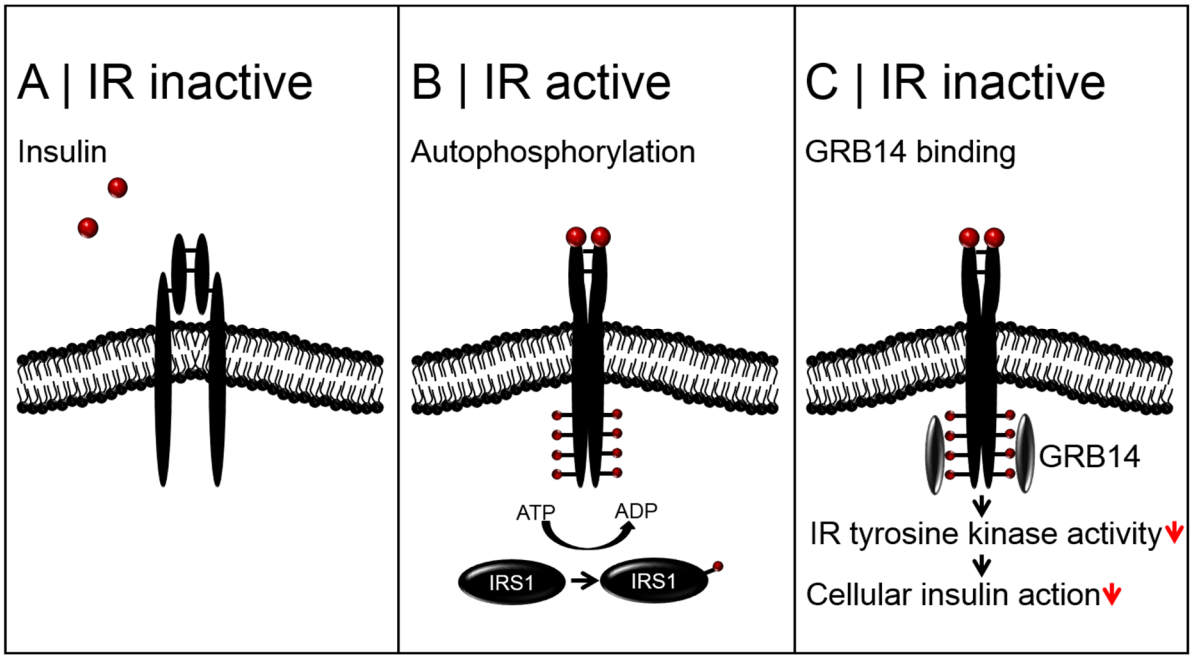
List of supplementary figures

List of supplementary tables

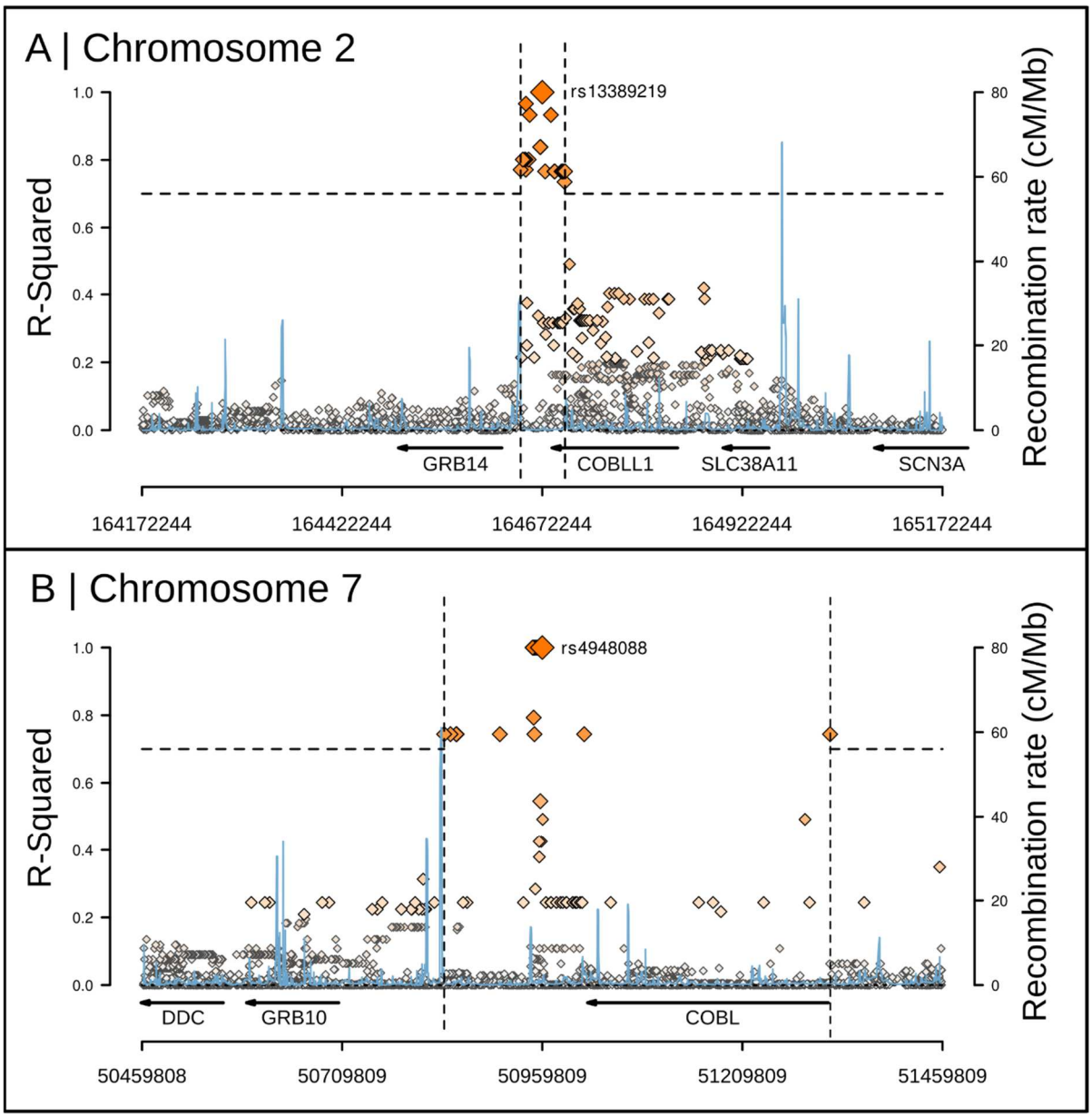
Supplementary results



Supplementary figure 1: Challenges inferring disease causal variant(s) from GWAS. The schematic draft shows a common GWAS signal. Variants in the human genome are inherited in haplotypes in which their recombination rate is decreased (Stranger et al. 2011). Haplotypes can be identified by the quantification of the occurrences of allele combinations in a specific population. Variants are in linkage disequilibrium (LD), when these occurrences are higher than expected by random. Here, the GWAS tagging variant (red dot) is in a haplotype, framed by recombination hot spots. In addition to many variants in LD (black dots), these haplotypes may potentially contain more than one cis-regulatory variant, with combinatorial or opposing effects. When genotype-information is incomplete and complex phenotypes like T2D are investigated, the identification of disease causal variant(s) only based on genome wide information becomes improbable. Thus, the combination of computational and experimental data is required to decipher GWAS.



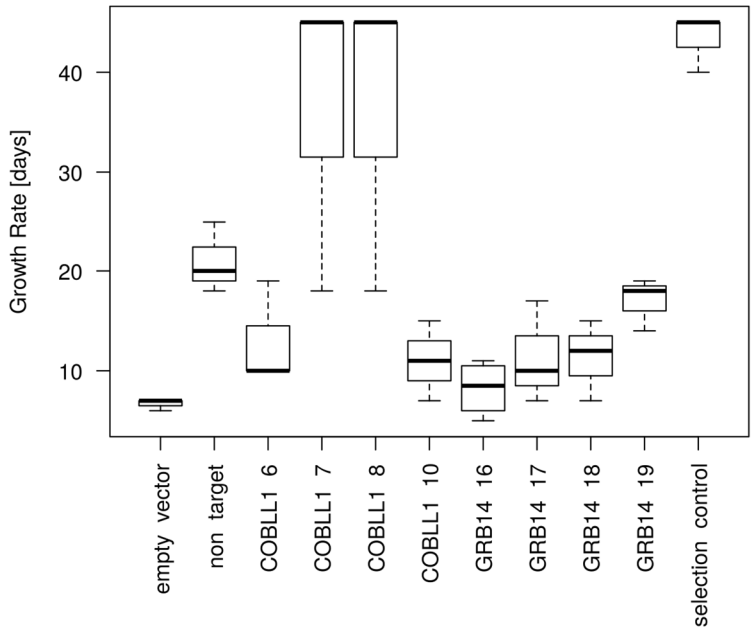
Supplementary figure 2: GRB14 is a negative regulator of the insulin receptor (IR). Panel A shows, binding of insulin at the IR, which leads to autophosphorylation of the intracellular tyrosine kinase. Its activity leads to phosphorylation of the, insulin receptor substrates 1–3 and Shc proteins, which enables downstream IR signaling (Panel B). GRB14 binding at the intracellular beta-subunit of the IR leads (Panel C) to a decreased IR tyrosine kinase activity and limits the cellular insulin action (Cariou et al. 2004; Kasus-Jacobi et al. 1998; Ceccarelli & Sicheri 2009).



Supplementary figure 3: Regional LD blots for the GRB14/COBLL1 and GRB10/COBL diabetes association signals. LD blots were generated and adjusted from the SNAP Proxy (broad institute) server. The depicted distance limit was set to 500 kb, the r2 (R-Squared) threshold to 0.7 and the 1000 Genomes Pilot 1 CEU population was used for calculations. The genome coordinates were lifted over from hg18 to the current genome build hg38 using the UCSC Genome Browser lift-over tool.

Supplementary table 1: Target Gene Sequences for stable knock down of COBLL1, and GRB14 in SGBS cells using the Lentiviral MISSION® shRNA library system from Sigma-Aldrich (Sigma-Aldrich, Steinheim, Germany). Clone ID represent the gene of interest and the here used ID of the cell clones, the Clone name represents the Sigma-Aldrich identification codes with the Target Sequence which is targeted in the respective clone.

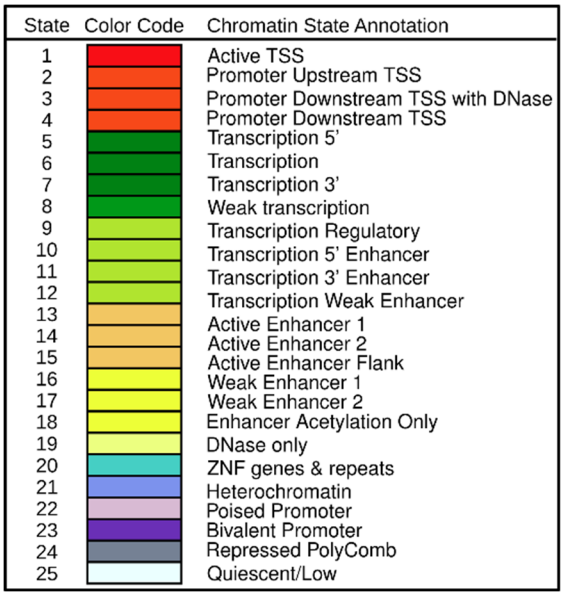
Clone ID	Clone name	Target sequence
COBLL1 6	NM_014900.2-3071s1c1	GCCTTGAAGAGATAGATGAAA
COBLL1 7	NM_014900.2-3859s1c1	GCACATGGTAATGATGATCTT
COBLL1 8	NM_014900.2-5159s1c1	GAAGGTCAAGACTCAGCCATT
COBLL1 10	NM_014900.2-4440s1c1	GAAGAGAGTATCGGGTCACTA
GRB14 16	NM_004490.1-1581s1c1	CCAGAATTATATGCATCCATA
GRB14 17	NM_004490.1-1243s1c1	CCTGAAATTCATGGTTTCTTA
GRB14 18	NM_004490.1-765s1c1	GCCATCTATTCCAAACCCTTT
GRB14 19	NM_004490.1-2101s1c1	CAACTCAATAAGGGCGTTCTT



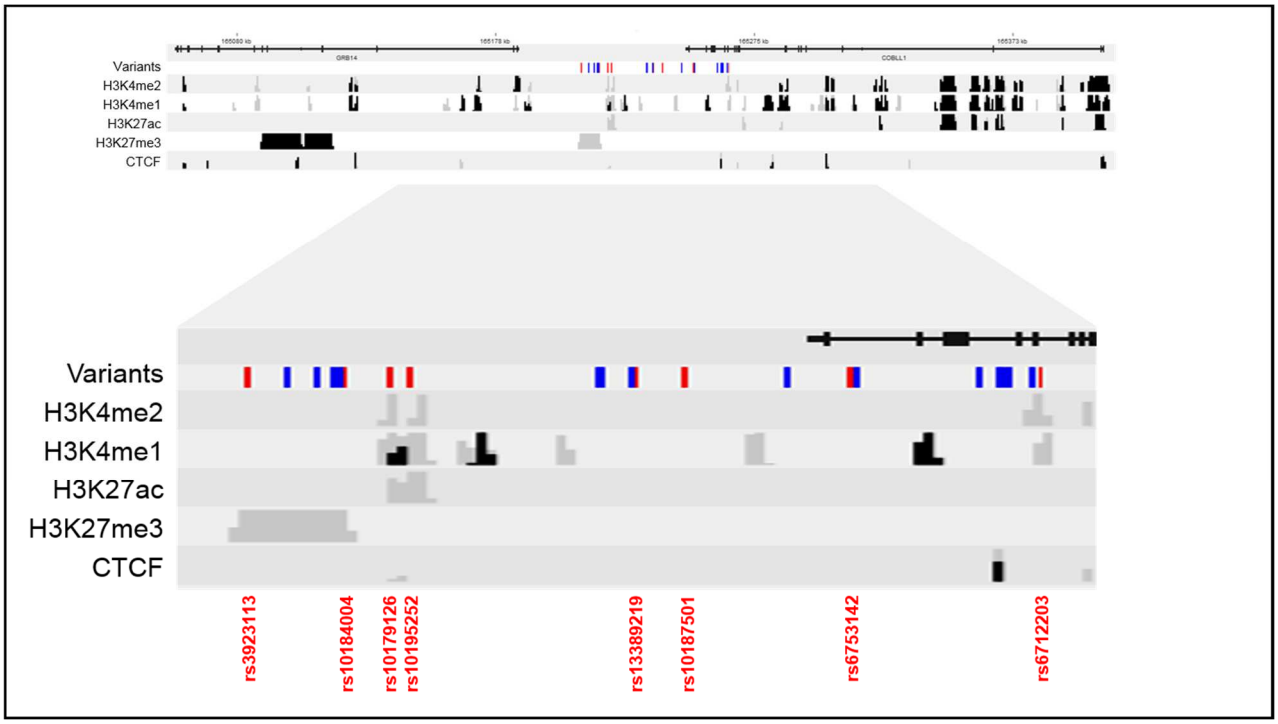
Supplementary figure 4: Growth rates of SGBS clones after lentiviral transduction. SGBS cells were transduced with lentiviral particles, against the target genes COBLL1 and GRB14, non-target or empty vector were used as controls. Non-transduced cells were included in the experiment as selection control to test for lentiviral infection. CloneIDs are given on the x-axis and growth rates are shown on the y-axis, in days until the cells reached confluency. Cells were sub-cloned three times. The number of days until they reached confluency is represented in boxplots. The cell proliferation in non-target vector cells and COBLL1 clones 7 and 8 were markedly decreased. They were excluded from further investigations. COBLL1 6, COBLL1 10 and non-target cells showed normal proliferation profiles. They were used to perform perturbation experiments.

Supplementary table 2: RT-qPCR primer sequences

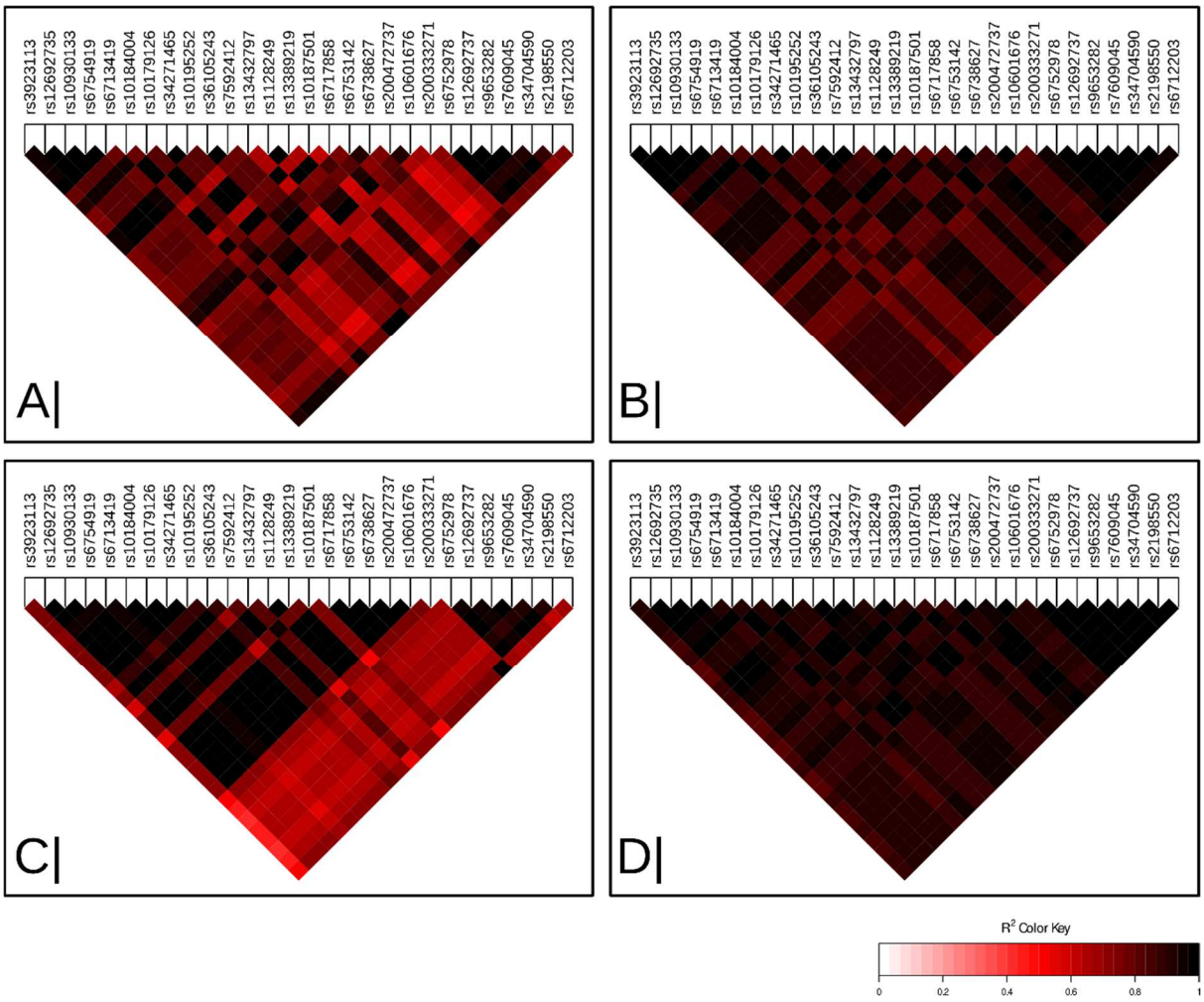
Target Gene	Forward sequence	Reverse sequence	Tm [°C]
GLUT4	CTGTGCCATCCTGATGACTG	CCAGGGCCAATCTCAAAA	60
COBLL1 pair 1	TGATCCTGACTCAGCCCTT	CAACTACCTTCTTTCCATTGC	58
COBLL1 pair 2	GGCCCATCATTCACTGTTATG	GGCTGAGTCTTGACCTTCCA	58
GRB14	CAATAGAAGACCACGAACTG	CGCATGTAAGAAACCATGAA	58
POU2F2	ACTCATGTTGACGGGCAGC	GATTTGGTGTCGGTAGCAGG	60
PLIN1	TTTCTGCCTGAGGAGACACTC	CCATCCTCGCTCCTCAAGT	60
LEPTIN	TGGGAAGGAAAATGCATTGGG	ATAAGGTCAGGATGGGGTGG	60
HPRT	TGAAAAGGACCCCACGAAG	AAGCAGATGGCCACAGAACTAG	60
YWHAZ	GCAACCAACACATCCTATCAGAC	TTCTCCTGCTTCAGCTTCGTC	60
PPIA	TGGTTCCCAGTTTTTCATC	CGAGTTGTCCACAGTCAGC	60
IPO8	CGGATTATAGTCTCTGACCATGTG	TGTGTCACCATGTTCTTCAGG	60



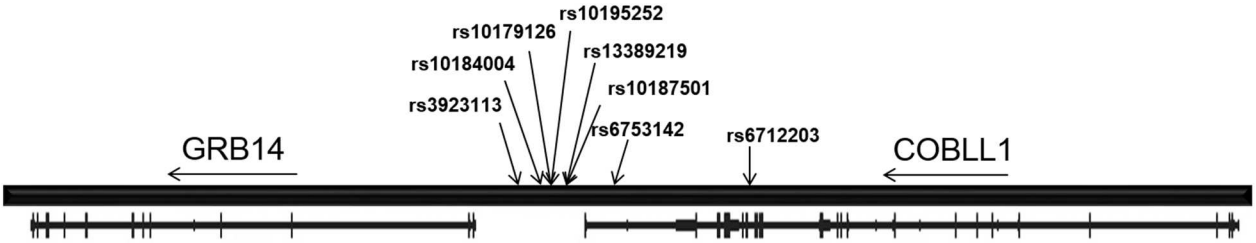
Supplementary figure 5: Color code of the 25-state chromatin annotations used to depict data from the Roadmap and ENCODE projects. The chromatin model was built from 12 imputed epigenomic marks learned jointly across 127 reference epigenomes (Claussnitzer et al. 2015). We primarily focused on enhancer states (states 9-18), which are characterized by high levels of enhancer-associated marks (H3K4me1 and H3k27ac) and lower levels of promoter-associated marks (H3K4me3, H3K4me2, and H3K9ac). These enhancer elements were divided in transcribed enhancers (states 9-12), which also show transcription-associated marks (H3K36me3, H4K20me1, H3K79me2), active enhancers (states 13-15), with stronger H3K27ac and weak enhancers (states 16-18), with weaker H3K27ac.



Supplementary figure 6: Chromatin marks in hASC and hAC at the GRB14/COBLL1 locus. The upper panel shows an overview of chromatin marks in hASC (grey) and hAC (black) over the whole range of the genes GRB14 and COBLL1. The lower panel shows an enlargement of the intergenic region between GRB14 and COBLL1. Variants are depicted as blue or red strokes. Red strokes are additionally annotated with their respective rs-numbers and represent the PMCA candidate variants rs6712203, rs10184004, rs10179126, rs10195252, rs10187501, rs6753142 and the control GWAS tagging variants rs3923113 and rs13389219. Shown are enhancer associated chromatin marks H3K4me2, H3K4me1, H3K27ac, repressed gene associated mark H3K27me3 and CTCF binding during adipogenesis from hASC (grey) to hAC (black). Data was derived from Mikkelsen and colleagues (Mikkelsen et al. 2010) and visualized in the genomatix browser (Genomatix, München, Germany).

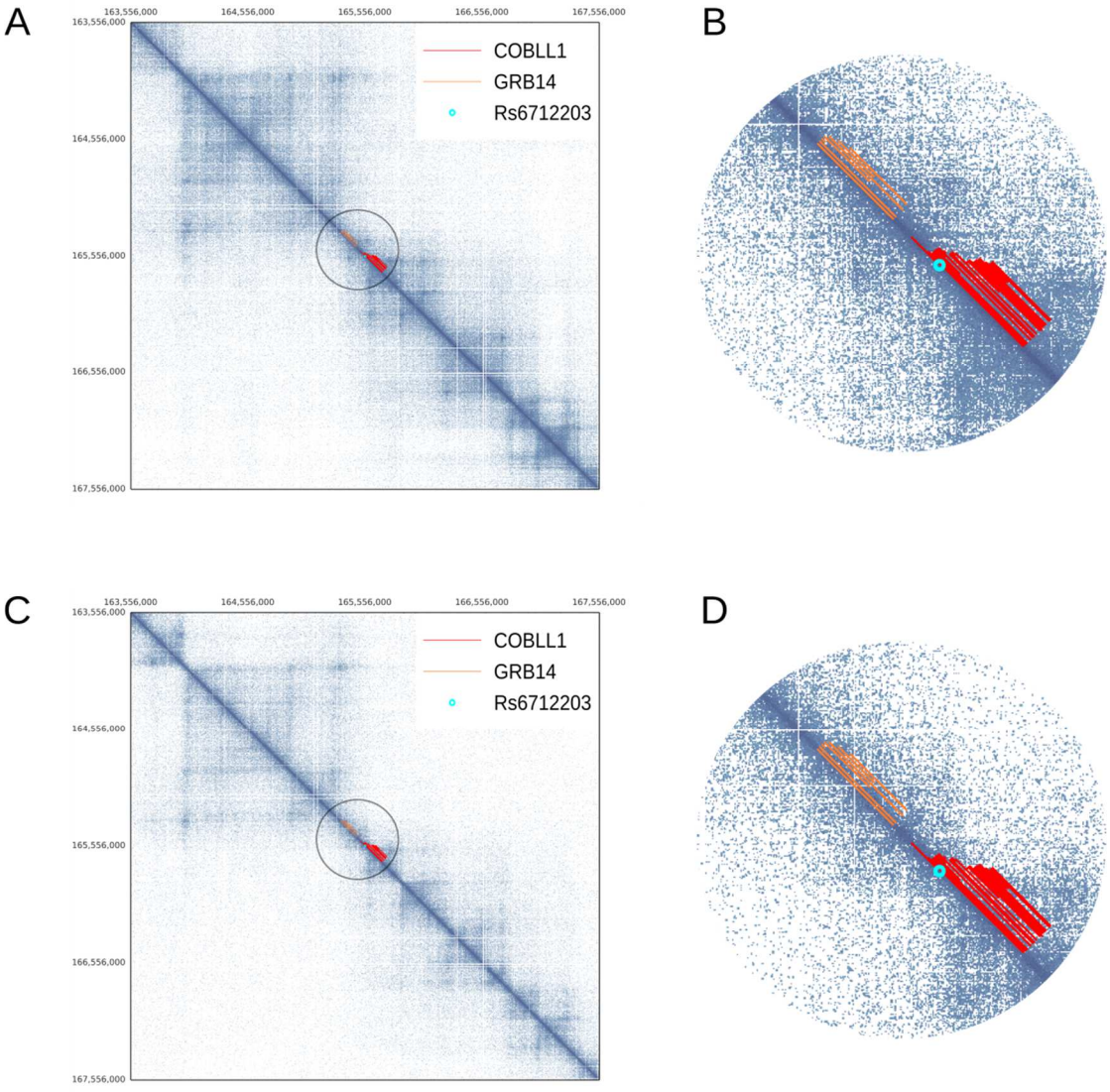


Supplementary figure 7: Linkage disequilibrium of T2D risk variants at the GRB14/COBLL1 locus in African (Panel A), European (Panel B), East Asian (Panel C) and Mixed Americans (Panel D). 1000 Genomes Phase 1 data was derived from haploreg v2; all variants in LD > 0.7 (Europeans) were visualized using the LDheatmap package in R.

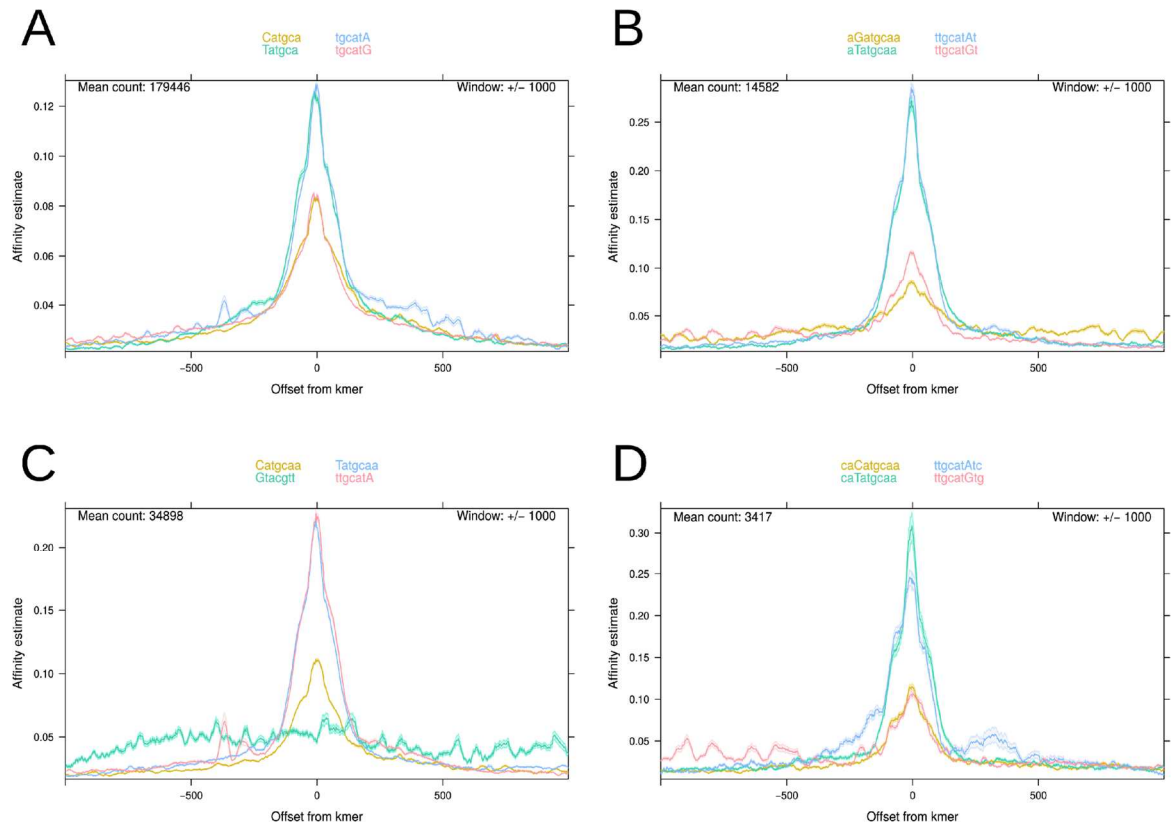


SNP	R2 of proxy to lead SNP	Risk Allele	Non-risk Allele	Risk allele frequency [1000 Genomes Pilote 1 CEU]	type of SNP	PMCA
rs3923113	1	T	G	0,642	INTERGENIC	not complex
rs10179126	0.965	C	G	0,633	INTERGENIC	complex
rs10187501	0.861	A	G	0,625	INTERGENIC	complex
rs6753142	0.861	T	C	0,625	INTRONIC	complex
rs6712203	0.861	C	T	0,625	INTRONIC	complex
rs10184004	0.809	C	T	0,592	INTERGENIC	complex
rs10195252	0.782	T	C	0,583	INTERGENIC	complex
rs13389219	0.771	C	T	0,6	INTERGENIC	not complex

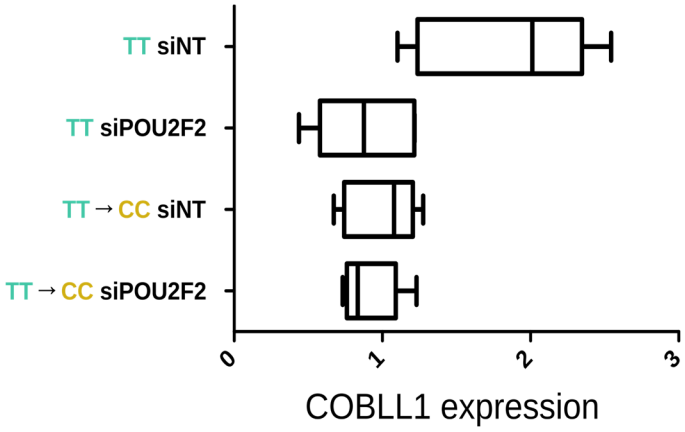
Supplementary figure 8: Schematic location of PMCA candidates' rs6712203, rs10184004, rs10179126, rs10195252, rs10187501, rs6753142 and control GWAS tagging variants rs3923113 and rs13389219. Shown are R2 to rs3923113, risk and non-risk alleles on the + strand of chr2, risk allele frequencies in Europeans (CEU), the location intergenic between GRB14/COBLL1 or intronic of COBLL1 and the PMCA classification in complex, with conserved TFBS and TFBS modules or not-complex (Claussnitzer et al. 2014).



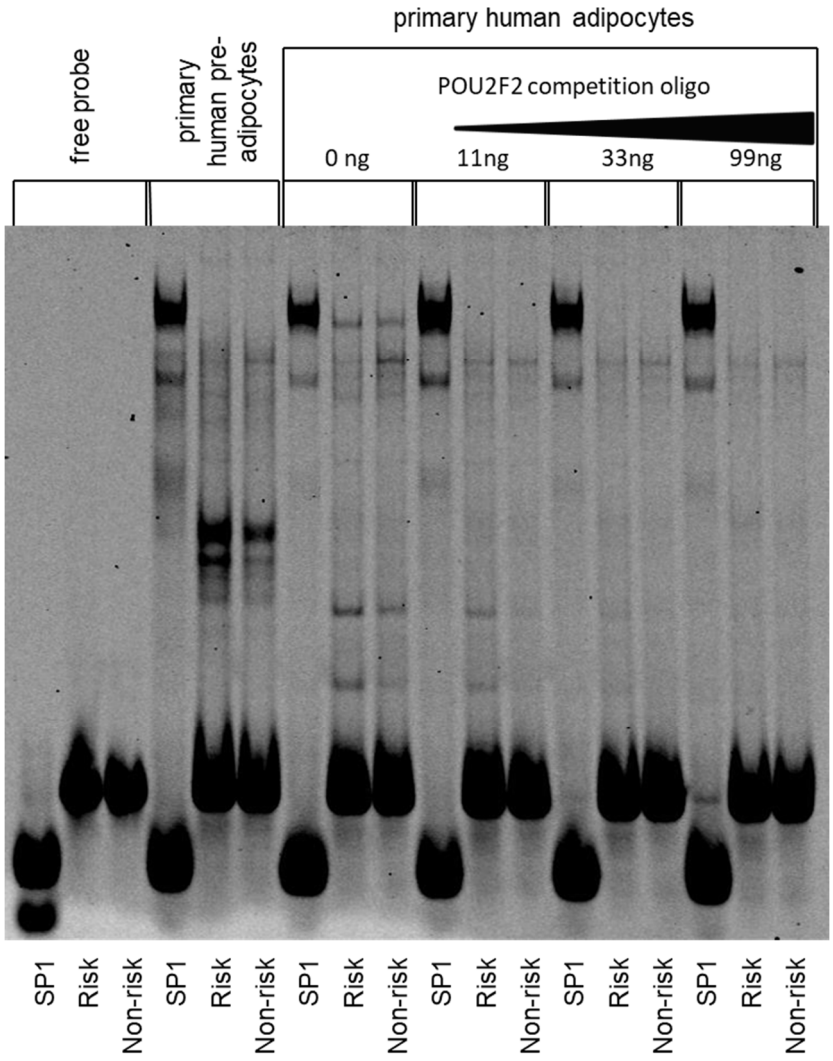
Supplementary figure 9: Putative cis-regulatory effect of rs6712203 on COBLL1. Shown are normalized Hi-C interaction frequencies displayed as a 2D heat-map. Panel A represents Hi-C interaction contact probabilities of human IMR90 myofibroblasts and Panel C NHEK primary normal human epidermal keratinocytes. Displayed in panel A and C, are the genomic regions on chr2: 163 556 000 - 167 558 000 (human genome build hg19), binned at 2kb resolution. For better visibility of the GRB14/COBLL1 area, panel B and C show the circled areas from Panel A and C fivefold increased. The variant rs6712203 is shown as light blue ring, sequencing reads of COBLL1 as red lines and of GRB14 as fuchsia lines. Rs6712203 is located in a boundary region between two topological domains, shown by a drop in interaction frequencies and at a control point between different COBLL1 isoforms, shown by differences in the length of the red lines.



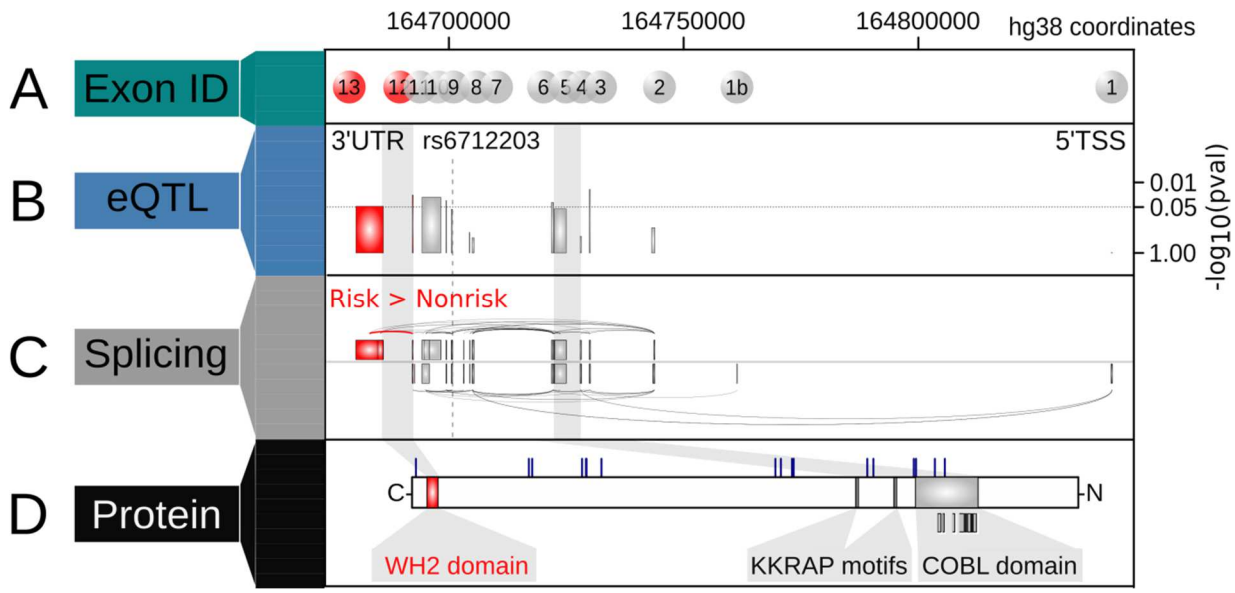
Supplementary figure 10: High binding affinity of POU2F2 to the rs6712203-T non-risk allele. The Figure shows different binding affinity estimate models of POU2F2 to the C risk and T non-risk allele, calculated using the IGR algorithm. Panel A shows the 6-mer model, Panel B the 8-mer model, Panel C the 7-mer model and Panel D the 9-mer model. The IGR affinity modeling reveals a higher affinity of POU2F2 to the rs6712203-T non-risk allele with a baseline scaled difference of -0.38 (9-mer model), when compared to the rs6712203-C risk allele.



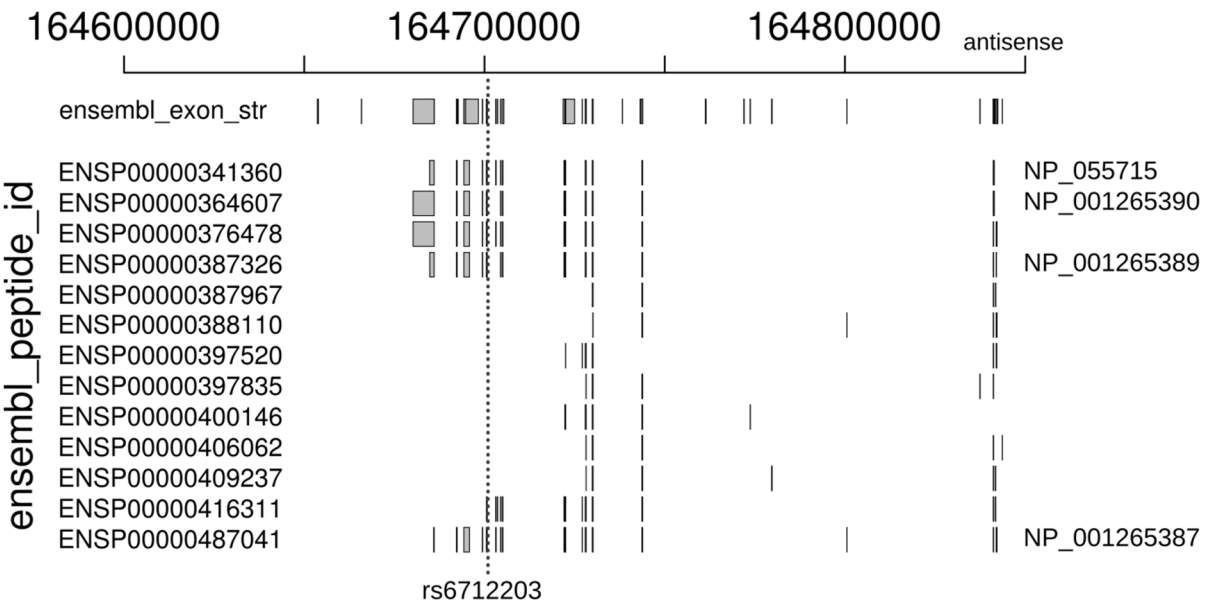
Supplementary figure 11: POU2F2 dependent expression of COBLL1 in rs6712203-T non-risk allele hASC. Shown are boxplots of the relative expression of COBLL1 exons 12 and 13 in the rs6712203 C-to-T (risk to non-risk) editing CRISPR/Cas9 experiment in hASC. The expression of POU2F2 was perturbed using POU2F2 targeting siRNA (siPOU2F2). Non targeting siRNA (siNT) was used as control. The expression of COBLL1 (COBLL1 pair 2) was measured by qRT-PCR.



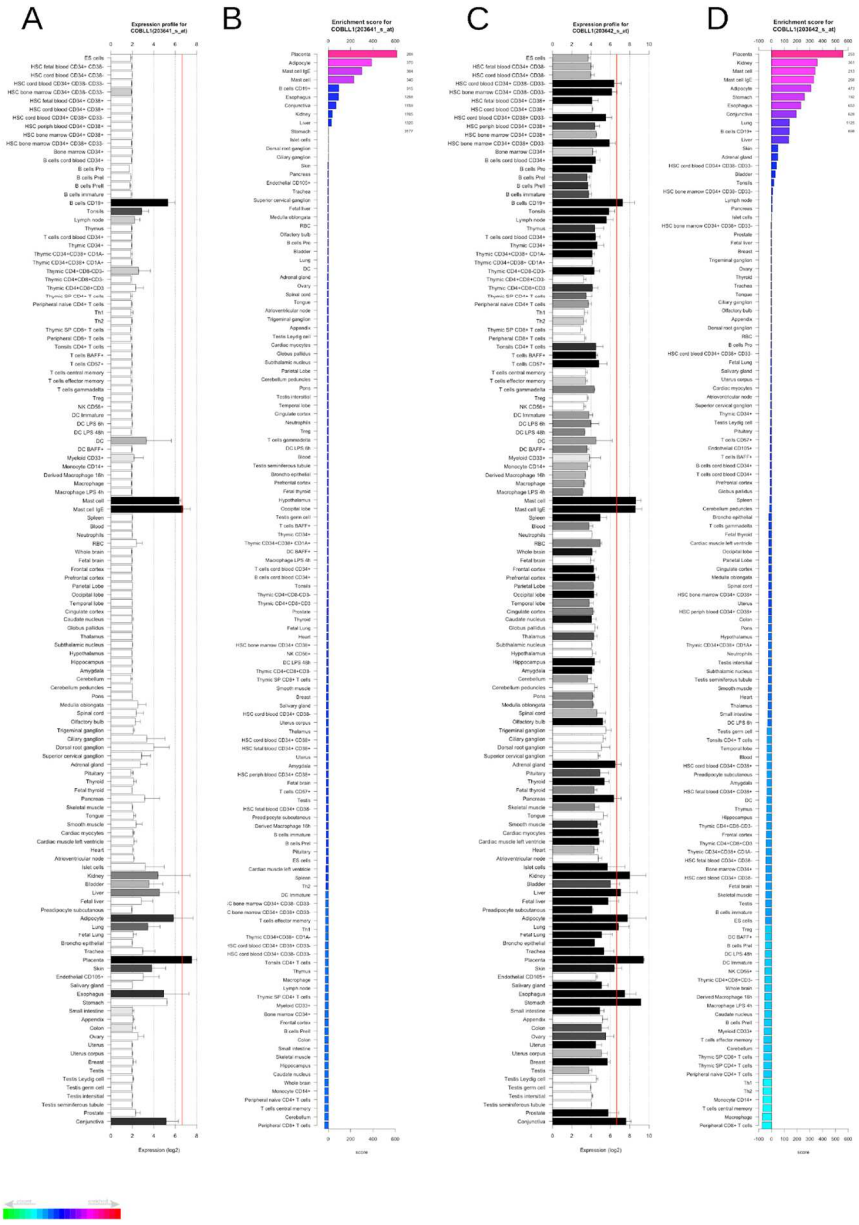
Supplementary figure 12: Differential nuclear adipocyte protein binding at rs6712203-T non-risk and rs6712203-C risk alleles. To assess the binding of POU2F2 to the rs6712203 region, we performed competition EMSA experiments, with rs6712203-T non-risk and rs6712203-C risk allele probes. As additional control we used SP1 probes. The competition was performed using unlabeled oligomer with the octamer DNA motif 5'-ATGCAAAT-3'. We used nuclear extract from hASC and mature hAC.



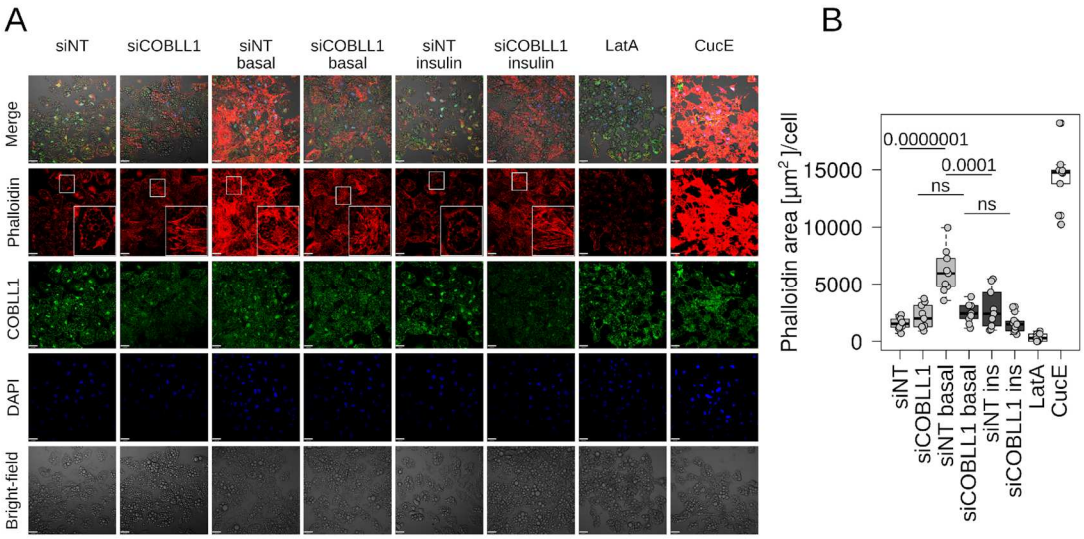
Supplementary figure 13: eQTL and splicing analysis reveals rs6712203-dependent expression of COBLL1-exons. Panel A shows COBLL1 mRNA exon IDs of adipose tissue RNA sequencing data from the EuroBATS cohort. Exon 12 and 13 (marked in red) encode for the WH2 domain of COBLL1. The bar plot in panel B shows eQTL p-values and the genomic localization of COBLL1 exons. The length of the bars represents the exon sizes; their height represents the negative log_{10} whole genome wide corrected p-values. The p-value 0.05 cut-off for significance is indicated as horizontal line. The six exons 3, 6, 10, 11, 12 and 13 were significantly different between rs6712203 risk and non-risk allele carriers. Panel C shows, the results from the relative splicing analysis Altrans (Ongen & Dermitzakis 2015), carried out with EuroBATS subcutaneous adipose tissue RNA sequencing data. Boxes represent exons, which may have alternative 5'- and 3'-exon boundaries and arcs represent detected links between the exons. The middle line separates exons and links which were detected in the forward reads (upper part) and exons and links detected in reverse reads (lower part). Only links, between annotated COBLL1 exons are displayed, links which were detected between COBLL1 exons and other genes or structures were disregarded and are not shown. In the genotype dependent analysis, the exon count for the link between exon 12 and exon 13 (red) of COBLL1 was significantly different between rs6712203 risk and non-risk allele carriers q-value = 0.0003 and β = - 0,21. Q-value represents p-values from an ANOVA of the full model against the model without rs6712203 adjusted for multiple testing. Panel D shows a schematic overview of annotated COBLL1 protein domains. This panel is not in the genomic scale used for the panels A-C. Domains are marked with their respective names. In the COBL-domain additional boxes indicate secondary structures within the domain. Blue lines represent predicted putative phosphorylation sites. Here, shown is the longest in ensembl (Ensembl release 81) annotated COBLL1 protein isoform.



Supplementary figure 14: Overview of ensembl annotated COBLL1 exons and protein isoforms. Genome coordinates were derived from the human reference genome assembly hg38. The first panel shows the COBLL1 exons, following are exons, translated into the 13 in ensembl annotated protein isoforms (downloaded in Mai 2015). On the right-hand side are ensembl ID's and on the right-hand side corresponding (if available) NCBI ID's shown. The location of rs6712203 on chr2 is depicted by a dashed line. We were not able to detect stop codons at each here depicted ORF, which means that some of the isoforms may be falsely annotated.



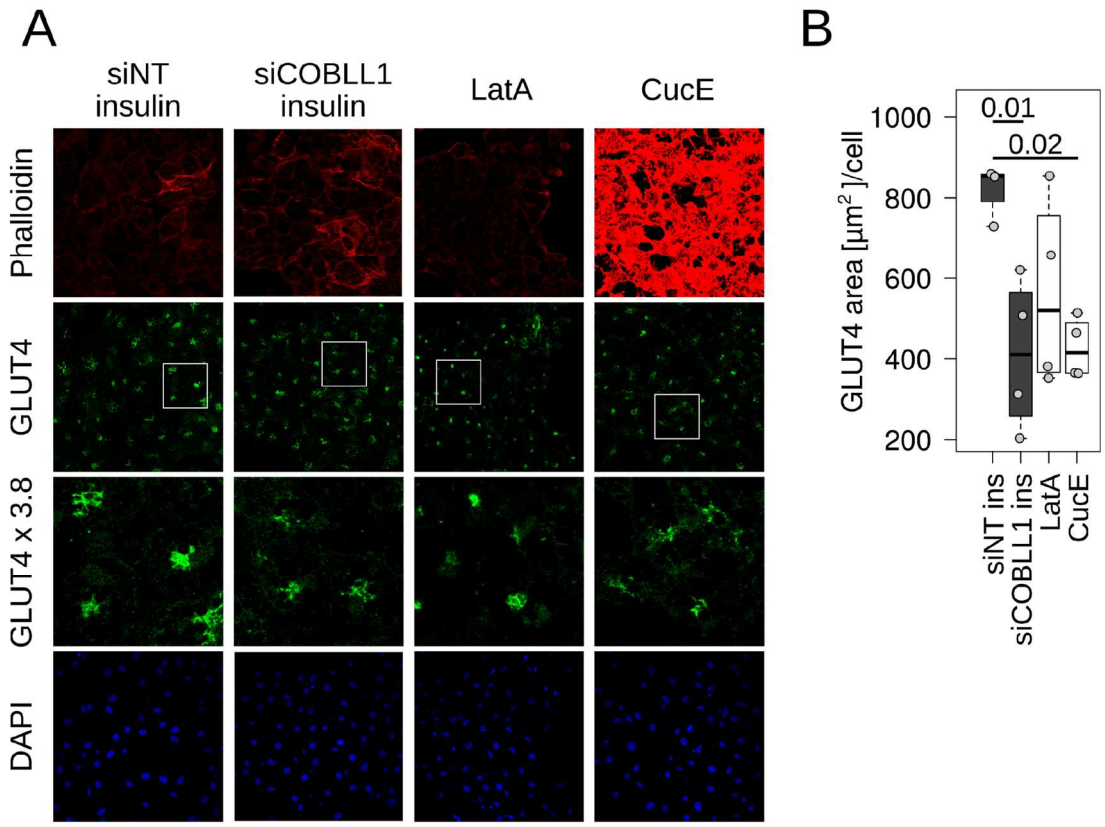
Supplementary figure 15: COBLL1 expression is enriched in adipocytes. The gene enrichment profiler (Benita et al. 2010), a database collecting and analyzing publicly available microarray data profiled on Affymetrix U133A chips, was used for this figure. 142 (16 cancer, 126 normal) raw CEL files of cells or tissues were obtained and normalized; the data was then filtered and processed for enrichment analysis. For COBLL1 the probes 203641_s_at and 203642_s_at were available for analysis. Adipose tissue was for both probes among the five highest ranked tissues enriched for COBLL1 expression. Leptin and ADAM12 were listed as co-regulated genes.



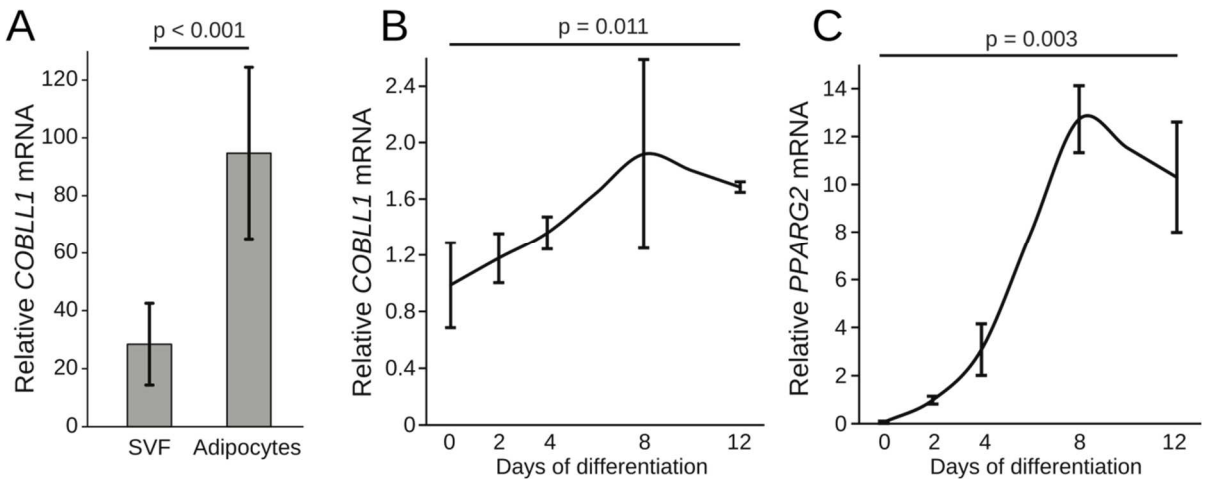
Supplementary figure 16: Actin filament staining in siNT or siCOBLL1 treated in vitro differentiated hASC. Shown are the staining of actin filaments (red), COBLL1 (green), nuclei (blue) and the bright-field of in vitro differentiated hASC. Merged pictures are a combination of pictures from actin filaments, COBLL1, nuclei and bright-field. To investigate the effect of insulin on actin filamentation in siNT or siCOBLL1 adipocytes, the cells were incubated overnight without insulin and glucose (basal) and stimulated with insulin and glucose for 30 minutes (insulin). We used the chemical compounds Latrunculin A (LatA) and Cucurbitacin E (CucE), which deplete and stabilize filamentous actin respectively, as controls for actin filamentation. For the reduction of COBLL1, target specific siRNA was used (siCOBLL1), and control cells were treated with non-targeting siRNA (siNT). For quantification the area in µm² per cell in three independent experiments with 3-4 images each was determined, using the Volocity 3D Image Analysis software version 5.4.1.

Supplementary table 3: Semi-quantification of Lipolysis Western Blots PLIN, ATGL, HSL, pHSL660, pHSL563 and pHSL565. Western Blots were performed with proteins from basal or Isoproterenol and IBMX stimulated shNT and shCOBLL1 SGBS cells (Figure 3F). The intensities of PLIN, ATGL, HSL, pHSL660, pHSL563 and pHSL565 were semi-quantified using the plot lanes tool from ImageJ (Rueden et al. 2017). Each signal is shown in relation to the respective GAPDH intensity.

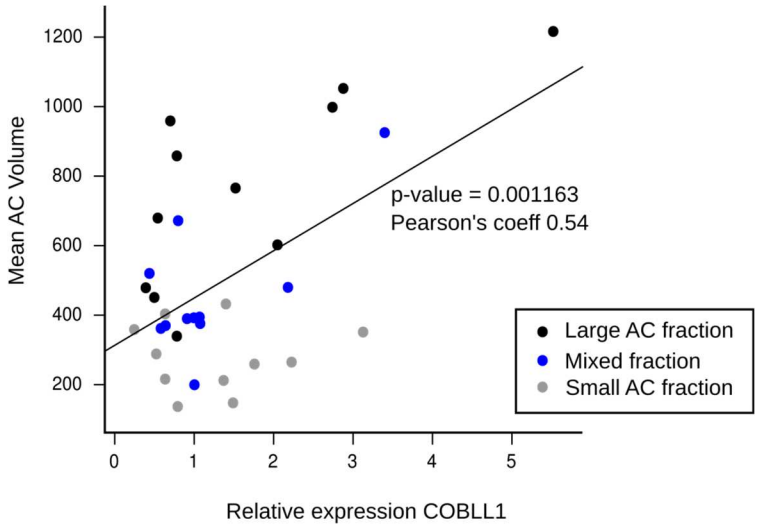
	shNT		**shCOBLL1**	
Iso/IBMX	*-*	*+*	*-*	*+*
PLIN	0,385	1,189	0,256	0,767
ATGL	1,035	0,943	0,551	0,449
HSL	1,006	0,973	0,720	0,564
pHSL660	0,112	1,020	0,030	0,775
pHSL563	0,057	2,423	0,002	1,355
pHSL565	1,087	0,520	0,771	0,328



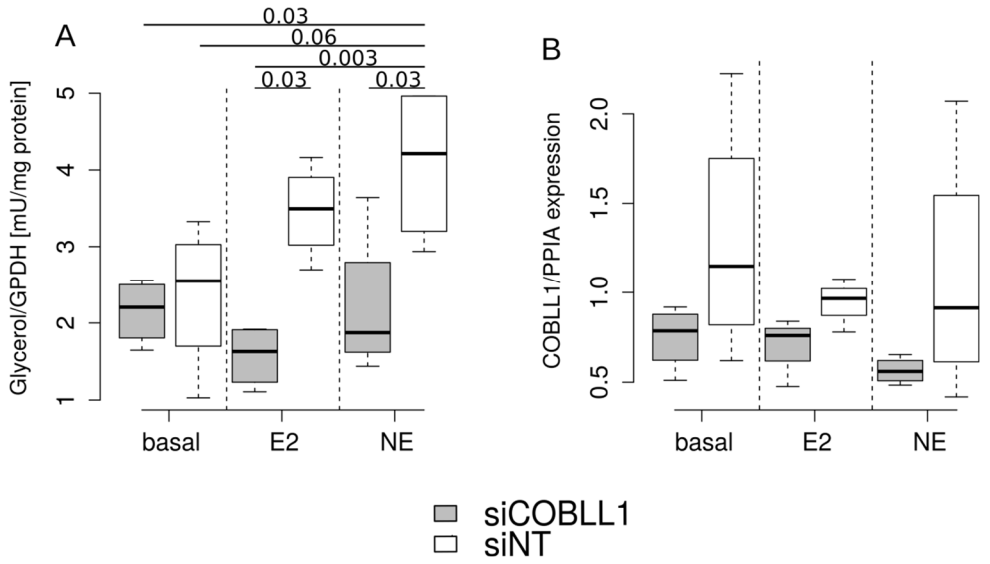
Supplementary figure 17: GLUT4 and actin filament staining in differentiated hASC. Shown are the staining of actin filaments (red), GLUT4 (green), 3.8-fold enlarged GLUT4 and nuclei (blue). To investigate the effect of insulin on actin filamentation in siNT or siCOBLL1 adipocytes, the cells were incubated for 12 hours without insulin and glucose and stimulated with insulin and glucose for 30 minutes (insulin). We used the chemical compounds Latrunculin A (LatA) and Cucurbitacin E (CucE), which deplete and stabilize filamentous actin respectively, as controls for actin filamentation. For the reduction of COBLL1 target specific siRNA was used (siCOBLL1), and control cells were treated with non-targeting siRNA (siNT). For quantification the area in µm² per cell of one experiment with 3-4 images each was determined, using the Volocity 3D Image Analysis software version 5.4.1.



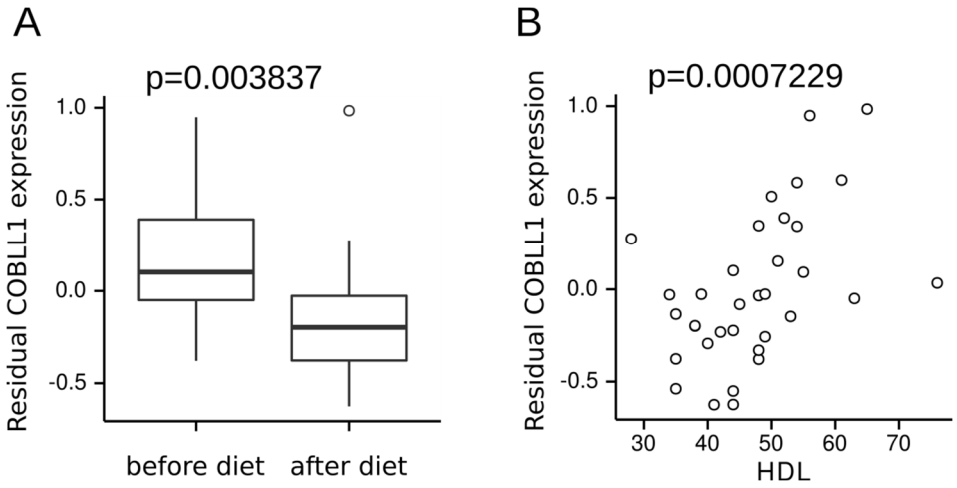
Supplementary figure 18: COBLL1 expression increases during adipogenesis. Shown on the y-axis is the relative expression of COBLL1 (PanelA and B) or PPARG2 (Claussnitzer et al. 2014) (Panel C) measured by qRT-PCR. Panel A shows a barplot with one group of cells from the stromal vascular fraction (SVF) enriched for preadipocytes and one group with adipocytes. COBLL1 is significantly higher expressed in adipocytes. Panel B and C show expression-profiles during in vitro differentiation of hASC. Like PPARG2, COBLL1 increases until day 8 after initiation of differentiation.



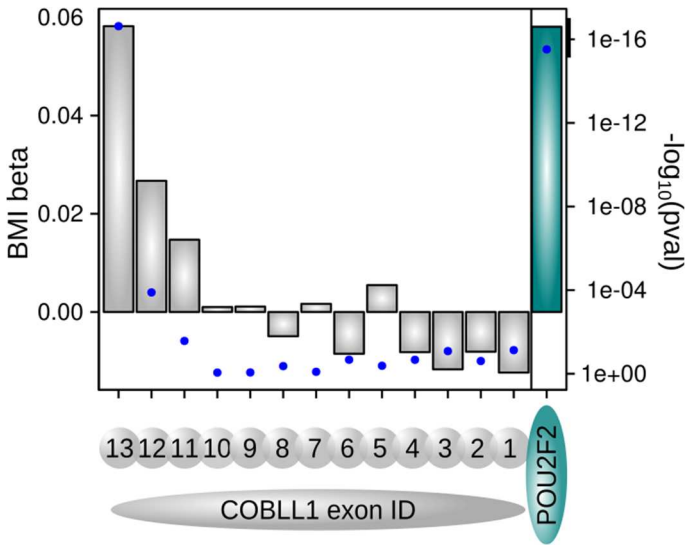
Supplementary figure 19: COBLL1 expression correlates with fat cell volume. Primary hAC from 11 donors were fractionated into small (grey) and large (black) cells or used as a control mixed fraction (blue). The x-Axis shows the relative expression of COBLL1 (COBLL1 pair 2) measured by qRT-PCR and the y-Axis shows the mean fat cell volume per fraction [pl]. The fractionation was used as tool, to ensure a mean fat cell volume with smaller standard deviations in each fraction. The correlation was calculated using Pearson's correlation coefficient.



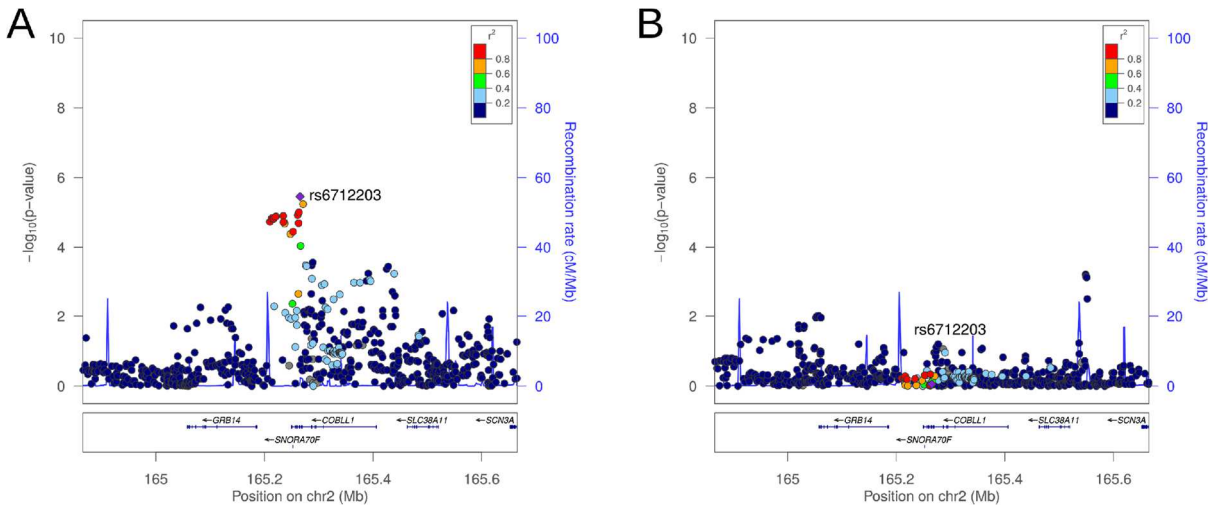
Supplementary figure 20: Stimulated lipolysis is decreased in adipocytes with a transient COBLL1 repression. Data was collected by Johanna Röttgen in the course of her Master Thesis. Glycerol measurements using in vitro differentiated SGBS cells, were normalized with the differentiation marker GPDH in mU per mg protein and assessed under basal, estradiol (E2) or noradrenalin (NA) stimulated conditions. Cells were incubated with COBLL1 siRNA (siCOBLL1) or control non-targeting siRNA (siNT) for 3 days. Reduction of COBLL1 expression was assessed by qRT-PCR.



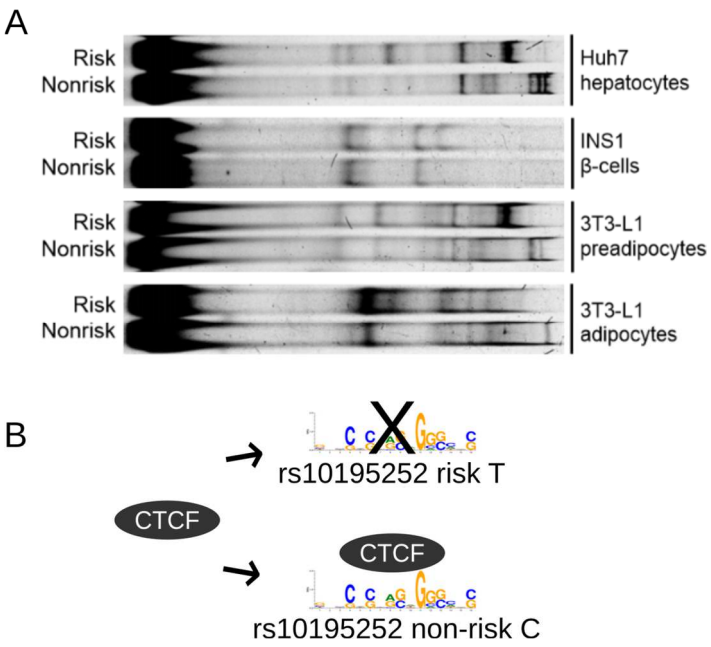
Supplementary figure 21: Caloric restriction reduces COBLL1 expression. Shown are results from a dietary intervention study, performed by Beate Ott at the Lehrstuhl für Ernährungsmedizin (TUM). Relative COBLL1 (COBLL1 pair 2) expression, measured by qRT-PCR was adjusted for BMI, WHR, age and amount of fat in a linear mixed effects model using the lmer package in R as described above. The full fitted model was compared against the null model using ANOVA and p-values were corrected for multiple testing using the QVALUE package. Residual COBLL1 expression is significantly reduced after caloric restriction (Panel A) and correlates with HDL cholesterol levels (Panel B).



Supplementary figure 22: POU2F2 and COBLL1 3'-exons increase with BMI. Shown are β coefficients of COBLL1 exons or POU2F2 with BMI (left y-axis; bars) and q-values (right y-axis; blue dots). The linear mixed effect model was calculated using EuroBATS RNA sequencing data as described above.



Supplementary figure 23: Rs6712203 is associated with BMI adjusted hip circumference in women. For the analysis (October 2014), the publicly released stage 1 GIANT data (Randall et al. 2013) was used, with haplotype information from Hap Map CEU in a display range of chr2:164865564−165665564 (genome build hg18). Shown is, on the left y-axis is the association with BMI adjusted hip measures in women (Panel A) and men (Panel B). The recombination rate (blue) is displayed on the right y-axis. In total 689 variants are plotted, the queried variant is rs6712203 (purple diamond) and variants in LD with rs6712203 ($r^2 > 0.6$) are indicated in red or orange spheres, whereas green and blue variants are not in LD. The p-value in females for chr2:165265564 (rs6712203) is 3.51 x 10^{-6}.



Supplementary figure 24: Differential nuclear protein binding at rs10195252. Panel A shows EMSA binding pattern with T risk and C non-risk allele probes of rs10195252 and nuclear extracts from Huh7 hepatocytes, INS1 β-cells, 3T3-L1 preadipocytes and adipocytes. Panel B shows an in-silico predicted CTCF binding site at the rs10195252-C non-risk allele (the CTCF_disc9 motif was extracted from the motif browser (Kheradpour & Kellis 2014).

Acknowledgement

I would like to express my sincere gratitude to my mentor Prof. Dr. Johann J. Hauner for the opportunity to write this PhD thesis. He continuously enabled an unconstrained working atmosphere, giving room for collaborations around the world and for ideas to grow. His immense knowledge, motivation and patience made this project possible. It was a pleasure and honor to work in his lab.

I would also like to give a big thank you to Prof. Dr. M. Claussnitzer. This PhD thesis would not have been possible without her support and guidance. She inspired me tremendously throughout my career and I am grateful for the opportunity to work with her.

My gratitude also goes to the members of my thesis committee Prof. Dr. M. W. Pfaffl, who introduced me to the “world of genomes” and Prof. Dr. med. H. Witt., who always encouraged a collaborative work environment in the “ZIEL”. Their insightful comments but also hard questions where greatly appreciated and let me to examine my research experiments from various perspectives.

I would also like to thank all collaboration partners. An integral part of this thesis was a collaboration with the Harvard Medical School (Gerontology Division, Beth Israel Deaconess Medical Center and Hebrew Senior Life) and the Broad Institute of MIT and Harvard. Data analysis and interpretations have been contributed, from M. Claussnitzer, A. Abdennur (Hi-C Data) and W. Meuleman (Roadmap Data). Furthermore, we received analysis and data contributions from P. Arner, I. Dahlman, S. Molnos and S. Dankel. I would also like to thank E. Dermitzakis with his team including A. Vinuela and H. Ongen, the EuroBATS consortium (T. Spector (Co-ordinator), King's College London) and the Erwin Riesch Stiftung (Universität Tübingen) for granting me a “Reisestipendium” and enabling a two-month research exchange. Additionally, I would like to thank M. Tschöp, P. Pfluger, K. Stemmer, P. Kotzbeck, F. Zani, and M. Bauer for enabling and helping with *in vitro* experiments. I have learned so much in their labs and loved working with their teams.

I also want to give a big thank you, to all of my former colleagues at the “Lehrstuhl für Ernährungsmedizin”. Thomas, Beate, Sylvia, Kerstin, Caroline, Christoph, Heekyong and many more. Their scientific input, the coffee break discussions, nightshifts and the fun in-between are sourly missed. My gratitude also goes to Helmut who helped to initiate this PhD thesis, and the students I have had the pleasure of working with - Johanna, Nina, Christian and Bahareh. I am also very grateful to Manuela and Elisabeth for their incredible technical support. It would not have been possible to stem the immense workload without their help, knowledge and dedication.

Finally, I would like to thank my precious family, their support cannot be put into words and means the world.