Technische Universität München
TUM School of Life Sciences

Investigation of the pathogenic mechanism underlying the common SPINK1 p.N34S pancreatitis risk haplotype

Lara Sophie Unger

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung einer Doktorin der Naturwissenschaften genehmigten Dissertation.

Vorsitz: Prof. Dr. Johann J. Hauner

Prüfer*innen der Dissertation:

1. Prof. Dr. Heiko Witt
2. Prof. Dr. Hana Algül

Die Dissertation wurde am 15.10.2021 bei der Technischen Universität München

eingereicht und durch die TUM School of Life Sciences am 26.02.2022 angenommen

Acknowledgements

Zunächst möchte ich mich ausdrücklich bei **Prof. Dr. Heiko Witt** bedanken. Ich blicke mit großer Wertschätzung auf lehrreiche Jahre, die neben exzellenter wissenschaftlicher Anleitung auch von außerordentlichem Vertrauen, gegenseitiger Achtung und unterhaltsamen Stunden geprägt waren.

Diesen Dank möchte ich auf die **gesamte Arbeitsgruppe** ausweiten. Gegenseitige Unterstützung, Wertschätzung, Vertrauen und Herzlichkeit, haben zu einer wundervollen Atmosphäre beigetragen, Freundschaften erwachsen lassen und uns viele unvergessliche Stunden beschert! Ein besonderer Dank gilt an dieser Stelle **Dr. Maren Ewers**, die meine Entscheidung für die Wissenschaft und diese Arbeitsgruppe entscheidend geprägt hat.

Zudem möchte ich mich besonders bei **meinen Eltern Ulrike und Jörg Unger** bedanken. Ihre uneingeschränkte Unterstützung, Zuspruch und Liebe sind von unbeschreiblichem Wert und haben mich weit getragen.

Und so möchte ich mich zu guter Letzt bei **all den weiteren lieben Menschen** bedanken, die mir in den letzten Jahren treu zur Seite gestanden sind, mit mir Höhen und Tiefen durchlebt, mir Zeit, Halt und Kraft gegeben haben.

Abstract

In chronic pancreatitis (CP), patients suffer from a relapsing or continuing inflammation of the pancreas. Alcohol abuse presents the major single etiological factor in adults, while genetic risk factors play an especially important role in children. The genetic variant p.N34S (*rs17107315*; c.101A>G) in the serine protease inhibitor Kazal type 1 (*SPINK1*) is the most relevant risk factor for idiopathic CP. SPINK1 inhibits prematurely activated trypsin. It contributes to the fine balance of proteolytic and anti-proteolytic activity within the pancreas and presents an undoubted protective factor against CP initiation and progression. Functional studies excluded a pathological consequence of *SPINK1* p.N34S due to an impaired protein function or the effects of four linked intronic variants. In the European population, *SPINK1* p.N34S is in high linkage disequilibrium with 26 variants. We narrowed our selection down to two final candidates, *rs148276928* (c.1-2090T>C) and *rs148911734* (c.1-7321C>T), in the upstream regulatory region of the gene. We postulated that these variants might affect transcription factor binding and thereby alter *SPINK1* expression. mRNA analysis of Paca44 cells and human intestinal biopsies confirmed a decreased expression of the risk allele in *SPINK1* p.N34S heterozygosity. Based on affinity chromatography and proteomics analysis, we focused on *rs148911734* as the potential pathological variant and GATA6 as the most promising factor and potential transcriptional repressor at this site. However, we were unable to establish a causal relationship of the selected transcription factors and *SPINK1* expression levels within the existing model of the *SPINK1*-expressing cells Caco2 (intestinal) and HepG2 (hepatic). Consequently, we turned towards a new model approach. The overexpression of E47, PTF1a, RBPJL, HNF1a and NR5A2 partially reconstituted an acinar-like phenotype in the PDA cell lines Panc1 and Paca44. Importantly, PTF1a, RBPJL and HNF1a induce *SPINK1* expression. This effect was independent of the previously published *SPINK1* p.N34S linked variant *rs142703147* or the *rs148911734* candidate. We used Paca44 as the adopted model because of its *SPINK1* p.N34S heterozygosity. Functional investigations of the proclaimed inhibitory effect of GATA6 on *SPINK1* expression from the risk allele, however, remained without effect so far. Instead, we discovered two molecular weight forms of GATA6 (90 kDa and 60 kDa). Their expression ratio was dependent on the endogenous or exogenous expression.

Further, endogenously expressed GATA6 showed DNA binding in EMSA, while the DNA binding ability was lost in exogenously expressed GATA6. We suspected posttranslational modification by SUMOylation and constructed GATA6 p.K484R to remove potential SUMO residues at this site. However, the molecular size, the DNA binding ability and the target gene activation remained unaffected. Future research is underway to investigate the nature of the different molecular weight forms of GATA6 in more detail and to elucidate their interaction with *rs148911734* and *SPINK1* expression in the context of the CP-associated *SPINK1 p.N34S* haplotype.

Zusammenfassung

Die chronische Pankreatitis (CP) ist eine rezidivierende oder kontinuierliche Entzündung der Bauchspeicheldrüse. Alkoholabusus ist bei Erwachsenen der wichtigste ätiologische Faktor, während bei Kindern genetische Faktoren bedeutend sind. Die Variante p.N34S (rs17107315; c.101A>G) im Serin-Protease-Inhibitor Kazal Typ 1 (*SPINK1*) ist der stärkste und relevanteste Risikofaktor für eine idiopathische CP. SPINK1 hemmt vorzeitig aktiviertes Trypsin und trägt zum feinen Gleichgewicht zwischen proteolytischer und antiproteolytischer Aktivität innerhalb der Bauchspeicheldrüse bei. Während die physiologische Rolle von SPINK1 als Schutzfaktor gegen die Entstehung und das Fortschreiten einer CP unbestritten ist, wurden eine gestörte Proteinfunktion der SPINK1 p.N34S Variante sowie der Effekt von vier gekoppelten intronischen Varianten ausgeschlossen. In der europäischen Population tritt *SPINK1* p.N34S mit 26 Varianten in fast vollständiger Kopplung auf. Unsere Untersuchungen grenzten die mit *SPINK1* p.N34S gekoppelten 5‘ gelegenen Varianten auf zwei Kandidaten ein: *rs148276928* (c.1-2090T>C) und *rs148911734* (c.1-7321C>T). Wir gingen von der Hypothese aus, dass diese Varianten die Bindung von Transkriptionsfaktoren beeinflussen und dadurch die *SPINK1* Expression verändern könnten. Durch mRNA Analysen von Paca44 Zellen und menschlichen Darmbiopsien konnten wir eine verminderte Expression des Risikoallels bei *SPINK1* p.N34S Heterozygotie bestätigen. Auf der Grundlage von Affinitätschromatographie und Proteomanalyse konzentrierten wir uns auf *rs148911734* als mögliche pathologische Variante und GATA6 als vielversprechendsten Faktor und potenziellen Transkriptionsrepressor an dieser Stelle. Im Rahmen des bestehenden Modells der *SPINK1* exprimierenden Zellen Caco2 (Darm) und HepG2 (Leber) konnte jedoch kein Effekt der ausgewählten Transkriptionsfaktoren auf die *SPINK1* Expression nachgewiesen werden. In einem neuen Modellansatz rekonstruierten wir einen azinusähnlicher Phänotyp in den PDA-Zelllinien Panc1 und Paca44 durch die Überexpression von E47, PTF1a, RBPJL, HNF1a und NR5A2. Die Überexpression von PTF1a, RBPJL und HNF1a erhöhte die *SPINK1* Expression. Dieser Effekt war unabhängig von der zuvor veröffentlichten und mit *SPINK1* p.N34S gekoppelten Variante *rs142703147* oder unserem Kandidaten *rs148911734*. Aufgrund der *SPINK1* p.N34S Heterozygotie haben wir Paca44 als weiteres Modell verwendet. Funktionelle Untersuchungen der

vermuteten hemmenden Wirkung von GATA6 auf die *SPINK1* Expression durch das Risikoallel blieben bislang jedoch ohne Ergebnis. Stattdessen entdeckten wir zwei Molekulargewichtsformen von GATA6 von ca. 60 und 90 kDa, deren Verhältnis von der endogenen oder exogenen Expression abhängig war. Während endogenes GATA6 in EMSA Experimenten eine DNA-Bindung zeigte, ging die DNA-Bindungsfähigkeit bei exogen exprimiertem GATA6 verloren. Wir vermuteten eine Proteinmodifikation der 90 kDa-Form und mutagenisierten eine potenzielle SUMOylierungsstelle (p.K484R). Allerdings unterschieden sich die Molekülgröße, die DNA-Bindungsfähigkeit und die Zielgenaktivierung nicht zwischen dem GATA6 Wildtyp und p.K484R. Unsere aktuellen Untersuchungen konzentrieren sich daher auf die Ursache der verschiedenen Molekulargewichtsformen von GATA6 und deren Zusammenhang mit *rs148911734* und der *SPINK1* Expression, um die Pathogenese des CP-assoziierten *SPINK1* p.N34S-Haplotyps weiter zu entschlüsseln.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	Adenine, Ampere
AC	Affinity chromatography
ACMS	Affinity chromatography mass spectrometry
ACP	Alcoholic Chronic Pancreatitis
ADM	Acinar-to-ductal metaplasia
AF	Allele Frequencies
Agp	Aquaporin
Amy	Amylase
APS	Ammonium persulfate
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
bp	base pairs
BBQ	Black Berry® Quencher, absorption wavelength 550 nm-750 nm, centered at 650 nm
BB	Binding Buffer
bHLH	basic helix-loop-helix
BHLHA15	Basic helix-loop-helix family member a15 (MIST1)
BiP	Immunoglobulin-binding protein
BSA	Bovine serum albumin
c	Centi
C	Cytosine
CCK	Cholecystokinin
cDNA	complementary DNA
CEL	Carboxyl ester lipase
CELA3A	Chymotrypsin-like elastase family member 3A
CELA3B	Chymotrypsin-like elastase family member 3B
CELA3AB	Chymotrypsin-like elastase family member 3A and 3B
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
ChiP	Chromatin immune precipitation
CHOP	C/EBP Homologous Protein
Chrdl2	Chordin like 2

CLDN2	Claudin-2
CLPS	Colipase
CoF	Co-factor
CP	Chronic Pancreatitis
Cpa/Cp	Carboxypeptidase
CPA1	Carboxypeptidase A1
Ct	Cycle threshold
CTRB	Chymotrypsinogen B
CTRC	Chymotrypsinogen C
CTRL	Chymotrypsin like
CUZD1	CUB and zona pellucida like domains
Da	Dalton
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
ds	Double stranded
DSMZ	German Collection of Microorganisms and Cell Cultures GmbH
DTT	Dithiothreitol
E47	E2A immunoglobulin enhancer-binding factors E12/E47 (*TCF3*)
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetra acetic acid
Ela	*Elastase*
ELISA	Enzyme Linked Immuno Sorbent Assay
EMSA	Electron Mobility Shift Assay
ER	Endoplasmic Reticulum
FBS	Fetal Bovine Serum
FC	Fold-change
FCPD	Fibrocalculous pancreatic diabetes
FDR	False discovery rate
FL	Fluoresceine
g	Gram, gravitational force
G	Guanin
Gal	Galanin and GMAP prepropeptide
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase

Gatm	Glycine amidinotransferase
Ghrelin	Growth Hormone Release Inducing
Gls	Glutaminase
Gpsm1	G-protein signaling modulator 1
GSK3	KT–glycogen synthase kinase 3
GTex	Genotype-Tissue Expression
GWAS	Genome Wide Association Study
HaPanEU	Harmonizing Pancreatitis across Europe
HEB	HeLa E-box binding protein
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEYL	Hairy/enhancer of split related with YRPW motif-like protein
HNF1a	Hepatocyte nuclear factor 1 homeobox A
HNF1b	Hepatocyte nuclear factor 1 homeobox B
HP	Hereditary Pancreatitis
ICP	Idiopathic Chronic Pancreatitis
ID3	Inhibitor of differentiation 3
Ifed1	Interferon related developmental regulator 1
IgG	Immunoglobulin G
kbp	Kilo base pairs
kDa	Kilo Dalton
Klk1	Kallikrein 1
LC	Light Cycler (Red)
L	Liter
LB	Lysogeny broth
LC3-II	Microtubule-associated protein light chain 3
LD	Linkage disequilibrium
m	Mili; meter
M	Molar
mAb	Monoclonal antibody
M-ANNHEIM	Multiple, Alcohol, Nicotine, Nutrition, Heredity, Efferent duct, Immunology, Miscellaneous
MEM	Minimum Essential Medium
Mgat4a	Mannoside acetylglucosaminyltransferase 4, isoenzyme A

mRNA	Messenger RNA
MODY	Maturity-onset diabetes of the young
MS	Mass spectrometry
n	Nano
NA	Not applicable
NE	Nuclear Extracts
NEAA	Non-essential amino acids
NEB	New England Biolabs
NIDDM	Non-insulin dependent diabetes mellitus
NP40	Nonyl Phenoxypolyethoxylethanol
NR5A2	Nuclear receptor subfamily 5 group A member 2
OR	Odds Ratio
PBS	Phosphate buffered saline
PAGE	Polyacrylamide-Gel-Electrophoresis
PCR	Polymerase Chain Reaction
PDA	Pancreatic Ductal Adenocarcinoma
PDX1	Pancreatic and duodenal homeobox 1
PMCA	Phylogenetic Module Complexity Analysis
PMSF	Phenylmethylsulphonyl fluoride
PNLIP	Pancreatic Lipase
PNLIPRP2	Pancreatic Lipase Related Protein 2
polydIdC	Poly-deoxy-inosinic-deoxycytidylic acid sodium salt
PP	Pancreatic polypeptide
PRSS1	Cationic Trypsinogen
PRSS2	Anionic Trypsinogen
PRSS3	Mesotrypsinogen
PTF1	Pancreas transcription factor 1
PTF1-L	Pancreas transcription factor 1 with RBPJL
PTF1a	Pancreas associated transcription factor 1a
qPCR	semi-quantitative Polymerase Chain Reaction
Rap1gap	Rap1 GTPase-activating protein
RBPJ	Recombination signal binding protein for immunoglobulin kappa J
RBPJL	RBPJ-like
Rbm38	RNA binding motif protein 38

REG1A	Regenerating family member 1 alpha
RIPA	Radio-immunoprecipitation assay
RNA	Ribonucleic acid
RT	Room Temperature
RX	Nuclear protein from cell culture extracts in the AC experiments
SDS	Sodium Dodecyl Sulfate
siRNA	small interfering RNA
SOC	Super Optimal broth (SOB) with catabolite suppression (glucose)
SPINK1	Serine protease inhibitor Kazal type 1
SN-cplx	Clear supernatant including all unbound proteins of the AC
SNP	Single Nucleotide Polymorphism
ss	Single -stranded
Sync	Syncollin
SYTL1	Synaptotagmin like 1
T	Thymidine
TAE	Tris base, acetic acid and EDTA
TBE	Tris base, boric acid and EDTA
TC	Tissue culture
TCF3	*Transcription factor 3* (E47)
TCP	Tropical calcific pancreatitis
Tdh	L-threonine dehydrogenase
TE	Tris and EDTA
TEMED	Tetramethyl ethylenediamine
TF	Transcription Factor
TFBS	Transcription Factor binding site
TIGAR-O	Toxic-Idiopathic-Genetic-Autoimmune-Recurrent, severe AP associated CP-Obstructive etiologic factors
Tmed11	Transmembrane p24 trafficking protein 11
Trp	Trypsin
TRPV6	Transient receptor potential cation channel subfamily V member 6
V	Volt
VNTR	Variable Number of Tandem repeats
v/v	Volume in volume

WB	Washing Buffer
WxH	Width x height
w/v	Weight in volume
XBP1s	Spliced form of X-box binding protein-1
µ	Micro
6-FAM	6-Carboxyfluorescein, fluorescent dye absorption wavelength at 495 nm, emission wavelength at 517 nm

1 Introduction

1.1 The pancreas: cellular compartments and organization

The cellular organization and compartmentalization of the pancreas mirror its endocrine and exocrine function.

The pancreas (in Greek pánkreas, pán - „all“, kréas - „flesh“) is located in the center of the upper abdomen. Its head is positioned in the duodenal loop where the major pancreatic duct enters the small intestine. The organ harbors both exocrine and endocrine capacities which together effectively orchestrate nutrient digestion and distribution. Endocrine islets are embedded within the exocrine tissue and constitute of alpha-, beta-, delta-, PP- and epsilon-cells. The largest cell population within the islet are the beta-cells which secrete insulin, followed by alpha-cells, which secrete glucagon. Both hormones contribute to blood glucose homeostasis and utter opposite functions. While insulin lowers the blood glucose level, glucagon increases it. The third largest population, delta-cells, secrete somatostatin and thereby exert an inhibitory effect on beta- and alpha-cells but also attenuate the release of cholecystokinin (CCK) and secretin from the duodenum. PP-cells and epsilon cells make up a minor fraction of the cellular islet population. While the pancreatic polypeptide (PP) from PP-cells is involved in the regulation of satiety, the secretion of ghrelin (short for *Growth Hormone Release Inducing*) from epsilon cells stimulates the feeling of hunger. The exocrine tissue consists of acinar and ductal cells. Upon nerval (Nervus Vagus, acetylcholine) innervation and hormonal (CCK) stimulation, acinar cells produce and secrete digestive enzymes. Stimulation of the ductal cells, on the other side, is predominately mediated through secretin which triggers the release of bicarbonate and water into the pancreatic duct. The latter helps the formation and flushing of the pancreatic juice into the duodenum. To protect the pancreas from intra-organic digestion, all digestive enzymes, including trypsinogen, are synthesized, and secreted as inactive precursors (zymogens). Trypsinogen turns into its catalytically active form trypsin upon the release of the activation peptide. Under physiological conditions, this process is mediated by the enteropeptidase in the duodenum and to a minor extent by trypsin itself (autoactivation). The active trypsin stands on top of the zymogen activation cascade

and carries out the activation of all other digestive enzymes (Rinderknecht *et al*., 1986).

1.2 Chronic Pancreatitis: Epidemiology and Etiology

CP is a disease of various etiological factors with a strong genetic component in children.

Chronic Pancreatitis (CP) in adults is characterized by a relapsing or continuing inflammation of the pancreas accompanied by irreversible morphological changes, pain, and impairment of the exocrine (*e.g.,* maldigestion) and/or endocrine (*e.g.* Diabetes mellitus) function (Witt & Becker, 2002). Worldwide, the CP prevalence ranges from 36-154 per 100,000 individuals. While the prevalence seems to underlie some regional variety, reports for the annual incidence are rather consistent around the globe and range from approximately 5 to 12 cases per 100,000 per year. Up to date, there is no uniform classification system for CP. Following the TIGAR-O classification system, CP is categorized based on the following risk factors: Toxic-metabolic, Idiopathic, Genetic, Autoimmune, Recurrent and severe acute pancreatitis-associated CP, and Obstructive etiologic factors. A second way of classification is the M-ANNHEIM (short for Multiple factors including Alcohol, Nicotine, Nutrition, Heredity, Efferent duct, Immunology, Miscellaneous) classification, which also takes multiple etiological factors into account (Conwell *et al*., 2014, Löhr *et al*., 2017). Regardless of the classification system, though, alcohol abuse is the major single etiological factor and accounts for approximately 45 – 75% of all CP cases in the USA and Europe. In China and India, idiopathic CP is more common (60 - 80%) and alcohol abuse plays a secondary role (Conwell *et al*., 2014**,** Löhr *et al*., 2017**,** Olesen *et al*., 2021**,** Hegyi *et al.*, 2020). Genetic mutations present a further important etiological factor. They play an especially important role in children since alcohol abuse can be neglected as a relevant contributor at that age (Witt & Becker, 2002). Correspondingly, alcoholic CP is diagnosed between the ages of 40 to 60 years while patients with genetic mutations typically present in the first four decades of life, many of them even before the age of 10. Based on a recent longitudinal study from Denmark, not only the overall but also the incidence for alcoholic CP peaked in the group of the 50 to 59 years old. The lowest number of overall CP cases was observed in the group from age 0 to 29 years with an

incidence rate of 1.1 per 100,000 person years (Olesen *et al*., 2021). According to the latest report of the Harmonizing Pancreatitis across Europe (HaPanEU) working group, the CP incidence for children was even reported as high as 4 - 13 in 100,000 (Löhr *et al*., 2017). Based on personal observations of Prof. Dr. Heiko Witt, among German pediatric CP patients, idiopathic and hereditary causes account for 40 - 50% and 20 - 25%, respectively. Other common etiological factors in children are anatomical abnormalities, traumata, inflammatory bowel disease or cystic fibrosis (Uc *et al*, 2019). For CP diagnosis in pediatric patients, cystic fibrosis needs to be ruled out. Based on investigation of Durno *et al*., 1-2% (19/1075) of all cystic fibrosis patients suffer from recurrent attacks of acute pancreatitis. In the group of cystic fibrosis patients with pancreatic sufficiency, the proportion is even higher and 10-15% (19/138) suffer from attacks of pancreatitis (Durno *et al*., 2002). In addition, genetic testing of associated variants in all patients with a family history or early onset disease (< 20 years) is highly recommended since genetic causes seem to be of much greater importance in children than in adults (Löhr *et al*., 2017). A recent review on pancreatitis in children further highlights the difference between pediatric and adult disease etiology. While environmental factors such as alcohol and smoking make up a significant contribution in adult patients, in pediatric patients, genetics and anatomic abnormalities seem to be the leading factors. For instance, while alcoholic CP is diagnosed in 45 - 75% of the adult cases, about 75% of the pediatric CP patients can be related to a genetic origin (Uc *et al*, 2019).

1.3 Genetic variants in CP: original discoveries and related disease models

The CP risk genes and related variants group into different categories based on cellular expression, functional consequence, and pathological model.

The genetic nature of pancreatitis was first described in 1952 by Comfort and Steinberg. They recorded the pedigree of a family with chronic relapsing pancreatitis and recognized the autosomal dominant inheritance pattern of the disease (Comfort & Steinberg, 1952). In the meantime, hereditary pancreatitis (HP) is characterized as a rare genetic disorder of the pancreas with about 70-80% penetrance. Clinically, HP is indistinguishable from pancreatitis of other etiologies. The only exception is that HP typically presents at a very young age. Starting with acute attacks in early adolescences, HP progresses from acute to recurrent acute attacks and eventually

to chronic disease already in early adulthood. Further, HP imposes a major risk factor for the later development of pancreatic ductal adenocarcinoma (PDAC) (Shelton & Whitcomb, 2016).

Over the course of the years, several risk genes and their related variants have been identified. The categorization of risk genes follows distinct premises based on *i.e.,* cellular expression (acinar *vs*. ductal), functional consequences of variants (gain- *vs*. loss-of-function) and contribution to the underlying disease mechanism. Two major pathological models were established so far: the trypsin-dependent and the misfolding-dependent pathway. Both models are well summarized in two recent overview articles from 2017 (Hegyi & Sahin-Tóth, 2017, Sahin-Tóth, 2017). The two models were pre-dominantly built on the cellular defects of acinar genes while ductal genes take a secondary role. An overview of the current CP models and involved risk genes is provided in Figure 1.

1.3.1 Genetic variants in acinar genes: the trypsin-dependent model

As early as in the year 1896, Hans Chiari postulated that pancreatitis resulted from an inappropriate activation of digestive enzymes and intra-pancreatic autodigestion (Chiari, 1896). Definite proof was only evident 100 years later when Whitcomb *et al.* identified the underlying genetic defect in the cationic trypsinogen (*PRSS1*): The p.R122H gain-of-function mutation leads to increased intra-pancreatic autoactivation of trypsinogen (Whitcomb *et al*., 1996). Elevated levels of trypsin and the fine balance of proteolytic and anti-proteolytic balance henceforth took up a central role and lead to trypsin-dependent pathological model of CP: While a low degree of trypsinogen autoactivation within the acinar cells is well tolerated, sustained trypsinogen autoactivation initiates a fatal cascade. Under physiological conditions, two major protection mechanisms exist. A first line of defense forms the serine protease inhibitor Kazal type 1 (SPINK1). The protein binds to prematurely activated trypsin and removes it from the autoactivation cascade by inhibition (Pubols *et al*., 1974). A second line of defense provides the degradation of trypsinogen by CTRC (referred to as enzyme Y by Rinderknecht). CTRC cleaves the peptide bond Leu^{81}-Glu^{82} much faster in trypsinogen than in trypsin. Hence, the protective action of CTRC exists in the removal of trypsinogen from the autoactivation cascade (Zhou *et al*., 2011**)**. The p.R122H mutations however

renders trypsin and trypsinogen resistant against such degradation (Whitcomb *et al*., 1996). With trypsin standing on the top of the zymogen activation cascade, increasing levels of active trypsin not only fuel the autoactivation of trypsinogen itself but also the activation of all trypsin-dependent zymogens. Consequently, a highly proteolytic and digestive cocktail results within the acinar cells that drive the organ towards autodigestion and inflammation.

The subsequent years, genes in the context of the trypsin-dependent mechanism of CP shifted into the focus of potential CP susceptibility genes. In fact, in the year 2000 Witt *et al.* associated variants in *SPINK1* with CP (Witt *et al*., 2000). The variant p.N34S showed the strongest association. A reasonable approach suggested a defect in trypsin inhibitory activity at first, the pathological role, however, remains enigmatic until today. While *PRSS1* p.R122H represents the strongest risk factor for HP, the *SPINK1* p.N34S imposes the largest and clinically most relevant risk factor for idiopathic chronic pancreatitis (ICP) (Witt *et al*., 2000). Strikingly, the *PRSS1* p.R122H presents a clear biochemical effect and cellular impact while the *SPINK1* p.N34S remains biochemically unaffected, and the functional consequences persist as a matter of current research.

Although less frequent, the role of chymotrypsinogen C (*CTRC*) variants, is functionally elucidated. In 2007, CTRC was identified to represent the formerly by Rinderknecht proposed enzyme Y. CTRC degrades all three human isoforms of trypsinogen (Szmola & Sahin-Tóth, 2007). Rosendahl *et al.* associated loss-of-function variants of the gene with CP (Rosendahl *et al*., 2008). A decreased inhibitory function on trypsin degradation was expected to facilitate sustained trypsinogen autoactivation and trypsin-dependent CP pathogenesis. The two variants p.R254W and p.K247_R254del both lead to severely diminished secretion levels of CTRC and thus explain the reduced inhibitory activity. While p.R254W was still functionally intact, p.K247_R254del was completely inactive. Moreover, the complex and multigenic nature of CP becomes evident upon the co-investigation of *PRSS1* and *SPINK1* variants in the same individuals. In the HP subgroup of all investigated German subject, 22.4% and 21% carried a *PRSS1* mutation or the *SPINK1* p.N34S variant, respectively, in a heterozygous state. In the ICP subgroup, however, only 2.6% of the patients were heterozygous carriers of a *PRSS1* mutation

while 12% carried the *SPINK1* p.N34S variant. The proportion of homozygous p.N34S carriers was measured at 1.4% and 2% in the HP and ICP subgroup, respectively, and none of the affected individuals carried a *CTRC* variant. In turn, 7.4% of the p.N34S heterozygotes also carried either the *CTRC* p.R254W or p.K247_R254del variant (Rosendahl *et al*., 2008). Interestingly, the analysis of a yet smaller cohort of tropical calcifying pancreatitis (TCP) patients revealed *CTRC* p.A73T as the most prevalent variant in the affected subjects (5.6%). The odds ratio (OR) of the most prevalent variants in the German ICP and HP cohort, p.R254W, accounted for an increased risk of 3.3 while the second most prevalent variant, p.K247_R254del, presented an increased risk of 11.5. An almost equally high risk was determined for the p.A73T variant in the TCP cohort with an OR of 11.3 (Rosendahl *et al*., 2008).

In contrast to the disease-associated variants of *PRSS1*, *SPINK1* and *CTRC*, the p.G191R variant of the anionic trypsinogen (*PRSS2*) was overrepresented in control subjects and thus categorized as a protective variant. The p.G191R induces high autoproteolytic activity through which PRSS2 loses all tryptic activity. Hence, the *PRSS2* p.G191R represent a loss-of-function mutation and stands in clear contrast to the *PRSS1* p.R122H gain-of-function mutations (Witt *et al*., 2006). It is noteworthy that despite a 90% identity between PRSS1 and PRSS2, a lack of PRSS2 association with CP can be explained by a tighter control through CTRC. An additional cleavage site and a lacking disulfide bond render PRSS2 more susceptible towards CTRC mediate degradation and thus mitigate the risk of fatal autoactivation events (Jancsó *et al*., 2016).

Besides the consequences of biochemical defects, a distinct cause of proteolytic imbalance was brought into play when a Genome Wide Association Study (GWAS) identified a variant in the *PRSS1–PRSS2* locus (*rs10273639*). The variant is in the proximal promoter region of *PRSS1* (c.1-408T>C) and was expected to confer the disease risk by altering trypsinogen expression levels. Indeed, messenger RNA (mRNA) analysis of pancreatic tissue of 69 individuals revealed a clear correlation between the allelic state and *PRSS1* transcripts. The highest *PRSS1* expression levels were detected in homozygous carriers (CC, risk allele), followed by intermediate levels in heterozygous (CT) carriers and the lowest levels in non-

carriers (TT, non-risk allele) (Whitcomb *et al*., 2012). The same GWAS identified a second locus, *CLDN2*, which maps to the X-chromosome. While the *PRSS1-PRSS2* locus showed a strong association with chronic and recurrent acute pancreatitis, the *CLDN2* locus predominantly associated with the chronic condition as well as alcohol consumption. Conclusively, the authors categorized variants at both loci as disease modifiers for sporadic (*PRSS1-PRSS2*) and alcoholic (*CLDN2*) CP (Whitcomb *et al*., 2012). Confirmatory and extending evidence was provided in a replication GWAS in European subjects. The key findings underlined the association of both loci (*rs10273639* at the *PRSS1–PRSS2* locus, *rs7057398* and *rs12688220* at the *CLDN2–MORC4* locus) to a lesser extent with CP and in a stronger extent with alcoholic CP (ACP). Sex-related influences of the X-chromosomal *CLDN2-MORC4* on an increased ACP risk in men were ruled out. It should be of note though, that *rs7057398* associated only in female but overall non-alcoholic CP patients, which leaves room for speculation on sex-related genetic risk factors (Derikx *et al*., 2014).

In addition, well-known risk genes were replicated in the same European GWAS and cohort of ACP subjects and confirmed *SPINK1* (*rs146437551*, OR=3.82), *CLDN2-MORC4* (*rs12688091*, OR=2.57), *PRSS1-PRSS2* (*rs2855983*, OR=1.84) and *CTRC* (*rs545634*, OR=1.83) while the inversion of *CTRB1-CTRB2* (*rs8055167*, OR=1.35) was added. Functional studies integrated the *CTRB1-CTRB2* inversion well into established model of the trypsin-dependent CP pathogenesis. CTRB2 was shown to degrade PRSS2 and to attenuate trypsin activity more effectively than CTRB1. In homozygous risk allele carriers, the expression ratio of *CTRB1/CTRB2* was significantly higher than in heterozygous risk allele carriers as measured from human pancreatic mRNA levels. Conclusively, protective PRSS2 degradation is impaired upon decreased *CTRB2* expression and thus facilitates increased trypsinogen activation in CP pathogenesis. Since *rs8048956* was in even higher LD with the *CTRB1/CTRB2* inversion, it proofed more suitable as a tag SNP in the analysis of three independent non-alcoholic CP cohorts. Both the lead (*rs8055167*) and the tag (*rs8048956*) SNP are in proximity to intron 1 (*CTRB2* major allele, *CTRB1* minor allele). The importance of the *CTRB1-CTRB2* inversion was replicated and thus established an association in non-alcoholic subjects as well (*rs8048956*, OR=1.62) (Rosendahl *et al*., 2018).

Even more recently, whole exome sequencing revealed the so far unknown c.268C>T transition in the chymotrypsin-like elastase family member 3B (*CELA3B*) as a predisposing variant for familial pancreatitis with diabetes and pancreatic adenocarcinoma. The discovery in the affected patient bridged the gap to a kindred of 71 individuals with HP from North Carolina with so-far unknown etiology reported in 1968. Functionally, the p.R90C gain-of-function mutation resulted in increased protein translation despite equal mRNA transcript levels. Also, increased protein levels were observed in total lysate and conditioned media of HEK293T cells. Interestingly, the ancestral p.R90L variant showed even higher translation rates and protein levels. The evolutionary transition of p.R90L could be viewed as a protective mechanism against increased CELA3B activity. The authors thus surmised an enhanced proteolytic activation of CELA3B by trypsin as the underlying pathomechanism. The p.R90C and p.R90L rendered the protein susceptibly to a much faster activation than their wild type control. While trypsin is not the primary cause in this scenario, the *CELA3B* p.R90C variant still falls into the category of increased proteolytic activity within the acinar cells and sheds light onto a similar and yet distinct mechanism of trypsin-dependent CP progression (Moore *et al*., 2019). Masson *et al*. confirmed the previous findings when they sequenced *CELA3B* in a French cohort of CP patients (n=644, control n=544) and found a c.269G>T transition in four patients but none of the controls. The variant affected the same amino acid (p.R90) and lead to the ancestral form which was described as p.R90L and characterized by Moore *et al*. (Masson *et al*., 2021).

1.3.2 Genetic variants in acinar genes: the misfolding-dependent model

A second pathogenic model arose upon the functional study of another HP associated trypsinogen variant identified in individuals of Thai, Turkish, French, and German origin. *PRSS1* p.R116C causes improper protein folding and cellular retention. Subsequent analysis of endoplasmic reticulum (ER) stress markers such as immunoglobulin-binding protein (BiP) and the spliced form of X-box binding protein-1 (XBP1s) confirmed elevated stress levels. Hence, a model complementing the trypsin-dependent pathway of CP was brought forth, placing protein misfolding in the center of second pathogenic pathway (Kereszturi *et al*., 2009). The functional analysis of *CPA1* variants in a CP early-onset cohort thereafter provided confirmatory and extending evidence for the new hypothesis. Overrepresentation of

loss-of-function variant (apparent activity < 20%) was found in a German discovery cohort, in a European replication cohort as well as in two cohorts of non-European decent with Indian and Japanese subjects. In the German subjects, the *CPA1* p.N256K was found in 7 out of 944 cases but in none of the 3,938 controls and showed strongest association. On a functional level, all apparent activity and secretion was lost for this variant in transfected HEK293T cells. Further, p.N256K transfected rat acinar AR47J cells responded with highly elevated levels of BiP and increased splicing of XBP1. Induction of ER stress was thus evident and confirmed protein misfolding as a pathological consequence of CP associated genetic variants (Witt *et al*., 2013).

The discovery of carboxyl ester lipase (*CEL*) as a novel CP susceptibility gene was made through a candidate gene approach. A hybrid allele (*CEL-HYB*) originating from a cross-over event of *CEL* and its adjacent pseudogene *CELP* has been associated with CP. As *CEL* is a lipase it expands the panel of previously addressed risk genes, which grouped into the category of proteases and inhibitors. Single base deletions in the *CEL* variable number of tandem repeat (VNTR) region in exon 11 were previously reported to cause maturity-onset diabetes of the young, type 8 (MODY8). Like CP, MODY8 is characterized by pancreatic exocrine dysfunction and morphological remodeling of the tissue. Screening of a discovery cohort of 71 German subjects with familial CP revealed a significantly enrichment of *CEL-HYB* allele in affected subjects compared to controls (OR=15.5; 95% CI=5.1–46.9; P=1.3×10^{-6}). Further screening in three independent cohorts of German and French origin confirmed the original findings (OR=5.2; 95% CI=3.2–8.5; P=1.2×10^{-11}). Control screening for known HP associated *PRSS1* variants revealed no co-occurrence with the *CEL-HYB* variant. Functional analysis in HEK293 and mouse acinar 266-6 cells showed secretion defects, impaired stimulation by bile salts and increased intracellular accumulation compared to the wild type. Those observations pointed towards an excess of defective proteins. Indeed, the authors measured increased levels of LC3-II, an autophagy marker, which went in line with the expected cellular response mechanism and the picture of autophagy as a previously suggested deleterious event in pancreatitis (Fjeld *et al*., 2015). Interestingly, an attempt to replicate the European findings in independent cohort of Chinese, Japanese and Indian subjects failed and the association of the *CEL-HYB* with CP

was not confirmed. Instead, an alternative hybrid allele (*CEL–HYB2*) was identified in Asian populations with a carrier frequency of 1.3%. Nonetheless, the association with CP remained absent. *CEL-HYB* was therefore postulated as an exclusively European risk variant (Zou *et al*., 2016).

After the establishment of CEL as a trypsin independent CP risk factor, pancreatic lipase (*PNLIP*) also shifted into the focus by a candidate gene approach. Even though the functional consequences of *PNLIP* variants remained unresolved, their association with early-onset CP was established without doubt. The most frequently found variant in the German discovery cohort was p.F300L (P=0.03). The variant was exclusively detected in patients and the enrichment across all European subjects was also highly significant (P=0.0001). Functional analysis proofed four more European variants (p.P245A, p.I265R, p.S304F, and p.F314L) besides p.F300L susceptible towards an increased degradation by trypsin, CTRC or CTRB2. Even though the subcategory of all protease-sensitive variants significantly associated with CP (OR=11.3, 95% CI=3.0–49.9, P<0.0001) in the European cohort, the pathogenic consequences remain elusive. Moreover, *PNLIP* missense variants were not enriched in Japanese CP subjects and the p.F300L was even completely absent. Furthermore, no missense variants were detected in additional cohorts from India and the USA. Like in the case of the *CEL-HYB* allele, the absence of pathogenic *PNLIP* variants in the Asian populations once more underpins the potential ethnic discrepancy in European and Asian genetic CP susceptibility (Lasher *et al*., 2019).

1.3.3 Genetic variants in ductal genes

After the initial association of *CFTR* variants with pancreatic disease, variants in another ductal gene recently emerged as significant susceptibility gene, namely the transient receptor potential cation channel subfamily V member 6 gene (*TRPV6*).

The first documentation of *CFTR* alterations in patients with HP dates to 1996. The authors analyzed two HP families and identified variant p.L327R which segregated with the disease in one of the families (Ravnik-Glavac *et al*., 1996). In two subsequent studies, the variant p.F508del was additionally found as the most frequently overrepresented variant in CP patients followed by the less frequent

p.R117H variant (Cohn *et al*., 1998, Sharer *et al*., 1998). Cohn *et al.* mechanistically suggested ductal obstruction due to impaired dilution and alkalinization of the pancreatic juice. In turn, Sharer *et al.* clearly distanced themselves from the latter explanation and argued in favor of direct acinar cell damage which was amplified by a reduced pH in the intra-acinar space and the ductal lumen. While the mechanistic consequences remain a matter of debate, the genetic association was consistent in both study cohorts. Testing for the so *5T* allele in intron 8 revealed an association with CP in both studies. Functionally, the variant impairs proper exon 9 splicing and reduces *CFTR* expression in subjects who are compound heterozygous for a second *CFTR* mutation (Cohn *et al*., 1998**,** Sharer *et al*., 1998). A later genetic study (ICP or HP patients n=660, controls n=1758) could not confirm the association of the *5T allele* and CP but revealed a comparable frequency of the variant in patients and controls. Moreover, the report highlighted that CFTR variants had an overall minor influence on CP development (Rosendahl *et al*., 2013).

The discovery of the two variants p.A210V and p.D324N in the *TRPV6* after whole exome sequencing (WES) of a 35-years old Japanese ICP patients gave incentive for the investigation of *TRPV6* variants in Japanese and European subjects. Subsequent analysis revealed an overrepresentation of p.A210V (*P*=0.48), p.I223T (OR=10.9; 95% CI=4.5-25.9; $P=7.4x10^{-9}$) and p.D324D (*P*=0.01) in the Japanese cohort. Variant p.A18S (OR=0.5; 95% CI= 0.3-0.7; $P=6.6\ x10^{-5}$) was significantly enriched in controls of the European cohort while p.L299Q (OR=3.0; 95% CI=1.9-4.8; *P*=0.003) and p.R483W *(P*=0.025) were significantly enriched in patients of the same study population. It should be of note, that p.L299Q was absent in the Japanese cohort and therefore exclusively occurred in the European cohort. Functional studies in HEK293 cells further revealed a significant decrease in TRPV6 activity for the p.A210V and p.R483W while such effect could not be demonstrated for the genetically most relevant variants mentioned above. Upon the further study of the voltage dependent Ca^{2+} flux of TRPV6, p.L299Q and p.D324N showed decreased activation kinetics while activation of the p.I223T and the wild type showed a similar response. As a resulting hypothesis the authors suggested constitutive suppression of the TRPV6 through a pore blockade mechanism which in the wild type achieved by intracellular Mg^{2+} (Masamune *et al*., 2020). Overall, 4.3% (13/300) of the Japanese patients and 2% (18/880) of the German patients

carried a functionally defective variant while these variants were found in only 0.1% (1/1070) of the Japanese controls and in none of the German controls (0/1320). In the view of the present work and CP as a multilayer genetic disease, the following aspect is of particularly interest: 20% (6/30) of the total patients not only carried a defective TRPV6 variant but also were trans-heterozygous for the *SPINK1* p.N34S variant. Compared to carriers of *CPA1* variants, only 10% of the cases were reported to be *SPINK1* p.N34S trans-heterozygous (Witt *et al*., 2013**,** Masamune *et al*., 2020). A second report of the same year confirmed the association of *TRPV6* loss-of-function variants in Chinese CP patients (OR=8.29, *P*=0.022). Loss-of-function was evidenced based on intracellularly reduced Ca^{2+} levels in HEK293T transfected with the variants compared to the wild type (Zou *et al*., 2020).

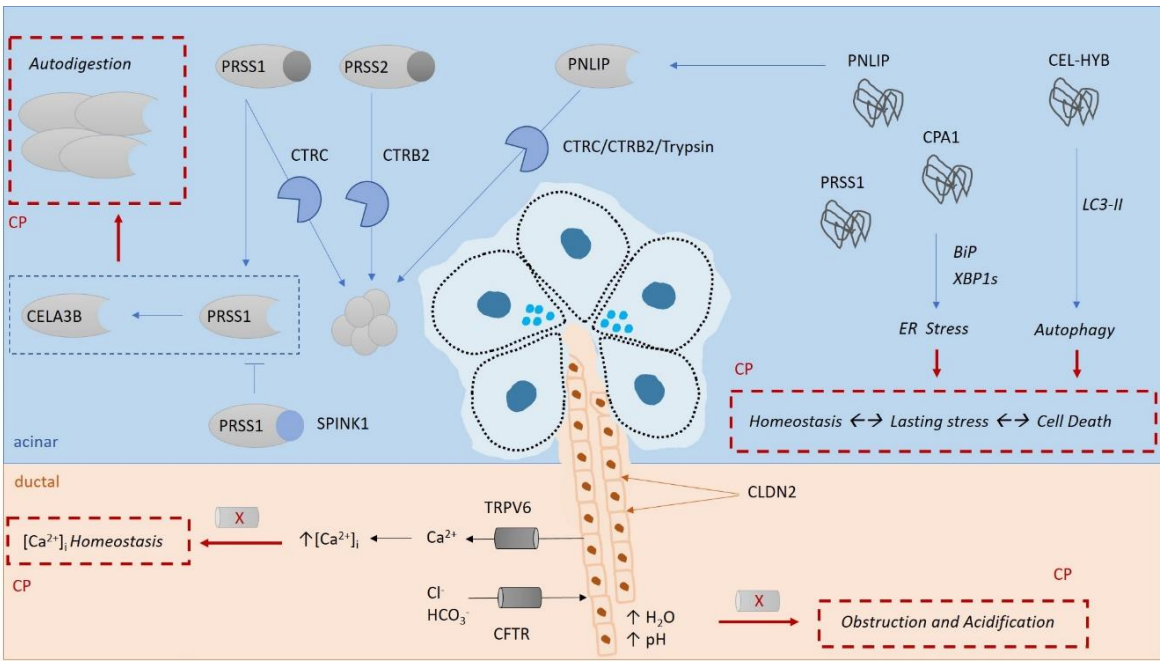


Figure 1 Overview of CP risk genes and related disease models.

Acinar risk genes associated with the trypsin-dependent pathway are presented in the upper left panel. PRSS1 (cationic trypsin) activation can lead to increased proteolytic activity, autodigestion and chronic pancreatitis (CP) if protective mechanism like degradation through CTRC or SPINK1 inhibitory activity fail. Active PRSS1 further activates the inactive precursors of digestive enzymes (zymogens), like CELA3B. Intra-pancreatic activation of zymogens further fuels autodigestion. A protective mechanism against premature PRSS2 (anionic trypsin) activation is the degradation by CTRB2. Loss-or gain of function mutation in the respective genes overall impact the fine balance of proteolytic and anti-proteolytic activity within the acinar cell. Acinar risk genes associated with the misfolding-dependent pathway are presented in the upper right panel. Misfolding of highly abundant acinar proteins like PRSS1, CPA1 and PNLIP induce ER stress, which is indicated by the increased levels of BiP and XBP1s. The CEL-HYB allele leads to autophagy, indicated by increased levels of LC3-II. If cell homeostasis cannot be restored, lasting stress and cell death can lead to the progression of CP. Ductal genes and related (dys)function are

presented in the lower panel. Through the secretion of chloride (Cl-) and bicarbonate (HCO3-) to the pancreatic duct, CFTR contributes to the fluidity and alkalization of the pancreatic juice. A defect in CFTR activity contributes to obstruction and acidification within the pancreatic duct and promotes CP. TRPV6 contributes to calcium homeostasis and increases intracellular calcium levels. Defects in TRPV6 and disturbances in the calcium homeostasis are associated with CP progression. CLDN2 locus is a tight junction protein but despite genetic association, its functional role in CP is not yet understood. CTRC: Chymotrypsin C, SPINK: serine protease inhibitor Kazal type 1, CELA3B: Chymotrypsin-like elastase family member 3B, CTRB2: Chymotrypsin B2, CPA1: Carboxypeptidase A1, PNLIP: Pancreatic Lipase, ER: Endoplasmic Reticulum , BiP: Immunoglobulin-binding protein, XBP1s: spliced form of X-box binding protein-1, CEL-HYP: Carboxyl Ester Lipase Hybrid, LC3-II: Microtubule-associated protein light chain 3, CFTR: Cystic Fibrosis Transmembrane Conductance Regulator, TRPV6: Transient receptor potential cation channel subfamily V member 6, CLDN2: Caludin-2

1.4 Genetic variants in CP: insights from animal models

Animal (rodent) models help to further assess the multi-layer pathogenesis in vivo.

Besides biochemical and cell culture studies, animal models form a third important pillar of research models. While animal models bear the advantage of remodeling the disease in the context of a whole organism, the reconstitution of a human phenotype including etiology, symptoms, histopathology, pathophysiology, and treatment effectiveness is usually incomplete and prone to challenges of its own. The development towards an ideal model is thus still underway (Klauss *et al*., 2018**,** Saloman *et al*., 2019, Semaniakou *et al*., 2019).

In the context of CP risk genes outlined in section 1.3, Archer *et al.* undertook the first approach to reconstitute the most prominent HP genetic variant *PRSS1* p.R122H and confirmed the postulated early acinar cell injury, pancreatic inflammation and tissue remodeling with increasing age of the animals (Archer *et al*., 2006). Furthermore, the transgene rendered animals more susceptible towards caerulein-induced pancreatitis. Those observations were mostly replicated by Athwal *et al*. It is of special interest, though, that in contrast to the murine transgene in the first model, human *PRSS1* wild type and p.R122H were introduced. Importantly, the expression of either the wild type or transgenic human *PRSS1* induced apoptosis in the acinar cells and lead to spontaneous pancreatitis (Athwal *et al*., 2014). A similar phenotype was observed in heterozygous *Spink1* c.194+2T>C mutant mice who also developed CP spontaneously (Sun *et al*., 2019).

Also, Moore *et al.* provided convincing evidence for their proclaimed pathologic *CELA3B* p.R90C variant, which they suggested to cause translational upregulation, increased activation by trypsin and uncontrolled proteolysis. Similar to the *PRSS1* p.R122H mouse model, the insertion of the *CELA3B* mouse homologue p.R89C by itself was not sufficient to trigger the pathogenic effect. Only upon the administration of caerulein, the homozygous mutant mice developed more severe pancreatitis than their matched controls by measures of immune infiltration, acinar dedifferentiation, and loss of lobular integrity (Moore *et al*., 2019).

Besides the reproduction of the trypsin-dependent model, investigations of transgenic *CPA1* p.N256K mice supported the misfolding-dependent model. Upon the introduction of the most frequent human mutation p.N256K into the mouse *Cpa1,* animals developed a clear CP phenotype including progressive acinar cell atrophy, inflammatory cell infiltration, fibrosis, and acinar-ductal metaplasia. Increased expression levels of BiP and C/EBP Homologous Protein (CHOP) further confirmed the induction of ER stress. Importantly, signs of pancreatic disease were absent in *CPA1 null* control animals (Hegyi & Sahin-Tóth, 2019).

With respect to ductal risk genes, mouse models harboring the *CFTR* p.F508del were found to only develop mild pancreatic disease (Semaniakou *et al*., 2019 and references therein). In contrast, investigations of Masamune *et al.* on a *TRPV6* p.D541A (corresponding to p.D581A in Uniprot entry Q91WD2) model underline the importance of this gene in proper pancreatic function. Similar to the observations made in the mouse models of *PRSS1* p.R122H and *CELA3B* p.R90C, the genetic variant only led to CP development upon the caerulein injections. Again, the combination of a genetic alteration combined with an additional insult by CCK-analogue overstimulation resulted in increased fibrosis, loss of acinar cells and the induction of acinar to ductal metaplasia (Masamune *et al*., 2020).

1.5 The *SPINK1* p.N34S haplotype as a risk factor for CP

SPINK1 p.N34S is an established risk factor for CP but the functional consequences remain enigmatic.

The *SPINK1* p.N34S (*rs17107315*; c.101A>G) variant is an established risk factor for CP. In the lead publication of Witt *et al.* in 2000, 23% of the children and young

adolescents with ICP of a German/Austrian cohort carried a mutation in *SPINK1*. Of all the variants, p.N34S showed the strongest association with the disease. The variant was found in 18/96 patients of which 6 individuals were homozygous. In contrast, p.N34S was only found in one heterozygous carrier out of 227 controls (Witt *et al*., 2000). Subsequent studies supported these initial findings (Pfützer *et al*., 2000, Chandak *et al*., 2002, Truninger *et al*, 2002, Bhatia *et al*., 2002). An overview is provided in table 1.

Table 1 Identification of SPINK1 p.N34S in patients and controls

Reference	Patients	Etiology	Controls	Origin
Witt *et al.* 2000	18/96 (6 hm)	Familial, ICP	1/227 (ht)	German, Austrian
AF	12.5%		0.2%	
Pfützer *et al.* 2000	28/112 (7 hm)	Familial, ICP	3/190 (ht)	American, European
AF	15.6%		0.8%	
Chandak *et al.* 2002	31/68 (8 hm)	TCP, FCPD	3/100 (ht)	Indian
AF	28.7%		1.5%	
Truninger *et al.* 2002	6/14 (1 hm)	ICP	4/397 (ht)	German, Swiss
AF	25%		0.4%	
Bhatia *et al.* 2002	29/66 (9 hm)	TCP, FCPD	2/92 (ht)	Indian
AF	28.8%		1.1%	

hm: homozygous, ht: heterozygous, ICP: Idiopathic Chronic Pancreatitis, TCP: Tropical Calcifying Pancreatitis, FCPD: Fibrocalculous pancreatic diabetes, AF: Allele Frequencies

SPINK1 is a polypeptide consisting of 79 amino acids. It contains a signal peptide of 23 amino acids, so that the secreted protein has 56 amino acids. The secreted protein is 6.5 kDa in size and makes up 0.1-0.8% of the total protein in the pancreatic juice. Its trypsin inhibitory capacity accounts for up to 20% (Pubols *et al*., 1974). The full amino acid sequence of the human protein was determined and published by Bartelt *et al*. in 1977 (Bartelt *et al*., 1977). Horii *et al.* identified the corresponding gene and reported multiple transcriptional start sites. *SPINK1* spans 7.5 kbp and is

located on the reverse strand of chromosome 5 (147,824,572-147,831,671(GRCh38.p13)). The coding sequence is divided into four exons (Horii *et al*., 1987).

The physiological relevance of SPINK1 to proper organ function and intra-pancreatic homeostasis was clearly demonstrated by the study of genetically engineered mice. Upon knock-out of the *SPINK1* mouse homologue (*Spink3*), homozygous animals showed severe autophagy and cell death in the pancreas shortly after birth. Intra-acinar trypsin activity was detected 0.5 to 1.5 days after birth in the *Spink3*$^{-/-}$ animals but not in heterozygous and homozygous wild type animals (Ohmuraya *et al*., 2006). The protective role of SPINK1 was further underpinned when Nathan *et al.* effectively increased the resistance against caerulein-induced pancreatitis in transgenic mice overexpressing the *SPINK1* rat orthologue *Psti-1*. Acinar expression of *Psti-1* significantly reduced histological severity of pancreatitis and trypsin activity in the transgenic animals while increasing the endogenous trypsin inhibitory capacity to 190% (Nathan *et al*., 2005). Moreover, the overexpression of *Psti-1* in *Spink3*$^{-/-}$ mice (*Spink3*$^{-/-}$ /TgN(*Psti-1*)) compensated for the loss of Spink3 protein and restored trypsin inhibitory activity to a comparable level observed in the wild type (Romac *et al*., 2010).

Astonishingly and despite the strong genetic association and physiological relevance of *SPINK1*, the biochemical and cellular consequences of the p.N34S variant remain unresolved. The exchange of asparagine to serin at amino acid position 34 suggested a protein defect at first. Computational analysis predicted the loss of a native turn structure between amino acid 34 (asparagine) and 35 (glutamine). Because of the structural proximity to the trypsin-binding site, a defective inhibitory capacity seemed to lie at hand (Kuwata *et al*., 2001). Subsequent functional analysis, however, failed to back up this assumption. When SPINK1 p.N34S was biochemically assessed for trypsin inhibitory activity under standard (pH 8.0, 0.02 M $CaCl_2$), pH-or calcium-adapted conditions (pH 5.0-9.0; 0-0.02 M $CaCl_2$, respectively) no differences compared to the wild type were detected. There was also no difference in susceptibility to trypsin degradation. (Kuwata *et al*., 2002). Further investigations of *SPINK1* missense variants (p.D50E, p.Y54H, p.R65Q, p.R67C) in addition to p.N34S also did not show any defect in trypsin

inhibitory activity. While decreased secretion was evident for the variants p.D50E, p.Y54H, p.R65Q and p.R67C and suggested intracellular retention and degradation, protein secretion of p.N34S remained unaffected (Király *et al*., 2007). SPINK1 protein defects of the p.N34S variant were therefore ruled out as a disease-driving factor.

Henceforth, the focus shifted onto p.N34S-linked non-coding variants. Linked variants always occur along with lead mutation and a set of linked variants is called a haplotype. Any variant within this haplotype could be the disease causing one. In a mini-gene approach, Kereszturi *et al*. investigated two common *SPINK1* haplotypes. The p.N34S haplotype consisting of the linked variants c.56-37T>C in intron 1, c.87+268A>G in intron 2, c.195-606G>A and c.195-66_65insTTTT in intron 3 and a second haplotype consisting of the promoter variant c.1-215G>A and c.194+2T>C. The variant c.194+2T>C was known to affect the 5' splice site in intron 3 which translated to a clear phenotype in transfected HEK293T cells with markedly reduced *SPINK1* mRNA levels, loss of SPINK1 protein and corresponding trypsin inhibitory activity. In turn, none of the p.N34S-associated variants showed any functional defect but behaved exactly as the wild type (Kereszturi *et al*., 2009).

As a matter of fact, after all linked intronic variants were found to be functionally innocuous, the focus for the search of the p.N34S associated pathogenetic variant turned towards the investigation of the upstream region of *SPINK1.* Assessment of *SPINK1* mRNA expression levels from the p.N34S heterozygous pancreatic cancer cell lines Paca44 and PancTu-I revealed an up to 6.7-fold decreased expression of the variant allele. This observation fostered the assumption that p.N34S linked variants in the 5' region of gene might exhibit a suppressive effect on the *SPINK1* expression (Kereszturi *et al*., 2017**)**.

First *in-silico* and *in-vitro* investigations proposed a functional relevance of *rs142703147C>A* (c.1-4141G>T). The variant was in full linkage with p.N34S in a population of 548 French ICP patients and 562 matched controls (r^2=1). Following a candidate-driven approach, visual inspection of the SNP spanning DNA sequence revealed a disruption of a putative HNF1A- and PTF1-L binding site (HNF1a: Hepatic Nuclear Factor 1 alpha, PTF1-L: adult Pancreas Transcription Factor 1

complex). In particular, the C>A transversion would alter the E-Box recognition site of the bipartite binding sequence (E-Box and TC-Box) to which the PTF1-L complex binds. The synergistic action of HNF1A and PTF1-L on *SPINK1* expression was effectively shown in *SPINK1* promoter transactivation assays in HEK293T cells. *SPINK1* luciferase reporter gene activity was increased by 6-fold upon the combined overexpression of HNF1a and PTF1-L compared to the empty vector transfected cells. To test the effect of *rs142703147*, the putative HNF1A-PTF1L core regulatory module was inserted downstream of the luciferase gene either harboring the non-risk (C) or risk allele (A). The combined overexpression of the transcription factors led to an 8-fold increases in luciferase reporter gene activity upon the presence of the C allele compared to the empty vector controls. With the A allele present, the reporter gene activity was only induced by 5-fold and did not differ from the reporter gene construct without the HNF1a-PTF1a core regulatory module. Electron mobility shift assay (EMSA) experiments with nuclear extracts from HEK293T with the transiently expressed PTF1-L complex further demonstrated the disruption of PTF1-L binding to the *rs142703147 (A)* allele (Boulling *et al*., 2017).

1.6 Aim of this work

The aim of this work set out to identify and functionally characterize the pathological variant in the regulatory region of *SPINK1* that underlies the CP-associated p.N34S haplotype. Preliminary research in our group narrowed p.N34S linked regulatory variants down to three final candidates. While functional analysis of one candidate, *rs17107316* (c.87+268T>C) was already published with negative results on *SPINK1* expression, the other two candidates, *rs148276928* (c.1-2090T>C) and *rs148911734* (c.1-7321C>T), were not investigated so far. We build on the hypothesis that instead of an altered protein function impaired gene regulation leading to altered *SPINK1* expression levels might be the underlying pathological mechanism. Altered levels of SPINK1 expression are expected to influence the fine balance of premature proteolytic activity within the pancreas and might therefore take up a role in CP initiation and progression.

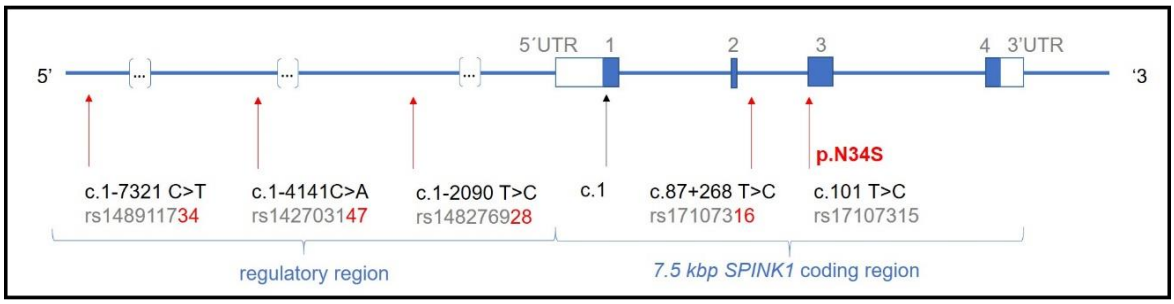


Figure 2 SPINK1 and p.N34S related variants under investigation in this thesis.

The indicated nucleotide represents the actual nucleotide of the gene, which is encoded on the reverse strand. The last two digits highlighted in red of each rs number are used as abbreviation for the corresponding SNP in this thesis. Empty box: untranslated region (UTR), filled box: exon 1-4. c.1 refers to the first nucleotide of the coding sequence.

2 Material

2.1 Human and mouse samples

The human complementary DNA (cDNA) of pancreatic tissue was sampled from 9 individuals and was a kind gift of Prof. Dr. Dr. Ihsan Ekin Demir (Chirurgische Klinik, Klinikum rechts der Isar der TUM).

Whole human blood of 638 Swiss individuals as wells as RNA from human intestinal biopsies of five *SPINK1* p.N34S heterozygous carriers and five wild type controls were kindly provided by PD Dr. med. Kaspar Truninger.

Whole tissue lysate for Western blot analysis of human and mouse pancreas was kindly provided by Prof. Dr. med. Hana Algül (Klinik und Poliklinik für Innere Medizin II, Klinikum rechts der Isar der TUM).

2.2 Cell lines

Table 2 List and specification of human cell lines and origin

Cell line	Origin
Caco2 Colorectal Adenocarcinoma (White, male)	Kind gift of Prof. Dr. Dirk Haller, TUM
HT29 Colorectal Adenocarcinoma (Caucasian, female)	Kind gift of Prof. Dr Hannelore Daniel, TUM
COLO357 Metastasis of Pancreatic Adenocarcinoma	Kind gift of Prof. Dr. Dr. Ihsan Ekin Demir, TUM
HepG2 Hepatocellular Carcinoma (Caucasian, male)	ATCC (HB-8065), Manassas, USA
Huh7 Hepatocellular Carcinoma (Japanese, male)	Sekisui XenoTech, LLC, Kansas City, USA
Capan1 Pancreatic Ductal Adenocarcinoma (Caucasian, male)	Kind gift of Prof. Dr. Güralp Ceyhan, TUM
Paca44 Pancreatic Ductal Adenocarcinoma	Kind gift of Prof. Dr. Matthias Löhr, Karolinska Institute, Huddinge
Panc1 Pancreatic Ductal Adenocarcinoma (White, male)	Kind gift of Prof. Dr. Jonas Rosendahl, Halle (Saale)

HEK293 established from a human primary embryonal kidney transformed by adenovirus type 5 (Ad 5)	DSMZ (ACC05), Braunschweig, Germany
HEK293T highly transfectable derivative of the human primary embryonal kidney cell line 293 (ACC 305) carrying a plasmid containing the temperature sensitive mutant of SV-40 large T-Antigen (tsA1609)	DSMZ (ACC-635), Braunschweig, Germany
RT4 Urinary bladder transitional cell carcinoma (male)	DSMZ (ACC-412), Braunschweig, Germany

2.3 Transfection

2.3.1 Human expression vectors

All human expression vectors were purchased from Genscript, Piscataway, USA. The pcDNA3.1_FLAG vector carries no insert and was used for control transfections.

Table 3 Vectors for transient overexpression experiments

Vector Name	Genscript ID	NCBI reference
pcDNA3.1_FLAG	-	-
pcDNA3.1_GATA4_FLAG	OHu41944D	NM_001308093.3
pcDNA3.1_GATA6_FLAG	OHu27933D	NM_005257.6
pcDNA3.1_HEYL_FLAG	OHu107321	NM_014571.4
pcDNA3.1_HNF1a_FLAG	OHu25248	NM_000545.6
pcDNA.3.1_NR5A2-wt-iso1_FLAG	OHu20723	NM_205860.3
pcDNA3.1_ONECUT1_FLAG	OHu00467	NM_001306179.2
pcDNA3.1_PTF1a_FLAG	OHu19480	NM_178161.3
pcDNA3.1_RBPJL-201_FLAG	OHu00805D	NM_001281448.2
pcDNA3.1_RBPJL-203_FLAG	OHu01387D	NM_001281449.1
pcDNA3.1_TCF3_FLAG	OHu23034	NM_003200.5

2.3.2 Reporter gene assay expression vectors

The vectors were provided by the laboratory of Prof. Witt as generated by others. The pGL4.22 SPINK1 promoter constructs were cloned as part of the Bachelor thesis of Andreas Schmidt. Table 4 lists the constructs. The corresponding vector maps are provided in the appendix.

Table 4 Vector constructs for reporter gene assays

	Vector	Reporter Gene	Promoter
1	pRL	Renilla luciferase	Human Ubiquitin C (UbC)
2	pGL4.22	Firefly luciferase	Human *SPINK1*
3	pGL4.22	Firefly luciferase	Human *SPINK1,* flanking sequence of *rs148911734* (C), non-risk allele, at 5' end
4	pGL4.22	Firefly luciferase	Human *SPINK1,* flanking sequence of *rs148911734* (T), risk allele, at 5' end

Note: Specification of vector name (map): (1) pRL-Ubi, (2) pGL4.22_SPINK1-Promotor_Luiferase, (3) pGL4.22_SPINK1-Promotor _flank-rs148911734-wt(C)-500bp_ Luciferase, (4) pGL4.22_SPINK1-Promotor _flank-rs148911734-mt(T)-500bp_Luciferase, Renilla luciferase: Coding sequence luciferase from the anthozoan coelenterate Renilla reniformis (sea pansy), Firefly luciferase: Coding sequence luciferase reporter gene *luc2* (Photinus pyralis) and is designed for high expression and reduced anomalous transcription

2.3.3 ON-TARGETplus siRNAs

All siRNAs were ordered as SMARTpool from Dharmacon Horizon (Cambridge, UK). siRNAs were resuspended in 1x siRNA buffer (diluted with phosphate buffered saline (PBS) from 5x siRNA buffer stock, B-0002000-UB-100, Dharmacon Horizon, Cambridge, UK) during a 30 min shaking incubation at room temperature (RT). The working concentration was set to 5 µM. The resuspended siRNA was aliquoted and frozen at −20°C.

Table 5 SMARTpool siRNAs for transient knock-down

Target	Dharmacon Horizon ID
Human GATA4 (2626)	L-008244-00-0005
Human GATA6 (2627)	L-008351-00-0005
Human YY1 (7528)	L-011796-00-0005
Non-targeting Pool	D-001810-10-05

2.3.4 Transfection reagents

Lipofectamine™2000 (#11668019, Invitrogen™, Waltham, USA) was used for all vector transfections and siRNA transfections of Paca44 cells. For siRNA transfections of Caco2 and HepG2 cells, DharmaFECT 1 (T-2001-02, Dharmacon

Horizon, Cambridge, UK) and DharmaFECT 4 (T-2004-01, Dharmacon Horizon, Cambridge, UK), respectively, proofed most suitable.

2.4 Primers

All primers and probes were ordered from TIB MOLBIOL (Berlin, Germany). Primers for polymerase chain reaction (PCR) and sequencing were ordered in GSF purification grade. Probes and primers for mutagenesis were ordered in HPLC grade.

Table 6 Primer for SPINK1 p.N34S RNA-Seq analysis

Name	Nucleotide Sequence 5'-3'
SPINK1-UTR-RevA	TCAAAACCTTGGTTCTCAGC
SPINK1-Q12F-Adaptor	TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGCCCTGTTGAGTCTATCTGGTAACAC
SPINK1-Q34R-Adaptor	GTCTCGTGGGCTCGGAGATGTGTATAAGA GACAGGAGGATAGAAGTCTGGCGTTTCCG
SPINK1-N34S-Seq-Exon3R	ACATAACACGCATTCATTGG
SPINK1-N34S-Seq-Exon3Ra	CACGCATTCATTGGGATAAG
Note: Adaptor Sequences for RNA-Seq are underlined	

Table 7 Primer for Sanger sequencing of the human expression vectors

Name	Nucleotide Sequence 5'-3'
pcDNA3.1-1F	GTCGCTGAGTAGTGCGCGAGC
pcDNA3.1-2F	CGCAAATGGGCGGTAGGCGTG
pcDNA3.1-3F	TGATCAGCCTCGACTGTGCC
pcDNA3.1-4R	CTACTCAGACAATGCGATGC
pcDNA3.1-5F	GCCCTTTGACGTTGGAGTCC
pcDNA3.1-6F	CTCTGCCTCTGAGCTATTCC
pcDNA3.1-7F	TCGATTCCACCGCCGCCTTC
pcDNA3.1-8R	TGGGGTGGGCGAAGAACTCC
pcDNA3.1-9F	CTGCGGCGAGCGGTATCAGC
pcDNA3.1-10F	GCTCTGCTGAAGCCAGTTACC
pcDNA3.1-11F	ATGGCTTCATTCAGCTCCGG
pcDNA3.1-12F	AGGGAATAAGGGCGACACGG
Note: The primers listed above sequentially allow to cover the whole vector sequence	

T7-F	TAATACGACTCACTATAGGG
M13rev	CAGGAAACAGCTATGAC
Note: The primers listed above cover the coding sequence of the insert	

Table 8 Primer for GATA6 p.P.K484R mutagenesis

Name	Nucleotide Sequence 5'-3'
pcDNA3.1-*HindIII*-F	AAACTTAAGCTTGGTACCGAGCTCGGATCC
GATA6-p.K484R-*EcoRI*-R	CTGGTTTGAATTCCCTCTTTT**C**TCATAGCAAGT GGTCTGGGCACCCC
GATA6-p.K484R-F-Seq	GTCACACCACAACTACCACC
Note: Restriction enzyme cutting sites are underlined. **C** marks the nucleotide change in the triplet code from lysine (K) to arginine (R)	

2.5 Synthetic oligonucleotides

The forward and reverse strand of a 22bp flanking sequence upstream and downstream of each SNP of interest (total length of 45bp) were order for the risk and non-risk allele from TIB MOLBIOL (Berlin, Germany) at HPLC purification grade. Table 9 lists the corresponding sequences. For electrophoretic mobility shift assays (EMSA), each forward strand (Forw) was labeled with a Cy5-fluorophore at the 5'-end. For affinity chromatography, the Cy5-label was replaced with a biotin residue. The complementary strand (Rev) was ordered without modification.

Table 9 Synthetic oligonucleotides for EMSA

Name	Nucleotide Sequence 5'-3'
rs17107316	
-ForwT	Cy5-GCTTCCACCCCAAATACTTGTC**T**ACAGCTATCACTAGTCTATATA
-RevT	TATATAGACTAGTGATAGCTGTAGACAAGTATTTGGGGTGGAAGC
-ForwC	Cy5-GCTTCCACCCCAAATACTTGTC**C**ACAGCTATCACTAGTCTATATA
-RevC	TATATAGACTAGTGATAGCTGTGGACAAGTATTTGGGGTGGAAGC
*rs14827692**8***	
-ForwT	Cy5-AATGCCATCCTTTCCCCCTGGG**T**TTCTGCATTTCCTGCAAAGAGT
-RevT	ACTCTTTGCAGGAAATGCAGAAACCCAGGGGGAAAGGATGGCATT
-ForwC	Cy5-AATGCCATCCTTTCCCCCTGGG**C**TTCTGCATTTCCTGCAAAGAGT
-RevC	ACTCTTTGCAGGAAATGCAGAAGCCCAGGGGGAAAGGATGGCATT

rs142703147	
-ForwC	Cy5-TGGAATGGGGATATTGGGCTGG**C**ACCTGTTGATTTTCACATTCTG
-RevC	CAGAATGTGAAAATCAACAGGTGCCAGCCCAATATCCCCATTCCA
-ForwA	Cy5-TGGAATGGGGATATTGGGCTGG**A**ACCTGTTGATTTTCACATTCTG
-RevA	CAGAATGTGAAAATCAACAGGTTCCAGCCCAATATCCCCATTCCA
rs148911734	
-ForwT	Cy5-TAATTCATCTTTCATCTACATT**T**ATCTTTCTATTCATCTTCCATC
-RevT	GATGGAAGATGAATAGAAAGATAAATGTAGATGAAAGATGAATTA
-ForwC	Cy5-TAATTCATCTTTCATCTACATT**C**ATCTTTCTATTCATCTTCCATC
-RevC	GATGGAAGATGAATAGAAAGATGAATGTAGATGAAAGATGAATTA
Note: nucleotide highlighted in **bold** marks mid position (SNP)	

2.6 PCR and probe-based assays

2.6.1 Assay for *SPINK1* p.N34S genotyping

Table 10 Assay for SPINK1 p.N34S genotyping

Name	Nucleotide Sequence 5'-3'	Specification
SPINK F3	CCAATCACAGTTATTCCCCAGAG	Primer
PSTI-I31R	CACCACTTATGGTCACTGAGG	Primer
PSTI-3FL	CCAAATGTTACA**A**TGAACTTAATGGATGC	Probe
PSTI-3LC	CCAAGATATATGACCCTGTCTGTGGGAC	Probe
Note: **A** marks c.101 A>G (p.N34S) of the wild type-specific probe		

2.6.2 Primers for qPCR with SYBR Green

Table 11 Primer for qPCR with SYBR Green

Name	Nucleotide Sequence 5'-3'
SPINK1-Q12F	CTGTTGAGTCTATCTGGTAACAC
SPINK1-Q34R	GGATAGAAGTCTGGCGTTTCCG
hActin-1F	GCGCCCCAGGCACCAGGGCG
hActin-1R	AGGTCTCAAACATGATCTGG
GAPDH-F	GATCATCAGCAATGCCTCCTGC
GAPDH-R	ACAGTCTTCTGGGTGGCAGTGA

2.6.3 Primers and probes for TaqMan assays

The TaqMan Assay were designed in close cooperation with Dr. Gerard Leboulle at TIB MOLBIOL (Berlin, Germany) and ordered from the same company. Assay establishing was performed out of the combination of the suggested forward (F, Fa, S, fwd, up) and reverse (R, Ra, A, rev, lo, B) primers on human pancreatic cDNA and cDNA of human cell lines. The human cell lines were selected based on the suggested target gene expression on the human protein atlas website (https://www.proteinatlas.org). The primer combination was selected based on their sensitivity (Cycle threshold (Ct) value), slope of the curve and reproducibility across human pancreatic tissue and cell line cDNA. The final primer combinations are listed in apathetical order in the tables below. Please note, P and TM specify probes, F- and --Q in the nucleotide sequence mark the donor (6-FAM) and acceptor (BBQ) of the hydrolysis probe, respectively.

Table 12 TaqMan assays for proteases and inhibitors

Name	Nucleotide Sequence 5'-3'
CELA3AB Q34F	GCCGGCCACTGCATCTCGAG
CELA3AB A	GATGTCATTGCCACAGGC
CELA3AB TM	F-CACACGAGCGGTTCCAGAGTGG--Q
CTRL S	AGTGGCGTGGGCAATG
CTRL A	TTCCCTTCTGGCAGACAAGA
CTRL P	F-CCTGCACAGATCATGGAGTCAGTGATACTT--Q
hCPA1 Fa	CTGGTTTGCAAAGAAGATCAC
hCPA1A	TTGGTGACGATCTCCAGGAA
hCPA1 TM	F-CAGCTTTCACCGCCATTCTCGAC--Q
hCTRC F	GCCCACTGCATCAGCAACAC
hCTRC-R	AGCTCCACATGCTCTGCAAGC
hCTRC TM	F-CCGTGGGAAAGAACAACCTGGAG--Q
hPRSS1 F	ACTGCTACAAGTCCCGCA
hPRSS1 A	GAGTCTTCCTGTCGTATTGG
hPRSS1 TM	F-CTGGGAGAGCACAACATCGAAGTCC--Q
SPINK1 Q12F	CTGTTGAGTCTATCTGGTAACAC
SPINK B(P1)	GTTCATTGTAACATTTGGCCTC
SPINK P1	F-CTTCCCAGGGAGTCAGCTCCAG--Q

Table 13 TaqMan assays for lipases and related proteins

Name	Nucleotide Sequence 5'-3'
CEL F	GCCAATCTGCCAGGTAACT
CEL A	GATATTCCTCTTCACCCAAGCAA
CEL P	F-CTTCGGGATCAGCACATGGC--Q
CLPS F	GAGTGCTCTGTCAAGACGC
CLPS R	GGGTGGGCAGTCTCACTG
CLPS P	F-CCACGCTCACAGGGACACTTGTAG--Q
hPNLIP Fa	TTGGACGCATCACAGGGTTG
hPNLIP Ra	CAAAGTCTCGAGTCCCTTCC
PNLIP TM	F-CAGAACCTTGCTTTCAGGGCACAC--Q
PLA2G1B S	GCTAGCTGTGCTGCTCACA
PLA2G1B R	CACGAGTATGAATAGGTGTGGG
PLA2G1B P	F-TCATAGCAGTTGTCATGTGTCTGGCAG--Q

Table 14 TaqMan assays for secreted proteins

Name	Nucleotide Sequence 5'-3'
GP2 F	CCATCCACTGTGGAGGACAA
GP2 A	GCAAACTGTGGACATCAGAACTA
GP2 P	F-CGAGGAGGAGTGCCTTGCCCT--Q
hCUZD1 Up	GCAGAAAACATCAACACTACATCTTTA
hCUZD1 Lo	GGTGATTATATTGGTGTAAGTAATTGACTG
hCUZD1 TM	F-CCACATCCATTAAGAGGGACAGAAAATTCC—Q
REG1A F	CTGACCTCAAGCACAGGATTC
REG1A R	GGGTAGGTAAAACTATTGAAGATTAACAA
REG1A P	F-CCTTGTGAAGACAAGTTCTCCTTTGTCTGC--Q
REG1B F	TCCATGACCCAAAAAAGAACC
REG1B A	GTTCTAGACATCCATTTTTCAGCTT
REG1B P	F-CTGGAGTAGTGGGTCCCTGGTCTCC--Q
SYNC S	GACCAGATCGCACTTGTGAG
SYNC A	CTCTGGCGTTCTTCCATTTC
SYNC P	F-ACGAGATGAAGAGGTGCAGCAGTACAG--Q
SYTL1 S	GGGCGGGTCAGCAAG
SYTL1 R	CCTCAGCCTCCCTGTCG
SYTL1 P	F-CCTGGTGCTCTTCTTCCTGCG--Q

Table 15 TaqMan assays for ductal proteins

Name	Nucleotide Sequence 5'-3'
hCFTR Fa	ACTGGAGCAGGCAAGACTTC
hCFTR Ra	GCAAACTTGGAGATGTCCTC
hCFTR TM	F-CTAGTTGGCATGCTTTGATGACGCTTC--Q
TRPV6 S	GGAAGCCTACATGACCCCTAAG
TRPV6 A	TTCTGAAGATGTCTGGAACCTCTAC
TRPV6 P	F-ACTGTCATTGGGGCTATCATCATCCTGC--Q

Table 16 TaqMan assays for transcription factors

Name	Nucleotide Sequence 5'-3'
BHLHA15 S	GATCCCCAGCTCCAAGG
BHLHA15 A	CCTGGACTGGCTGTCCTAG
BHLHA15 P	F-TCACCTTCCTGCCGCCACC--Q
HEYL F	GCTCTTCCAAGCTGGAGAAAG
HEYL A	CCAGGTACCTGATGACCTCAG
HEYL P	F-CTTGAAAATGCTCCATGCCACTGG--Q
HNF1A F	CTCCCCCAGTAAGGTCCACGGTG
HNF1A R	CCCCCAGCTGCTGAGACCAG
HNF1A P	F-CTATGGACAGCCTGCGACCAGTG--Q
HNF1B F	TCCAAACAAGCTGTCAGGAG
HNF1B A	GTGACTGATTGTTGAGGAGGAAG
HNF1B P	F-CTACAGCCAGCAGGGAAACAATGAGATC--Q
GATA4 S	CGTGTCCCAGACGTTCTCA
GATA4 A	GGGAGACGCATAGCCTTGT
GATA4 TM	F-CCATGGGCCCTCCATCCACC--Q
GATA6 F	CTCCAACTTCCACCTCTTCTAACT
GATA6 A	GCTCGCTGTTCTCGGGAT
GATA6 P	F-CTCTCTCCCGCACCAGTCATCAC--Q
NR5A2 S	AATGTCTAAGTGTTGGAATGAAGCTA
NR5A2 A	GCCCAAACTTATTCCTTCCTC
NR5A2 P	F-ACGCATTCGGTCGGCCCTTACA--Q
PTF1A F	GCCCAGAAGGTCATCATCTG
PTF1A A	CCATGAGAGAGAGTGTCCTGCTA
PTF1A P	F-CCCAGCGACCCTGATTATGGC--Q

RBPJ S	ACATCCATTATGGACAAACAGTC
RBPJ A	CCCATTCCCTCATAAAATGTATACTCT
RBPJ TM	F-CTCAGTTACTGGCATGGCACTCCC--Q
RBPJL S	CTACCTCCCATGATCATCCG
RBPJL A	GCCTGGAAACTGGAATGC
RBPJL P	F-CAGTGTGCGCTCCTTGATGTGGA--Q
TCF3 S	GCCTCATGCACAACCACG
TCF3 R	TCGCCCTAGCCCACTG
TCF3 P	F-CTGACCTGTCTCGGCCTCCC--Q
TCF7 F	CCCCAACTCTCTCTCTACGA
TCF7 R	GCAGAGGCCTGTGAACTT
TCF7 P	F-CCCTGCACCTGCGGACATCA--Q
TCF12 S	AGAGCTTGGCCGAATGTG
TCF12 A	GCTTTGGGGTTAAGGTTCCTC
TCF12 P	F-CAAGCCGTGGCAGTCATCCTTAGTCTAG--Q
YY1 S	GAGGGCGAGTTCTCGGT
YY1 A	AGGAGGTGAGTTCTCTCCAATG
YY1 P	F-CATCTGAGGACCACATGGTGACC--Q

Table 17 TaqMan assays for BiP

Name	Nucleotide Sequence 5'-3'
HSPA F	AAATTTGAAGAGCTCAACATGGATCT
HSPA A	GGTTTATGCCACGGGATGGT
HSPA5 Probe	F-CCGGTCTACTATGAAGCCCGTCCAGAAAGTG--Q

Table 18 TaqMan assays for house keepers

Name	Nucleotide Sequence 5'-3'
ACTIN F	AGCCTCGCCTTTGCCGA
ACTIN R	CTGGTGCCTGGGGCG
ACTIN TM	F-CCGCCGCCCGTCCACACCCGCC--Q
GAPDH fwd	GAAGGTGAAGGTCGGAGTC
GAPDH rev	GAAGATGGTGATGGGATTTC
GAPDH TM	F-CAAGCTTCCCGTTCTCAGCC--Q

2.7 Antibodies

Table 19 Primary antibodies for Super-Shift EMSA

Target	Dilution/Amount	Species, Isotype	Company ID	Company
GATA6	1:10	Rabbit mAb IgG	D61E4	Cell Signaling, Danvers, USA
GATA4	1:10	Rabbit mAb IgG	D3A3M	Cell Signaling, Danvers, USA
SATB2	1:10	Rabbit mAb IgG	E8R8H	Cell Signaling, Danvers, USA
Isotype control	1:10	Rabbit mAb IgG	DA1E	Cell Signaling, Danvers, USA
YY1	2 µg	Mouse mAb IgG	sc-7341	Santa Cruz, Dallas, USA
Isotype control	2 µg	Mouse mAb IgG	sc-2025	Santa Cruz, Dallas, USA
ZNF568	1:10	Rabbit mAb IgG	PA61798	Invitrogen™, Waltham, USA
Isotype control	1:10	Rabbit mAb IgG	02-6102	Invitrogen™, Waltham, USA
Note: Total reaction volume is 11.5µL				

Table 20 Primary antibodies for Western blot

Target	Dilution	Species, Isotype	Company ID	Company
GATA6	1:1,000	Rabbit mAb IgG	D61E4	Cell Signaling, Danvers, USA
GATA4	1:1,000	Rabbit mAb IgG	D3A3M	Cell Signaling, Danvers, USA
ANTI FLAG®	1:1,000	Mouse mAb	F1804	Sigma-Aldrich, Darmstadt, Germany
β-Actin	1:2,500	Goat polyclonal	ab8229	Abcam, Cambridge, UK
Lamin B1	1:1,000	Rabbit monoclonal	D9V6 h	Cell Signaling, Danvers, USA

Sumo 1	1:1,000	Rabbit mAb	ab32058	Abcam, Cambridge, UK
Sumo 2 + 3	1:1,000	Mouse mAb	ab81371	Abcam, Cambridge, UK

Table 21 Secondary antibodies for Western blot

Target	Dilution	Species, Isotype	Company ID	Company
Anti-rabbit	1: 10,000	IRDye® 680RD donkey anti-rabbit IgG (H+L)	926-68073	Licor, Lincoln, USA
Anti-rabbit	1:20,000	IRDye®800CW Donkey anti-Rabbit IgG	926-32211	Licor, Lincoln, USA
Anti-mouse	1: 10,000	IRDye® 680RD donkey anti-mouse IgG (H+L)	926-68072	Licor, Lincoln, USA
Anti-goat	1: 10,000	IRDye® 800CW donkey anti-goat IgG (H+L)	926-32214	Licor, Lincoln, USA

2.8 Enzymes

Table 22 List of commercial enzymes

Enzyme	Company ID	Company
Platinum™ SuperFi II DNA Polymerase	12361010	Thermo Fisher Scientific, Waltham, USA
MyTaq™ DNA Polymerase	BIO-21107-BL	Biocat, Heidelberg, Germany
OneTaq® Hot Start DNA Polymerase	M0481X	NEB, Ipswich, USA
Phusion™ High-Fidelity DNA Polymerase	F530L	Thermo Fisher Scientific, Waltham, USA
Exonuclease I	M0293L	NEB, Ipswich, USA
Antarctic Phosphatase	M0289L	NEB, Ipswich, USA

Fast Digest *HindIII*	FD0505	Thermo Fisher Scientific, Waltham, USA
Fast Digest *EcoRI*	FD0274	Thermo Fisher Scientific, Waltham, USA
T4 DNA-Ligase	EL0011	Thermo Fisher Scientific, Waltham, USA

2.9 Ready-to-use kits

Table 23 Ready-to-use kits

Kits	Company ID	Company
Protein Assay Dye Reagent Concentrate	500-0006	BioRad Laboratories, Hercules, USA
Human SPINK1 DuoSet ELISA Kit	DY7496-05	RD Systems, Minneapolis, USA
QIAamp DNA Blood Mini Kit	51106	Qiagen, Hilden, Germany
QIAshredder	79656	Qiagen, Hilden, Germany
RNeasy® RNA Extraction Kit	74106	Qiagen, Hilden, Germany
QuantiTect® Reverse Transcription	205314	Qiagen, Hilden, Germany
Nuclear Extract Kit	40010	Active Motif, Carlsbad, USA
qPCR BIO SyGreen Mix Lo-ROX	PB20.11-20	Nippon Genetics, EU, Düren, Germany
SensiFast Probe No-Rox Kit	Bio-86020-BL	Biocat, Heidelberg, Germany
Wizard SV Gel and PCR kit	A9282	Promega, Madison, USA

2.10 Others

Table 24 E.coli for vector amplification

Material	Company ID	Company
Stellar™ Competent Cells	636763	Clontech Laboratories/ Takara Bio Company, EU, Göteborg, Sweden

3 Experimental Work

3.1 Human expression vectors

Human overexpression vectors were ordered from Genscript, Piscataway, USA. The appropriate coding sequence (isoform) of the corresponding gene was chosen based on the listed transcripts on Ensemble Genome Browser (https://www.ensembl.org/index.html), the suggested isoform expression levels in the human pancreas (Genotype-Tissue Expression (GTex) https://www.gtexportal.org/home/) or after direct identification of the human pancreatic isoform (*i.e.,* RBPJL) from human pancreatic cDNA (see appendix). The vector constructs are listed in Table 3. All coding sequence inserts were embedded into the pcDNA3.1+/C-(K)DYK (listed as pcDNA3.1_FLAG) backbone vector. A representative vector map is provided in the appendix.

3.1.1 Control sequencing

Upon arrival of the vector constructs, the correct coding sequence was controlled by Sanger sequencing using the T7-F and pcDNA3.1-4R primer (see Table 7). If applicable, the primer listed in Table 7 allowed full sequencing of the whole vector construct in a sequential manner. Sanger sequencing was carried out at Eurofins Genomics (Ebersberg, Germany). In a total volume of 17 µL, 2 µL of the vector (500ng/µL) were diluted in 13 µL water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany) and submitted to sequencing with 2 µL primer (10 µM).

3.1.2 Transformation of human expression vectors into *E. coli*

For the amplification of the vector constructs, 0.5-1.0 µL of vector were added to 25 µL of freshly thawed Stellar™ Competent Cells (Clontech Laboratories/Takara Bio Company, EU, Göteborg, Sweden) and incubated on ice for 5 min. Following a heat shock at 42°C for 1 min, a 1 min cool-down on ice and the addition of 250 µL room-temperature super optimal broth with catabolite suppression (SOC) outgrowth medium (#B9020S, NEB, Ipswich, USA), the mixture was incubated shaking at 37°C for 20 min. Of the re-transformation mixture, 5-10 µL were transferred to a 9 cm lysogeny broth (LB) Agar plate (Lennox, Roth, Germany) with 0.1% ampicillin (Roth, Germany) and incubated overnight at 37°C. The remaining re-transformation mixture was added to 30 mL LB-medium (Roth, Germany) with freshly added

ampicillin (final concentration 0.1%). On the next day, the agar plates were checked for single colonies, sealed, and stored at 4°C. Single colonies from those plates were re-grown in 30 mL of 0.1% ampicillin LB-medium upon requirement.

3.1.3 Vector-purification from an overnight culture of transformed *E. coli*

In a large-scale approach (Maxi-Prep), the overnight culture (30 mL) was transferred to large collection tubes and centrifuges at 21,000 xg at 4°C for 1 min (Sorwvall LYNX 400 centrifuge, Thermo Fisher Scientific, Waltham, USA). The supernatant was removed thoroughly, and the bacterial pellet was resuspended in 2 mL of solution 1 (100 mM Tris, 10 mM EDTA) including freshly added RNAse A (100 mg/mL, Qiagen, Hilden, Germany) at a dilution of 1/1,000. After completely dissolving the pellet, 2 mL of solution 2 (200 mM NaOH, 35 mM sodium dodecyl sulfate (SDS)) were added and mixed with the suspension by carefully inverting the tube top to bottom for 5x. The incubation was timed to exactly 5 min after which 2 mL of solution 3 (3 M potassium acetate, 2 M glacial acetic acid) were added. The inversion of the tube was repeated as described before and the mixtures was subjected to centrifugation at 21,000 xg at 4°C for at least 7 min until a sharp separation of a milky pellet from a clear supernatant was achieved. The supernatant was carefully transferred to a 15 mL collection tube. Collection tubes (1.5 mL capacity) were prepared holding 420 µL of isopropanol (≥ 99.9%, VLSI Grade, Roth, Germany) per tube to which 500 µL of the supernatant were added. All tubes were vigorously mixed and incubated for a minimum of 20 min at RT. Following centrifugation at 21,000 xg for 10 min at RT, the supernatant was thoroughly sucked off by a vacuum pump. Residual isopropanol was removed by centrifuging the tubes with an open lid for 2 min under vacuum pressure (960 mbar) or by leaving the tubes standing with an open lid at the working bench for at least 10 min. The remaining DNA pellets form each tube were pooled by consecutive resuspension in ddH_2O (total volume of 200 µL for 9-10 tubes, 180 µL for 7-8 tubes). The total DNA was further purified using the Wizard SV Gel and PCR kit (Promega, Madison, USA). All steps were conducted as instructed by the manufacturer. Final elution of the DNA was carried out with 35-60 µL of nuclease free water, re-applying the flow-through for a second elution. The vector concentration was measured at the NanoDrop™ 2000 (Thermo Fisher Scientific, Waltham, USA) and working aliquots of 0.5 µg/ µL were prepared.

In a small-scale approach (Mini-Prep), the overnight culture (3 mL) was added into one 2 mL collection tubes by two separate centrifugation steps. Pellets were dissolved in 200 µL of solution 1, lysis was performed by adding 200 µL of solution 2 and neutralization by adding 200 µL of solution 3. After centrifugation, supernatant was mixed with 420 µL isopropanol, and all other steps were performed as outlined above.

3.2 Cell culture

3.2.1 Cell lines and culturing

All human cell lines were cultured in T75 or T175 tissue culture (TC) flask (Sarstedt, Nümbrecht, Germany) at 37°C and 5% CO_2. The cells were transferred to a new TC flask upon covering an estimated 80-90% of the total surface area. An overview of the corresponding growth media for each cell line is listed in Table 25.

Table 25 List of human cell lines and corresponding culture media

Cell line	Culture Media
Caco2	500 mL MEM (M4655) 50 mL FBS Superior (S0615) 5 mL NEAA 100x (M7145) 1 mL Gentamycin 50 mg/mL, final 0.05% (G1272)
HT29 COLO357 Capan1 RT4	500 mL RPMI-1640 Medium (R8758) 50 mL FBS Superior (S0615) 5 mL Penicillin/Streptomycin 100x, final 1% (P4333)
HepG2 Huh7	500 mL DMEM, low glucose (D6046) 50 mL FBS Superior (S0615) 5 mL Penicillin/Streptomycin 100x, final 1% (P4333) 5 mL NEAA 100x (M7145)
Paca44 Panc1 HEK293 HEK293T	500 mL DMEM, high glucose (D6429) 50 mL FBS Superior (S0615) 5 mL Penicillin/Streptomycin 100x, final 1% (P4333)

Note: all culture media components were purchased from Sigma-Aldrich (Darmstadt, Germany); MEM: Minimum Essential Medium, FBS: Fetal Bovine Serum, NEAA: Non-essential Amino Acids, DMEM: Dulbecco's Modified Eagle's Medium

For the transfer of cells to a new flask or to a TC-plate at the beginning of an experiment, cells were washed with 10 mL (T75) or 15 mL (T175) pre-warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany), detached through a 5 min incubation in 1 mL or 3 mL trypsin ethylenediaminetetraacetic acid (EDTA) (Sigma- T3924, Aldrich, Darmstadt, Germany), respectively at 37°C, 5% CO_2 and re-suspended in 9 mL or 17 mL total growth media, respectively. The total number of cells was determined with a Neubauer-improved cell-counting chamber (0.0025 mm^2, Brand, Wertheim, Germany). For maintaining permanent cultures, a fraction of the previous culture was transferred to a fresh TC flask with fresh growth media (total volume: T75: 10 mL, T175: < 20 mL). The fraction of the previous culture was adjusted accordingly for each cell line to split the cells no more than every three to four days. For the start of a new transfection experiment, cells were seeded to reach 80% confluence on the next day. Table 26 provides an overview of the most frequently used conditions if not specified otherwise.

Table 26 Number of seeded cells by type of cell line and experimental format

Format	48 well	12 well	6 well	10cm dish
Surface area	1.1 cm^2	3.5 cm^2	9.6 cm^2	56.7 cm^2
Caco2	$0.05x10^6$	$0.125x10^6$	---	$3.4x10^6$
HepG2	$0.08x10^6$	$0.2x10^6$	---	$5.25x10^6$
Paca44	---	$0.075x10^6$	$0.35x10^6$	$2.8x10^6$
Panc1	---	---	$0.5x10^6$	$3.0x10^6$
HEK293	---	---	---	$3.3x10^6$
HEK293T	---	---	---	$2.9x10^6$

3.2.2 *SPINK1* RNA expression levels in different human cell lines

At the start of the experiment, all available cell lines were seeded in duplicates onto 6-well plates (TPP, Trasadingen, Switzerland). The cells were harvested either shortly before reaching full confluence (<100% of surface area) or after exceeding full confluence (covering >100% of surface area).

3.2.3 SPINK1 protein abundance in selected human cell lines

Huh7, Caco2, HepG2 and RT4 cells were cultured in T175 TC flask (Sarstedt, Nümbrecht, Germany) to a confluence of 70-80% until seeded on to 150 mm TC-dishes (Sarstedt, Nümbrecht, Germany) for the experiment. The SPINK1 protein

abundance was assessed over different incubation times (48 h, 72 h, 120 h) and while the cells were either in a dividing or a confluent state. For the dividing state, the cells were seeded with an initial cell number of $0.9x10^6$ cells/dish (Huh7, Caco2) $6x10^6$ cells/dish (HepG2) and $10x10^6$ cells/dish (RT4). For the confluent state, twice as many cells as for the dividing state were seeded. The cells were rinsed with 15 mL of pre- warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany) 24 h after seeding and 10 mL of serum free OptiMEM™ (GIBCO™, Amarillo, USA) was added to each dish.

3.2.4 Cell culturing for large scale nuclear extract (NE)

Cells were cultured in T175 TC flasks (Sarstedt, Nümbrecht, Germany) to a confluence of 70-80%. At the day of seeding (day 1), all cells from one T175 TC flask were transferred to five TC-dishes (150 mm, Sarstedt, Nümbrecht, Germany). Growing to a confluence of 80-90%, cells were harvested at day 4. This procedure was strictly applied to all rounds of nuclear extract (NE) preparation. Each round included 20 to 40 TC-dishes. For EMSA pre-experiments, three rounds of NE preparation were completed for each cell line and compared as biological replicates. To obtain enough NE material for the affinity chromatography (AC), five rounds were required for the HepG2 cell line whereas three rounds of NE preparation were sufficient for Caco2 cells. The total protein content of each NE of one round was measured by BioRad Protein Assay (BioRad Laboratories, Hercules, USA, see section 3.3.8), prior to pooling the NE for affinity chromatography (AC) into one sample. After confirming the total protein content of the pooled NE, all consecutive AC experiments were conducted using the pooled material.

3.2.5 Transient overexpression of acinar transcription factors

Panc1 and Paca44 cells were cultured as describe before. For the experiment, $0.5x10^6$ cells were seeded into one well of a 6-well plate (TPP, Trasadingen, Switzerland) with each well containing a total volume of 2 mL growth media. Following a 24 h incubation, the cells were transfected with the over-expression vectors by lipofection using Lipofectamine™2000 (Invitrogen™, Waltham, USA). The lipofection mixture was prepared as follows: working at RT, two 1.5 mL collection tubes were each prepared with 250 µL OptiMEM™ (GIBCO™, Amarillo,

USA). To the first tube, 10 µL of Lipofectamine™2000 were added whereas a total amount of 4 µg (0.5 µg/ µL working stock) vector was added to the second tube. For co-transfection experiments, the total amount of vector was kept constant whereas the co-transfected vectors were adjusted to be included at an equal proportion (*e.g.,* 2 vectors: each 2 µg; 3 vectors: each 1.3 µg; 4 vectors: each 1 µg). Both mixtures were incubated for 5 min at RT. Afterwards, the content of the first tube was added to the second tube and mixed well. The combined mixtures were incubated for 20 min at RT. Please note that if using several conditions of vector combinations in parallel, a master mix of the OptiMEM™ (GIBCO™, Amarillo, USA) and Lipofectamine™2000 (Invitrogen™, Waltham, USA) mixture was prepared and distributed to each combination of vectors. Before adding 500 µL of the final transfection mixture to each well of 6-well plate, 500 µL of the growth media were removed. Following a 6 h incubation, the growth media including the transfection mixture was removed carefully and each well was washed with 1 mL of pre-warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany). Finally, 2 mL of fresh and pre-warmed OptiMEM™ (GIBCO™, Amarillo, USA) were added to each well and the cells were incubated for 72 h in total after transfection. When working in larger formats like *e.g.,* 100 mM TC-dishes for NE preparation, the volumes were scaled up accordingly. Table 3 provides an overview on the used expression vectors.

3.2.6 Transient knock down of *rs148911734*-associated transcription factors

3.2.6.1 Single siRNA knock downs in Caco2, HepG2 and Paca44

One day before transfection, cells were seeded on 12-well plates (compare Table 26) to reach 80% confluence the next day. The cells were prepared for transfection 24 h after seeding. Each well was washed with PBS (D8537, Sigma-Aldrich, Darmstadt, Germany) and antibiotics-free growth medium was added.

Caco2 cells were transfected with a final concentration of 100 nM siRNA using the DharmaFECT 1 transfection reagent and harvested for total RNA extraction 48 h-post transfection.

HepG2 cells were transfected with a final concentration of 50 nM siRNA using the DharmaFECT 4 transfection reagent. The cells were harvested 72 h-post transfection.

siRNA transfection of Paca44 cells was carried out in a two-step-manner using Lipofectamine™2000 (Invitrogen™, Waltham, USA). In a first step, 24 h after seeding, 25 nM siRNA were applied. Following a 48 h incubation, cells were transfected with an additional 25 nM siRNA and incubated for another 48 h before cell harvest.

Following a 6 h incubation after the addition of the transfection mix, each well was washed with pre-warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany) and whole growth medium was added. The experiments were carried out in technical duplicates and the transfection conditions refer to one well. The siRNAs are listed in Table 5.

3.2.6.2 Combined siRNA knock downs with luciferase expression vectors in Caco2 and HepG2

One day before transfection, cells were seeded on 48-well plates (compare Table 26) to reach 80% confluence the next day.

After 24 h, HepG2 cells were prepared for transfection. Each well was washed with PBS (D8537, Sigma-Aldrich, Darmstadt, Germany) and antibiotics-free growth medium was added. HepG2 cells were transfected with a final concentration of 12.5 nM siRNA and DharmaFECT 4. Following an incubation of 24 h, 50 ng of the pRL-Ubi and 100 ng of the corresponding luciferase-firefly vector were transfected using Lipofectamine™2000 (Invitrogen™, Waltham, USA).

For Caco2 cells, the transfection to a final siRNA concentration of 12.5 nM, 50 ng of the pRL-Ubi and 100 ng of the corresponding luciferase-firefly vector was carried out in one step using Lipofectamine™2000 (Invitrogen™, Waltham, USA) as transfection reagent. After a 6 h incubation, the cells were washed with PBS and fresh OptiMEM™ (GIBCO™, Amarillo, USA) was added.

Luciferase activity was measured 24 h after the last transfection in both cell lines. The experiments were carried out in biological and technical triplicates and the transfection conditions refer to one well. The vector constructs are listed in Table 4.

3.2.6.3 Combined siRNA knock downs with acinar transcription factors in Paca44

One day before transfection, cells were seeded on 6-well plates (compare Table 26) to reach 80% confluence the next day. A two-step approach was followed for the combined transcription factor overexpression and siRNA knock down experiments. First, 24 h after seeding, each well was washed with PBS (D8537, Sigma-Aldrich, Darmstadt, Germany), antibiotics-free growth medium was added and cells were transfected with 50 nM siRNA using Lipofectamine™2000 (Invitrogen™, Waltham, USA). Following a second 24 h incubation, the medium was replaced with fresh full growth medium. Second, 24 h after the medium change, the cells were transfected with 50 nM siRNA and a total of 1 µg of overexpression vector. After 6 h, the cells were washed with PBS (D8537, Sigma-Aldrich, Darmstadt, Germany) and fresh OptiMEM™ (GIBCO™, Amarillo, USA) was added. Cell were harvested 72 h after the second transfection for total RNA extraction. The transfection conditions refer to one well.

3.3 Sample extraction and preparation

3.3.1 Genomic DNA

Whole human blood of 638 Swiss individuals was provided by PD Dr. med. Kaspar Truninger (Basel, Ethics Approval EK: 276/13). For genomic DNA extraction from whole human blood the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) was used according to the manufacturer's instructions (appendix) with the following exceptions: 400 µL of whole blood from EDTA-collection tubes were vigorously mixed and combined with 40 µL of Protease and 400 µL of AL (lysis) buffer. After a 10 min incubation at 56°C, 400 µL of 100% ethanol (analytical grade, Roth, Germany) were added to the lysis reaction and the total volume (1,240 µL) was loaded in two steps onto the same column. Centrifugation was performed at 6,000 xg and the applied volume held 620 µL for each step. DNA elution was carried out in three steps. During the first elution 100 µL AE (elution) buffer was applied, while

50 µL were used for the subsequent second and third elution. Elution 1 and 2 were pooled. Elution 3 was transferred to a separated collection tube and used for genotyping as outlined in section 3.6.3. The DNA concentration was determined at the NanoDrop™ 2000 (Thermo Fisher Scientific, Waltham, USA) and stored at -20°C until further use.

3.3.2 Total RNA

For total RNA extraction, the RNeasy® kit (Qiagen, Hilden, Germany) was used. Prior to addition of 350 µL RLT buffer to each well of 6-well plate (TPP, Trasadingen, Switzerland), the conditioned media was carefully removed, and the cells were carefully washed with 1 mL of pre-warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany). Upon adding 350 µL RLT buffer, the cells were immediately scratched off the TC-plate, transferred to a 1.5 mL collection tube and placed on ice until the collection of each sample was completed. For cell homogenization, the collected material was transferred onto a QIAshredder spin column (Qiagen, Hilden, Germany) placed in a 2 mL collection tube. The sample was spun through during 2 min centrifugation at full speed. The flow through was mixed with 1 volume (350 µL) of 70% ethanol and applied onto a gDNA eliminator spin column. All further steps were carried out according to the manufacturer's instructions (appendix). Centrifugation steps were carried out at 15°C whereas all other steps were issued at RT. For the final elution, 35 µL of RNase-free water were applied onto the column. The first eluent was re-applied onto the column for a second elution to increase the RNA yield. RNA concentrations were measured at the NanoDrop™ 2000 (Thermo Fisher Scientific, Waltham, USA) and adjusted to a 100 ng/µL working solution. RNA samples were shock frozen in liquid nitrogen and stored at -80°C.

3.3.3 Reverse transcription of total RNA and cDNA synthesis

For the reverse transcription of RNA into cDNA, the QuantiTect® Reverse Transcription Kit (Qiagen, Hilden, Germany) was used, including 5 µL of the 100 ng/µL working stock (0.5 µg) of the RNA sample in a total reaction volume of 20 µL. All instructions were followed as provided by the manual (appendix). The gDNA elimination reaction was set to 5 min within the recommended range of 2 to 10 min. The final complementary DNA (cDNA) was diluted in a 1/10 ratio with water

(ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany) to obtain the working stock.

3.3.4 Reverse transcription of *SPINK1* RNA and cDNA synthesis

RNA of human intestinal biopsy of five *SPINK1* p.N34S carriers and five wild type controls was provided by PD Dr. med. Kaspar Truninger. For the human samples, 100 ng of RNA were reverse transcribed and used undiluted for further experiments. For total RNA samples of cell culture experiment, 500 ng were used (5 µL of 100 ng/µL stock) and the cDNA was diluted 1/10 with water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany). The instructions of the QuantiTect® Reverse Transcription Kit (Qiagen, Hilden, Germany) were followed (appendix). Only instead of the provided RT Primer Mix, a *SPINK1*-specific primer (2 µM, SPINK1-UTR-RevA, see Table 6) was used.

3.3.5 Total protein from cell culture supernatant

At least 1 mL of the conditioned media was collected, transferred to a fresh 1.5 mL collection tube and placed on ice. Cell debris was collected at the bottom of the tube after centrifugation at 200 xg for 5 min, the supernatant was transferred to a fresh 1.5 mL collection tube, shock frozen in liquid nitrogen and stored at -80°C until further use.

3.3.6 Total protein from cell lysate

After removing the conditioned media and washing the cells of each well of a 6-well plate with 1 mL pre-warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany), the cells were scratched off the plate in 0.5 mL of PBS with a cell scraper, transferred to a 1.5 mL collection tube and immediately placed on ice. Centrifugation at 500 xg for 3 min was used to collect the cell material at the bottom of the tube and to remove the PBS supernatant. The cell pellets were shock frozen in liquid nitrogen and stored at -80°C. For total lysate preparation, 50 µL of radio-immunoprecipitation assay (RIPA) lysis buffer (stock buffer: 10 mM Tris-HCL pH 7.4, 150 mM NaCl, 5 mM EDTA, freshly added: 1% TritonX100, 10 mM phenylmethylsulphonyl fluoride (PMSF)) were added to each sample while thawing the samples on ice. Exposure to sonification (UP 200s, Dr. Hielscher GmbH, 5 beats, cycle 0.5, amplitude 40) was

carried out two times, including a sample cool down on ice after each sequential exposure of 5 beats. Cell debris was collected at the bottom of each tube after 5 min centrifugation at 1,000 xg. The supernatant was transferred to a new collection tube, shock frozen in liquid nitrogen and stored at -80°C until further use. Alternatively, 1 mL insulin syringes (B. Braun, Melsungen, Germany) were pre-rinsed with RIPA buffer and the cell pellet was re-suspended in 50 µL of RIPA buffer. Always keeping the sample on ice, the cells were passed through the syringe 20 times. After centrifugation at 500 xg for 3 min (4°C) the lysate was transferred to a fresh collection tube.

3.3.7 Nuclear protein

At the day of harvest, PBS (D8537, Sigma-Aldrich, Darmstadt, Germany) was pre-warmed to 37°C whereas buffers and collection tubes were kept on ice and centrifuges were cooled to 4°C. The cells were harvested in batches of maximal 10 TC-dishes at a time. Using a vacuum pump, the conditioned media of each TC- dish was removed thoroughly and 15 mL of pre-warmed PBS were added to wash off dead cells and residual culture media. Working dish by dish, the pre-warmed PBS was removed completely, the dish was transferred onto ice and 1 mL of buffer A (hypotonic buffer: swelling of the cytoplasm of the cell; 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Hepes) pH 8, 10 mM KCl, 0.1 mM EDTA, 0.1 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetra acetic acid (EGTA), 0.5 mM PMSF, 1 mM dithiothreitol (DTT), 1x Protease/Phosphatase Inhibitor Cocktail (Roche, Basel, Switzerland) was added. The cells were scraped off the dish and immediately transferred to a collection tube on ice. Half of the dishes were pooled into one shared collection tube, leaving two collection tubes for subsequent centrifugation at 21,000 xg at the Sorvall LYNX 400 centrifuge (Thermo Fisher Scientific, Waltham, USA). Following 30 min incubation in buffer A on ice, 10% nonyl phenoxypolyethoxylethanol (NP40) as detergent was added and the cells were vigorous shaken with the help of a vortexer for 10 sec. The total amount of 10% NP40 was calculated based on the assumption to use 40 µL of 10% NP40 for the cells of one TC-dish. To separate the intact cell nuclei from the cell debris and to remove the NP40, centrifugation at 21,000 xg for 5 min at 4°C was carried out. The supernatant was removed, and 2 mL of fresh buffer A were added to the remaining pellet. To wash the cell nuclei, the pellet was thoroughly re-suspended with a 10 mL

serological pipette (Sarstedt, Nümbrecht, Germany) and centrifuged again at 21,000 xg for 5 min. The washing step was repeated, however this time the suspension was transferred in 500 µL aliquots to 1.5 mL collection tubes and centrifuged at the centrifuge 5810 / 5415R at 21,000 xg for 5 min (Eppendorf, Hamburg, Germany). After completely removing buffer A, buffer C (hypertonic buffer, release of nuclear proteins; 20 mM Hepes, 0.4 mM NaCl, 1 mM EDTA, 1 mM EGTA, 2.7 M glycerol (20%), 1 mM PMSF, 1 mM DTT, 1x Protease/Phosphatase Inhibitor Cocktail (Roche, Basel, Switzerland) was added at an estimated volume of one-fold of the pellet size. The pellets were re-suspended thoroughly and incubated on a shaker at 2,000 xg at 4°C for a minimum of 45 min. A final centrifugation step at 21,000 xg for 5 min was carried out. The supernatant containing the nuclear proteins was transferred and pooled into a fresh collection tube prior to shock freezing in liquid nitrogen. The NE were stored at -80°C until further use. The total protein concentration was determined by BioRad Protein Assay (BioRad Laboratories, Hercules, USA) which is based on the method of Bradford. The procedure was followed as described in section 3.3.8 The bovine serum albumin (BSA) standard curve was prepared in buffer C.

Please note, that the above-described procedure applies for the large-scale preparation of NE for EMSA and AC. For follow-up experiments such as *e.g.,* Super-Shift EMSA or NE from overexpression experiments, we used the Nuclear Extraction Kit (Active Motif, Carlsbad, USA) according to the manufacturer's instructions (appendix).

3.3.8 Determination of total protein concentration

The total protein content was determined by BioRad Protein Assay which is based on the method of Bradford (Bradford *et al*., 1976). Working in technical triplicates, 1 µL of the sample and 2 µL of the BSA standard (BSA fraction V, blotting grade, Applichem Lifescience, Darmstadt, Germany) were added to one well of a 96-well plate (TPP, Trasadingen, Switzerland) on ice. The BSA standard curve was prepared in the corresponding lysis buffer of the sample. In a serial of 1:1 dilution starting with S1 (400 µL of 10 mg/mL BSA stock in 600 uL lysis buffer) a range of reference values from S1 (4 mg/mL) to S7 (0.0674 mg/mL) was covered. S8 constituted of lysis buffer only and was used for blank measurements. The

application of 200 µL of a 1/5 dilution of the Protein Assay Dye Reagent Concentrate to each well were followed by a 10 min incubation at RT. The binding of the dye to the protein causes a shift in the absorbance maximum from 465 nm to 595 nm (Bradford *et al*., 1976) which was measured at 595 nm at the VarioScan™ Multimode Reader (Thermo Fisher Scientific, Waltham, USA) and the protein concentration of each sample was determined with the help of the BSA standard curve. The blank values were subtracted from all measurements. Accounting for the different volumes of the sample and the standard, the concentration of each sample was multiplied by the dilution factor 2.

3.4 Enzyme-based assays

3.4.1 Enzyme-linked immunosorbent assay (ELISA)

For SPINK1 protein measurements from cell culture supernatant and total lysate, the Human SPINK1 DuoSet ELISA Kit (DY7496-05, RD Systems, Minneapolis, USA) was used. All instructors were followed as provided by the manufacturer (appendix). All samples, including the standard (SPINK1 recombinant protein from *E. coli*), were measured in duplicates. Instead of 400 µL wash buffer as suggested in the manufacturer's protocol, 300 µL were applied. The supernatant was diluted in OptiMEM™ (GIBCO™, Amarillo, USA) in three dilution steps (DS1-DS3). DS1 refers to a 1/100 dilution of the original sample. D2 is the 1/5 dilution of DS1 (1/500 dilution of original sample) whereas DS3 was obtained through a 1/2 dilution of DS2 (1/1,000 dilution of the original sample). Of DS2 and DS3, 100 µL were measured with the kit. For the total lysate samples, the following dilution steps were applied (DL1-DL3). First, the original sample was diluted 1/10 (DL1). From DL1, a second 1/10 dilution was derived by mixing a total of 10 µg (approx. 50 µL) in a total volume of 500 µL (DL2, total protein 1 µg/µL, 1/100 dilution of original sample). DL2 was further diluted 1/5 for DL3 (total protein 0.2 µg/µL, 1/500 dilution of original sample). DL4 (total protein 0.1 µg/µL, 1/1,000 dilution of original sample) was obtained by another 1/2 dilution of the DL3. Of DL3 and DL4, 100 µL were measured with the kit. The above-described dilution steps were applied to stay above a minimal pipetting volume of 10 µL and to normalize the lysate samples to 10 µg of total protein with as little pipetting steps as possible. In doing so, the accuracy and standardization of sample preparation was handled in the best possible way.

3.4.2 Luciferase reporter gene assay

Prior to cell lysis, the cells were washed in pre-warmed PBS (D8537, Sigma-Aldrich, Darmstadt, Germany). For each well of a 48-well plate, 40 µL of 1x passive lysis buffer (5x, diluted to 1x with PBS) were added and cell lysis was carried out during a 10 min incubation at RT at 1,100 rpm. Subsequently, the luciferase activity measurement was performed at RT with the SIRIUS Luminometer (Serial No. 0627, Berthold Detection Systems, Pforzheim, Germany) as follows: after blank measurement of the empty tube (PS tubes Catalog No. 115101, Greiner Bio-One, Kremsmünster, Austria) cells were scratched from the plate and 5 µL were transferred to the plastic tube. After the addition of 25 µL of the **firefly** buffer (D-luciferin: 470 µM; adenosine triphosphate (ATP): 530 µM; coenzyme A: 270 µM; DTT: 33.3 mM; tricine: 20 mM; $MgSO_4Mg(OH)_2$ x $5H_2O$: 2.67 mM; EDTA: 0.1 mM; pH 7.8) were added and light emission through firefly activity was measured. Immediately after, 25 µL of the **renilla** buffer (coelenterazin: 1.43 µM; Na_2EDTA: 2.2 mM; K_xPO_4 (pH 5.1): 0.22 M; BSA: 0.44 mg/mL; NaCl: 1.1 M; NaN_3: 1.3 mM; pH 5.0) were added to the tube and the renilla activity was measured. The measurement time was set to 10 sec and the delay time to 3 sec.

The promoter activity was calculated by first subtracting the blank value from both luciferase (firefly and renilla) activity measurements. Second, the firefly activity was divided by the renilla activity (internal normalization for the vector transfection efficiency). Third, the average of the normalized firefly activity of each experiment was referenced to the enzyme activity of the non-target siRNA and pGL4.22_SPINK1-Promotor_Luiferase transfected cells (baseline condition).

3.5 Synthetic oligonucleotides

3.5.1 With Cy5-label for EMSA

The forward and reverse strand of a 22 bp flanking sequence upstream and downstream of each SNP of interest (total length of 45 bp) were order for the risk and non-risk allele from TIB MOLBIOL, Berlin, Germany (Table 9). Each forward strand was labeled with a Cy5-fluorophore at the 5'-end. In order to obtain a double-stranded (ds)-oligonucleotide, 7 µL of the forward and 7 µL of the reverse strand (both 100 µM) were mixed with 6 µL of Tris EDTA (TE) buffer (10 mM Tris, 1 mM Tris HCL, 1 mM Na_2EDTA, pH 7.6), heated to at least 10°C above the strands

melting temperature (here: 95°C for *rs142703147* and *rs148276928*, 88°C for *rs17107316* and *rs148911734*) and subjected to a controlled cool down at 3°C/10 min to a final temperature of 10°C. To purify the ds-oligonucleotides from un-annealed single-stranded (ss)-oligonucleotides, the annealing reaction (20 µL) was mixed with 4 µL of 10x loading buffer (250 mM Tris-HCL pH 7.5, 40% glycerol, 0.2% Orange G (Roth, Karlsruhe, Germany) and loaded onto a 12% acrylamide gel (WxH 17.0 x 18.0 cm, 1x tris base, boric acid and EDTA (TBE) pH 8.3 (90 mM Tris base, 88.8 mM boric acid, 0.1 M EDTA (pH 8.5); 12% acrylamide (v/v, ROTIPHORESE®Gel 40 (37.5:1), Roth, Karlsruhe, Germany), 2.5% glycerol (v/v), 0.075% ammonium persulfate (APS) (v/v), 0.5% tetramethyl-ethylenediamine (TEMED) (v/v)). Further, ss-oligonucleotide controls (7 µL of ss-reverse strand, 13 µL water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany), 4 µL 10x loading buffer) were loaded next to the corresponding annealing reaction. Please note that, to avoid cross contamination, only every third lane of the gel was used, leaving at least two pockets empty. During a pre-run, the acrylamide gels were calibrated at 200 V and 4°C for 30 min prior to loading the samples. Electrophoresis was performed under the same conditions for at least 3 h. Subsequently, the corresponding ds-oligonucleotides bands were visualized by UV light exposure on a silica plate, cut out of the gel and transferred to a light protected 1.5 mL collection tube. TE buffer was added just in the right amount to fully cover the gel piece (approx. 120 µL) to wash out the ds-oligonucleotides during an overnight incubation at 37°C and 1,000 rpm. The next day, the tip of a 1 mL filter tip was removed, and the remaining filter was placed into a fresh light protected 1.5 mL collection tube. The sample was added on top of the filter and spun through the filter during a 1 min centrifugation step at 3,381 xg. Subsequently, 30 µL fresh TE buffer were added onto the filter and centrifugation was repeated. Concentrations of ds-oligonucleotides were measured at the Nanodrop™ 2000 (Thermo Fisher Scientific, Waltham, USA). Aliquots were adjusted to precisely 40 ng/µL. From the 40 ng/µL stock, a final 1 ng/µL working solution was generated.

3.5.2 With biotin label for AC

The procedure was followed as described in section 3.5.1 with the following differences: instead of a Cy5-label, synthetic oligonucleotides carried a biotin label at the 5'-end. Also, instead of 1 ng/µL working solution, an 8.2 ng/µL solution of the

biotinylated oligonucleotides was used for the binding reaction to the streptavidin beads.

3.5.3 Agarose gels

3.5.3.1 Agarose gel preparation and electrophoresis

Agarose gels were prepared in 1x tris base, acetic acid and EDTA (TAE) buffer (50x TAE: 2 M Tris, 64 mM EDTA, 5.7% v/v acetic acid) and the corresponding amount of agarose (Nippon Genetics, Düren, Germany, w/v). The mixture was heated in a microwave until the agarose had dissolved completely. HDGreen+ (Intas, Göttingen, Germany) was added to the mixture to later stain the DNA and the mixture was left to polymerize at RT for at least 30 min. PCR products were mixed with DNA Gel Loading Dye (6x, Thermo Fisher Scientific, Waltham, USA). Alongside the sample, a DNA ladder of choice *e.g.,* 1-5 µL of a GeneRuler DNA Ladder Mix (Thermo Fisher Scientific, Waltham, USA) were loaded. Gel electrophoresis was performed at 100 V for 30-75 min depending on the size and percentage of the gel and the size of the PCR product.

3.5.3.2 PCR product clean up

Using the FastGene FAS V Gel Documentation System (Nippon Genetics, Düren, Germany) PCR products were visualized, and the DNA band of interest was cut out of the gel with a razor. The gel slice was collected in a 1.5 mL collection tube and the PCR product was extracted following the instructions of the Wizard SV Gel and PCR kit (appendix, Promega, Madison, USA).

3.5.4 SDS-Polyacrylamide-Gel-Electrophoresis (PAGE) and Western blot

From single transfection experiments, 10 µg of total protein from cell extracts were loaded. The corresponding volume of each cell extract was topped up with double distilled water to a total volume of 15 µL. To each sample, 5 µL of 4x Laemmli buffer (227 mM SDS, 75 mM Tris, 20% glycerol (v/v), 20% β-mercaptoethanol (v/v), 0.4% bromphenol blue (m/v)) were added and mixed well. A final volume of 20 µL per sample were loaded in each lane along with 0.5 µL of PAGE ruler (Thermo Fisher Scientific, Waltham, USA) in a separate lane. The SDS-polyacrylamide gel consisted of a 10% separation gel (buffer separation gel: 372 mM Tris (pH 8.8),

0.1% SDS (v/v); 10% acrylamide (v/v, ROTIPHORESE®Gel 30 (37.5:1), Roth, Karlsruhe, Germany), 0.08% TEMED (v/v), 0.2% APS (v/v)) and a 5% stacking gel (buffer: 140 mM Tris (pH 6.8), 0.1% SDS (v/v); 5% acrylamide (v/v, ROTIPHORESE®Gel 30 (37.5:1), Roth, Karlsruhe, Germany), 0.08% TEMED (v/v), 0.06% APS (v/v). SDS-PAGE was performed at 120 V for initial 10 min followed by 160 V for 45 min (SDS running buffer: 25 mM Tris, 192 mM glycine, 3.5 mM SDS). The separated proteins were transferred onto a nitrocellulose membrane (GE Healthcare, Solingen, Germany) at 360 mA for 30 min by wet blot (transfer buffer: 20 mM Tris, 150 mM glycine, 20% methanol (v/v), 0.02% SDS (v/v)). The membrane was incubated at RT in 5% BSA blocking solution (5 g of BSA fraction V, blotting grade (Applichem Lifescience, Darmstadt, Germany)) in a final volume of 100 mL PBS-T (1x PBS with with 0.1% Tween20 (Serva, Heidelberg, Germany)). The primary antibodies were prepared in a total volume of 5 mL of 5% BSA blocking solution and incubated overnight at 4°C. The following day, the membrane was washed 3 times for 10 min in 10 mL PBS-T and incubated with the secondary antibodies in a total volume of 5 mL PBS-T or 5% BSA blocking solution with 0.1% Tween20. Following a 2 h incubation at RT, the membrane was washed twice for 10 min in 10 mL PBS-T and once for 10 min in 10 mL PBS. All incubation steps were placed on a shaker. Imaging of the protein bands was performed with the LI-COR Odyssey fluorescence scanner (Licor, Lincoln, USA) or the Sapphire Biomolecular Imager (Azure Biosystems, Dublin, USA). An overview of the applied primary and secondary antibodies and corresponding dilutions is provided in Table 20 and Table 21.

3.5.5 Electrophoretic mobility shift assay (EMSA)

3.5.5.1 Baseline condition

In a total volume of 9 µL per reaction, 3 µg (exception Paca44: 5 µg) of NE were diluted with water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany) to a total volume of 6.2 µL and combined with 350 ng (0.5 µL) of poly-deoxy-inosinic-deoxycytidylic acid sodium salt ((polydIdC), Sigma-Aldrich, Darmstadt, Germany, 700 ng/µL working stock)) and 2.3 µL 5x binding buffer (BB: 20% glycerol, 5 mM $MgCl_2$, 50 mM Tirs-HCl pH 7.5, 2.5 mM EDTA, 2.5 mM DTT). After a short spin down at 4°C, the mixture was incubated for 10 min on ice.

Subsequently, 1 µL of the Cy5-labeled ds-oligonucleotides (1 ng/µL working stock) were added, the sample was spun down at 4°C and incubated light protected for 20 min on ice. After the addition of 1.5 µL of the 10x loading buffer (250 mM Tris-HCl, pH 7.5; 0.2% Orange G (w/v); 40% Glycerol(v/v)) and another spin down of the sample, the total reaction volume of 11.5 µL was loaded onto a 5.3% acrylamide gel (WxH 17.0 x 18.0 cm, 1x TBE buffer pH 8.3 (90 mM Tris base, 88.8 mM boric acid, 0.1 M EDTA (pH 8.5); 5.3% acrylamide (v/v, ROTIPHORESE®Gel 40 (37.5:1), Roth, Karlsruhe, Germany) 2.5% glycerol (v/v), 0.075% APS (v/v), 0.5% TEMED (v/v)). For each set of samples, the reaction mixture including the non-risk allele was always applied first, followed by the risk allele in the next lane. The samples were subjected to a native gel electrophoresis at 200-250 V and 4°C for at least 3 h until the orange dye of the loading buffer reached the lower end of the gel. Please note that before loading the samples, the gels have been pre-run, in 0.5% TBE buffer (diluted with double distilled water from 5x TB: pH 8.3: 450 mM Tris base, 444 mM boric acid, 0.5 M EDTA (pH 8.5)) for at least 30 min under the same conditions. Protein-DNA complexes and unbound ds-oligonucleotides (reaction and loading control) were imaged at high sensitivity at 633 nm and 100 µm resolution on the Typhoon TRIO+(R) Variable Mode Imager (Amersham Biosciences Europe, Freiburg, Germany).

3.5.5.2 Post-affinity chromatograpy control

Control aliquots of the RX (all nuclear proteins), SN-cplx (all unbound nuclear proteins) and elution fractions (nuclear proteins eluted from ds-oligonucleotides at the given salt concentration) were placed aside and kept at 4°C along the AC procedure. For the EMSA control run, 1.5 µL of the RX and 2 µL of the SN-cplx were combined with 10 µL of washing buffer (WB), whereas 7 µL of the elution fractions were combined with 15 µL of 1x BB. Finally, 1.5 µL of the Cy5-labeled ds-oligonucleotides (1 ng/µL) of the non-risk or risk allele were added to the corresponding reaction. All following steps were performed as described in the previous section.

3.5.5.3 Super-shift assay

The sample preparation was conducted as outlined in section 3.5.5.1 and expanded by the following step: After the 10 min incubation at 4°C of 5x BB, polydIdC and NE, primary antibodies directed against the protein of interest were added to the reaction and the whole mixture was incubated for an additional 40 min at 4°C. Subsequently, the Cy5-labeled ds-oligonucleotides (1 ng/µL) were added and the procedure was followed as described above. A control reaction with an unspecific antibody from the same species and isotype was included for each reaction with a targeted primary antibody. The corresponding antibodies are listed in Table 19.

3.6 PCR-based assays

3.6.1 PCR clean up and Sanger sequencing

For Sanger sequencing, 0.25 µL of Exonuclease I (NEB, Ipswich, USA) and 0.25 µL of Antarctic Phosphatase (NEB, Ipswich, USA) were combined with 1.5 µL water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany) and directly added to the PCR products. Digestion of PCR residues was performed at 37°C for 40 min and enzymes were inactivated at 85°C for 20 min. Of the digested PCR product, 2 µL were diluted in 13 µL water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany) and submitted to sequencing with 2 µL of the corresponding sequencing primer at Eurofins Genomics (Ebersberg, Germany) on an ABI 3730xl fluorescence sequencer.

3.6.2 Cloning of GATA6-p.P.K484R

3.6.2.1 PCR mutagenesis

For the exchange of lysine (K) at the amino acid position 484 (*JASSA* (http://www.jassa.fr) predicted SUMOylation site) to arginine (R) in the GATA6 coding sequence, mutagenesis primer were designed as follows: the forward primer was located upstream of the indicated position and carried a *HindIII* restriction site (pcDNA3.1-*HindIII*-F). The reverse primer harbored a nucleotide exchange at the indicated position, changing the triplet code from TTT (lysin) to TCT (arginine) and carried a *EcoRI* restriction site (GATA6-P.K484R-*EcoRI*-R). Both primers were designed with a 6-nucleotide-overhang next to the restriction site at the 5’ position (for primer sequence see Table 8).

The mutagenesis PCR was composed in a total volume of 100 µL. As template DNA, 1 µL of the pcDNA3.1_GATA6_FLAG vector (200 ng/µL) was used. The reaction mixture included 68.2 µL of water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany), 22 µL of the SuperFi II buffer (Thermo Fisher Scientific, Waltham, USA) 2.2 µL of dNTPs (10 mM, dNTP-Set1, Catalogue No. K039.1/2, Roth, Karlsruhe, Germany), 2.2 µL of each the forward and the reverse primer (pcDNA3.1-*HindIII*-F and GATA6-P.K484R-*EcoRI*-R, 10 µM) and 2.2 µL of the SuperFi II DNA polymerase (Thermo Fisher Scientific, Waltham, USA).

The PCR conditions were set to an initial denaturation at 98°C for 30 sec, followed by 45 cycles of denaturation at 98°C for 10 sec, primer annealing at 60°C for 10 sec and primer extension at 72°C for 105 sec, terminating with a final elongation step at 72°C for 5 min.

The PCR product was mixed with 6x DNA Gel Loading Dye (R1161, Thermo Fisher Scientific, Waltham, USA) and loaded onto a 2% agarose gel. Gel electrophoresis was carried out at 95 V for approximately 1 h. The expected PCR product size was 1,508 bp. The corresponding band was cut out of the gel and purified using the Promega Wizard clean up kit (appendix) according to the manufacturer's protocol. The final elution volume was 38 µL.

3.6.2.2 Digestion of the mutated insert sequence and the vector

In a total volume of 40 µL, the 33 µL of mutagenesis PCR product or 6 µg of the pcDNA3.1_GATA6_FLAG vector, respectively, were digested with each 1.5 µL of the restriction enzymes *HindIII* and *EcoRI* (both Fast Digest, Thermo Fisher Scientific, Waltham, USA) and 4 µL of the enzyme buffer provided by the supplier during 1 h incubation at 37°C.

The digested products were separated on a 2% agarose gel for 1 h at 95 V. The expected product sizes for the vector (5,670 bp) and the insert (1,580 bp) were cut out of the gel and purified with the Promega Wizard clean up kit (appendix). The final elution volume was 30 µL. The corresponding DNA concentrations were determined at the Nanodrop™ 2000 (Thermo Fisher Scientific, Waltham, USA).

3.6.2.3 Ligation of the digested vector and the mutated insert

The ligation mix was prepared on a total volume of 20 µL and constituted of 100 ng digested vector and 21 ng insert (100 ng vector*insert [kbp])/vector [kbp]). For the negative control, water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany) instead of the insert was used. For the overnight incubation at RT, 2 µL of the ligase buffer and 1 µL of the T4 ligase (both Thermo Fisher Scientific, Waltham, USA) were included.

3.6.2.4 Transformation and vector isolation

For the vector transformation, 3 µL of the ligation product were carefully added to 25 µL of Stellar™ Competent Cells (Clontech Laboratories/Takara Bio Company, Göteborg, Sweden) and incubated on ice for 30 min. A heat shock of 42°C for 1 min was followed by a 1 min incubation on ice and the subsequent addition of 250 µL room-temperature SOC outgrowth medium (#B9020S, NEB, Ipswich, USA). The starting culture was incubated for 1 h at 37°C and 165 rpm on a shaker and distributed onto 1% ampicillin containing LB bacterial culture plates for overnight incubation at 37°C.

The next day, single colonies were inoculated in 3 mL LB media and 0.1% ampicillin for a second overnight incubation at 37°C shaking at 165 rpm.

In the morning, 3 µL of each overnight culture was transferred onto 0.1% ampicillin containing LB bacterial culture plates (master plate). The plate would serve to save some of the material for future vector amplification upon verification of the correct vector construct. To this end, the vector was purified from the remaining culture as outlined in section 3.1.3 and sent for control sequencing to Eurofins Genomics (Ebersberg, Germany). To cover the mutagenesis site, the sequencing primer GATA6-P.K484R-F-Seq (see Table 8) was used.

3.6.3 *SPINK1* p.N34S genotyping

PCR was performed in a total volume of 22 µL as follows: 2 µL DNA template (Elution 3), 1.0 µL of the forward and the reverse primer (10 µM, see Table 10), 4 µL of MyTag® 5x Buffer, and 0.15 µL of MyTag® polymerase (both Bioline, London,

UK). The PCR conditions were set to an initial denaturation at 95°C for 1 min, followed by 50 cycles of denaturation at 95°C for 20 sec, primer annealing at 56°C for 40 sec and primer extension at 72°C for 1 min, terminating with a final elongation step at 72°C for 10 min.

For the melting curve analysis at the LightCycler480-II (Roche, Basel, Switzerland), 6 µL of the PCR product were combined with each 0.5 µL of the LC and FL probe (0.3 OD in 120 µL, see Table 10) to a total reaction volume of 7 µL. Melting curve conditions were set to an initial denaturation at 95°C for 30 sec, followed by a cool-down to 35°C for 1 min and a gradual increase to 80°C at a ramp rate of 2°C/sec.

3.6.4 Real time polymerase chain reaction (qPCR)

3.6.4.1 SYBR Green

In a total reaction volume of 10 µL, each reaction was composed of 5 µL of the 5x qPCR BIO SyGreen Mix Lo-ROX mix (Nippon Genetics, Düren, Germany), 1.0 µL of the forward and the reverse primer (25µM; see Table 11) and 2 µL template cDNA. The PCR conditions were set as follows: initial denaturation at 95°C for 7 min, followed by 45 cycles of denaturation at 95°C for 10 sec, primer annealing and extension at 60°C for 15 sec, elongation at 72°C for 15 sec. The melting curve conditions were set to an initial heating to 95°C for 10 sec, followed by a cool-down to 60°C for 1 sec. Subsequently, the temperature was increased to 95°C at a ramp rate of 0.11°C/sec. The program terminated at 95°C for a final hold of 20 sec. All reactions were run at the LightCycler480-II (Roche, Basel, Switzerland). All measurement were performed in technical triplicates.

3.6.4.2 TaqMan

Each reaction constituted of 1.8 µL water (ROTISOLV® HPLC Gradient Grade, Roth, Karlsruhe, Germany), 5 µL SensiFast Probe No-Rox Kit (Biocat, Heidelberg, Germany), 0.5 µL of the forward and the reverse primer (10 µM) and 0.2 µL of the hydrolysis probe. As template, 2 µL cDNA was used. The PCR conditions were set as follows: initial denaturation at 95°C for 2 min, followed by 55 cycles of denaturation at 95°C for 10 sec and primer annealing and extension at 60°C for 30 sec. All reactions were run at the LightCycler480-II (Roche, Basel, Switzerland) in

technical triplicates. The relative fold change in gene expression was calculated based on the delta-delta-Ct (ΔΔCt) method (Livak & Schmittgen, 2001). The average Ct across the two housekeeper genes β-Actin and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was subtracted from the Ct of each individual gene of interest (Δ-Ct). The Δ-Ct of the mock-transfected control was further substracted from the Δ-Ct of the different transfection conditions (ΔΔCt). Using the formula $2^{-\Delta\Delta Ct}$, the fold gene expression of the mock-transfected control was set equal 1 whereas the gene expression of all other conditions was expressed as the fold change of the pcDNA3.1_FLAG (no insert) transfected control.

3.6.5 RNAseq

The PCR reaction for RNA-Seq amplicon preparation was composed in a total volume of 20 µL as follows: 2 µL DNA template (cDNA 5 ng/µL), 1.0 µL of the forward and the reverse primer with adaptor sequences (10 µM, see Table 6), 4 µL of 5x SuperFi Buffer (Thermo Fisher Scientific, Waltham, USA), 11.2 µL of water (ROTISOLVE® HPLC Gradient Grade, Roth, Karlsruhe, Germany), 0.4 µL of dNTPs (10 mM, dNTP-Set1, Catalogue No. K039.1/2, Roth, Karlsruhe, Germany) and 0.4 µL of Platinum™ SuperFi II DNA Polymerase (Thermo Fisher Scientific, Waltham, USA). The PCR conditions were set to an initial denaturation at 98°C for 30 sec, followed by 35 cycles of denaturation at 98°C for 10 sec, primer annealing at 56°C for 10 sec and primer extension at 72°C for 30 sec, terminating with a final elongation step at 72°C for 5 min.

3.7 Affinity chromatography (AC)

In a first step, the biotinylated ds-oligonucleotides were bound to streptavidin coupled magnetic beads. The Dynabeads® MyOne™ Streptavidin C1 (Thermo Fisher Scientific, Waltham, USA) were vigorously resuspended in the original package of which 50 µL were transferred into a DNA- LoBind 1.5 mL collection tube (Eppendorf, Hamburg, Germany). The tube was placed onto the DynaMag™-2 magnet stand (Thermo Fisher Scientific, Waltham, USA) for 1 min to separate the beads from the original medium. The clear supernatant was carefully removed and 100 µL of 1x bind and wash buffer (1x BW: 1 M NaCl, 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA) was added. The collection tube was removed from the magnet and the

beads were well resuspended before placing the tube back onto the magnet for 1 min. After removing the supernatant, the washing step was repeated. The beads were dissolved in 105 µL of 2x BW (2 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA) and for each allele, 50 µL were transferred to a fresh DNA LoBind collection tube (Eppendorf, Hamburg, Germany). In a next step, 50 µL of the biotinylated ds-oligonucleotides (8.2 ng/µL) including either the non-risk or risk allele were added to the beads and incubated for 1 h at RT on a roller bank. The tubes were covered in aluminum foil to protected them from light and direct heat. For overnight incubation, the tubes were placed horizontally onto a rotating overhead shaker at 4°C. The next day, the beads were washed in 100 µL of 1x BW buffer. In a second step, beads with un-bound streptavidin binding sites were blocked in 100 µL of 1x BW containing 2 ng/µL biotin during a 1 h incubation at RT on a roller bank. In the meantime, the reaction mixture including the NE (RX) was prepared. Working on ice, 600 µL of 5x BB, 5.4 µL of 2.5% 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS, final 0.0045%) and 2 mg of the NE (*i.e.,* HepG2: 454.5 µL, stock 4.6 µg/µL; Caco2: 512µL, stock 3.9 µg/µL) were combined in a total volume of 3 mL. Please note, the RX is prepared for both alleles, leaving 1 mg of NE for each allele. To save material, only 0.5 mg NE were used per allele during the test runs. After 1 h of blocking, all consecutive steps were carried out at 4°C (cold room) and all working material including plastic ware, pipettes and buffers were pre-cooled before the start of the experiment. In a first step, the beads were washed in 200 µL washing buffer (WB: 1x BB (4% glycerol, 1 mM $MgCl_2$, 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.5 mM DTT), 50 mM NaCl). In a second step, the beads were washed in 200 µL of 1x BB. The supernatant was removed thoroughly before applying 725 µL of the RX twice, dissolving the beads well in between. For the binding reaction of nuclear proteins to the ds-oligonucleotides, the reaction mixture was incubated at 4°C on an overhead shaker for 20 min. To reduced unspecific protein DNA interaction and to highlight SNP specific protein binding, polydIdC was added to the reaction mixture to compete for protein binding with the ds-oligonucleotides. To do so, the polydIdC amount was adjusted to 10% of the amount established in the EMSA experiments (*i.e*., 23.3 µg polydIdC (700 ng/µL stock) per 1 mg NE). The reaction was incubated shaking for another 10 min at 4°C before the reaction tubes were placed onto the magnet (10 min). The clear supernatant (SN-cplx), including all unbound proteins, was removed carefully and transferred to a

collection tube. To wash off further unbound or weakly associated proteins, the beads were resuspended in 200 µL of WB, incubated for 1 min off the magnet and for 1 min on the magnet. The clear supernatant was transferred to a new collection tube. This washing step was repeated for a total of three times, pooling all supernatants for one condition (W1). Subsequently, the stepwise elution of proteins through increasing salt concentration was performed. The beads were well dissolved in 100 µL elution buffer (EB) and incubated for 2 min off the magnet and 1 min on the magnet before the clear supernatant, containing the eluted proteins, was collected in a new collection tube. During the test run, consecutive elution was performed with increasing salt concentrations (NaCl) of 200 mM, 300 mM, 400 mM, 500 mM, 600 mM, 800 mM and 1,000 mM. After careful evaluation of the pre-experiment, elution was performed at one low salt condition (400 mM) followed by one elution at a high salt condition (800 mM) in the actual AC experiment. Please note, that the described procedure applies to one reaction of one allele (compare *Figure 3*). However, all experiments were conducted in four technical replicates for both alleles in parallel on one day with the same RX mixture. Please note further, that no centrifugation steps can be carried out through the whole AC procedure. The mixtures should therefore be always handled carefully avoiding disturbances or placing the tube upside down.

The collected eluents were shock frozen in liquid nitrogen and sent to our cooperation partner Dr. Christine von Törne (Helmholtz Zentrum München, German Research Center for Environmental Health, Research Unit Protein Science) for mass spectrometry (MS) analysis. A description of the sample processing and data evaluation is provided in the appendix.

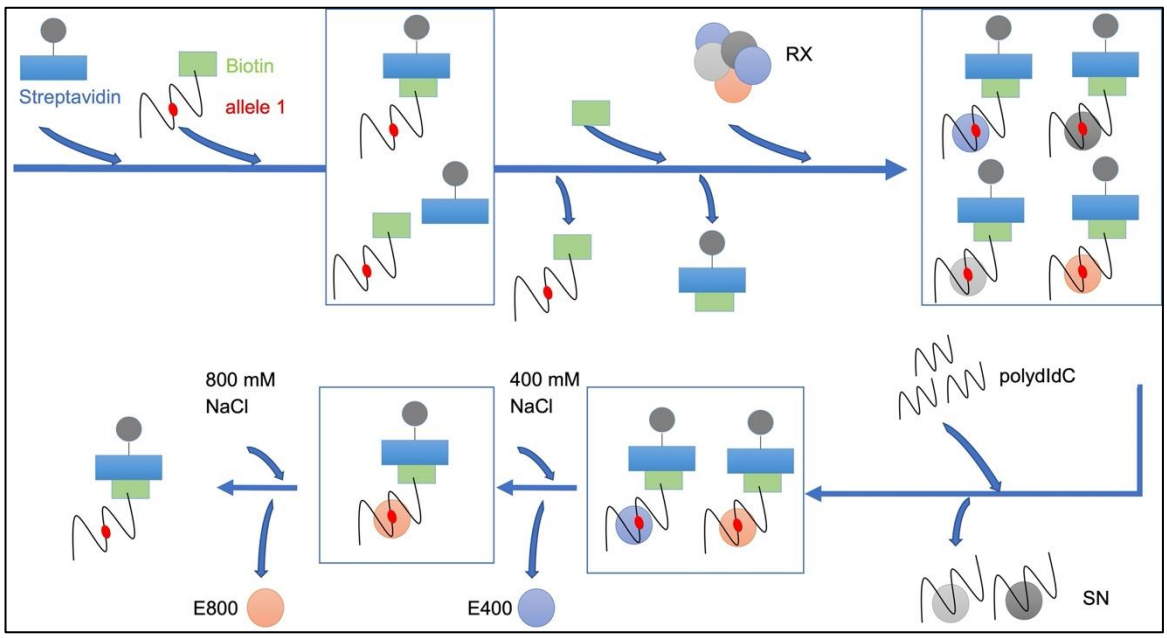


Figure 3 Affinity chromatography (AC) workflow.

The illustration depicts the workflow for one of each two alleles per SNP. Biotinylated oligonucleotides harboring the allele of interest in mid position of a 45bp DNA fragment are bound to magnetic beads via streptavidin during an overnight incubation. Unbound oligonucleotides are washed off and residual streptavidin sides are blocked by the addition of biotin. Nuclear protein (RX) from cell culture extracts binds to the bead-coupled synthetic oligonucleotides. A mixture of unspecific DNA fragments (polydIdC) is added as a competitor to bind weak and unspecific DNA binding proteins (SN). Increasing salt concentrations (NaCl) are used to disrupt the protein-DNA interaction based on the strength of binding affinity. While the oligonucleotides remain attached to the bead, the proteins are eluted from the DNA region of interest. Finally, the protein fractions are subjected to mass spectrometry analysis (E400, E800) to identify the differentially binding transcription factors.

3.8 Data analysis proteomics

We obtained data from proteomics analysis of *rs148276928* (c.1-2090T>C), *rs142703147C>A* (c.1-4141G>T) and *rs148911734* (c.1-7321C>T) and further processed the results in R (V. 4.0.5) for candidate protein prioritization. We removed all proteins, which were not identified by mass spectrometry (MS) in at least three replicates and filtered for candidate proteins passing a false discovery rate (FDR) adjusted p-value threshold of 0.01 for differential binding in both investigated cell lines. Next, we annotated and filtered for transcription factor and cofactor annotation from the HumanTFDB database (Hu *et al*., 2019) to obtain potential regulatory proteins. For transcription factors we annotated SNP transcription factor binding site (SNP-TFBS)-disruption information based on the hg19 reference genome (Team TBD, 2020) and dbSNP Build 142 (Pagès *et al*., 2015) with the *motifbreakR* function from the *motifbreakR* R-package (Coetze *et al*., 2015), if a predicted TFBS binding

p-value was below 0.05 (filterp = TRUE & threshold = 0.05). We applied this function for all human TFBS entries (prefiltered for organism = “Hsapiens”) from the *MotifDB* R package (Shannon *et al*., 2021). This subset encompasses a large universe of TFBS from different datasets and databases (Jolma *et al*., 2013, Weirauch *et al*., 2014, Kulakovskiy *et al*., 2016, Kulakovskiy *et al*., 2018, Xie *et al*., 2009**,** Mathelier *et al*., 2014**,** Mathelier *et al*., 2016, Khan *et al*., 2018, Neph *et al*., 2012**,** Pachkov *et al*., 2013, Hume *et al*., 2015, Scharer *et al*., 2009, Berger *et al*., 2006). Some of the integrated databases represent different versions of the same source (e.g., different versions of the Jaspar database). To prevent redundant annotations, we removed TFBS, if they were also present in a newer version of the same database. Furthermore, if for a certain TF-SNP scenario multiple TFBS were available, we kept only the matrix leading to the strongest difference in binding (by absolute difference between the scaled binding scores at the reference (pctRef) and alternative (pctAlt) sequence). In case of a tie, we gave preference to Jaspar and Hocomoco TFBS, as we considered Jaspar and Hocomoco to be the most prominent databases in the field. We plotted SNP-TFBS-Disruptions if they were predicted to be “strong” by motifbreakR with the plotMB function and considered this information for further protein prioritization. The bioinformatic analysis was performed by Andreas Schmidt.

3.9 *SPINK1* p.N34S haplotype frequency estimation

We obtained genotype data from 253 cases and 256 controls for *rs17107315T>C* (p.N34S), *rs142703147C>A* and *rs148911734C>T* from Dr. Giriraj Ratan Chandak, Scientist and Group Leader at the CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India. Genotyping was based on Sanger sequencing. We estimated haplotype frequencies with the *haplo.em* function from the *haplo.stats* (V. 1.8.7) R-package (https://cran.r-project.org/web/packages/haplo.stats/index.html) using default parameters. Furthermore, we used the *haplo.score* function, which applies a score test (Schaid *et al*., 2002) to estimate haplotype-disease associations. When using the *haplo.score* function, we set the *trait.type* argument to "binomial", to account for thecase-control scenario. The bioinformatic analysis was performed by Andreas Schmidt.

4 Results

4.1 Preliminary work

4.1.1 *SPINK1* expression in p.N34S heterozygous carriers

We extracted genomic DNA (gDNA) from whole human blood samples of 638 Swiss individuals (PD Dr. med. Kaspar Truninger, Basel) and determined their p.N34S genotype by melting curve analysis. We identified five heterozygous carriers and requested RNA from intestinal biopsies of those individuals together with five samples of wild type as controls. Total RNA was reverse transcribed with *SPINK1*-specific primers and either sent for Sanger sequencing at Eurofins Genomics or analyzed with RNA-Seq at the *ZIEL Institute for Food & Health, Core Facility Mikrobiom,* PD Dr. Klaus Neuhaus.

Electropherograms of intestinal mRNA samples of *SPINK1* p.N34S heterozygous carriers show a decreased expression of the risk allele. mRNA and gDNA samples of heterozygous and wild type subjects were Sanger sequenced with two different reverse primers located downstream of p.N34S (c.101 T>C) (Figure 2). The electropherograms of the gDNA samples confirmed the previously determined genotype and showed a single peak (TT) for individuals with the wild type and a double peak (TC) for heterozygotes. As expected, the heterozygous gDNA samples showed an equally high peak for the non-risk (T) and risk allele (C). However, in the corresponding RNA samples, the heterozygotes showed a weaker signal for the risk allele (C) in 7 out of 10 electropherograms. The electropherograms of subjects with the wild type, in turn, showed the same pattern in the gDNA as in the RNA sample. We thus obtained a first and valuable hint, that variants of the p.N34S haplotype in the regulatory region of the gene might influence the expression of *SPINK1* in an allele-dependent manner.

Unfortunately, the results of the RNA-Sequencing at *ZIEL Institute for Food & Health, Core Facility Mikrobiom,* PD Dr. Klaus Neuhaus were not available upon the finalization of this thesis. Due to material shortage in connection with the Corona pandemic the analysis is still pending, and results are expected by the end of 2021.

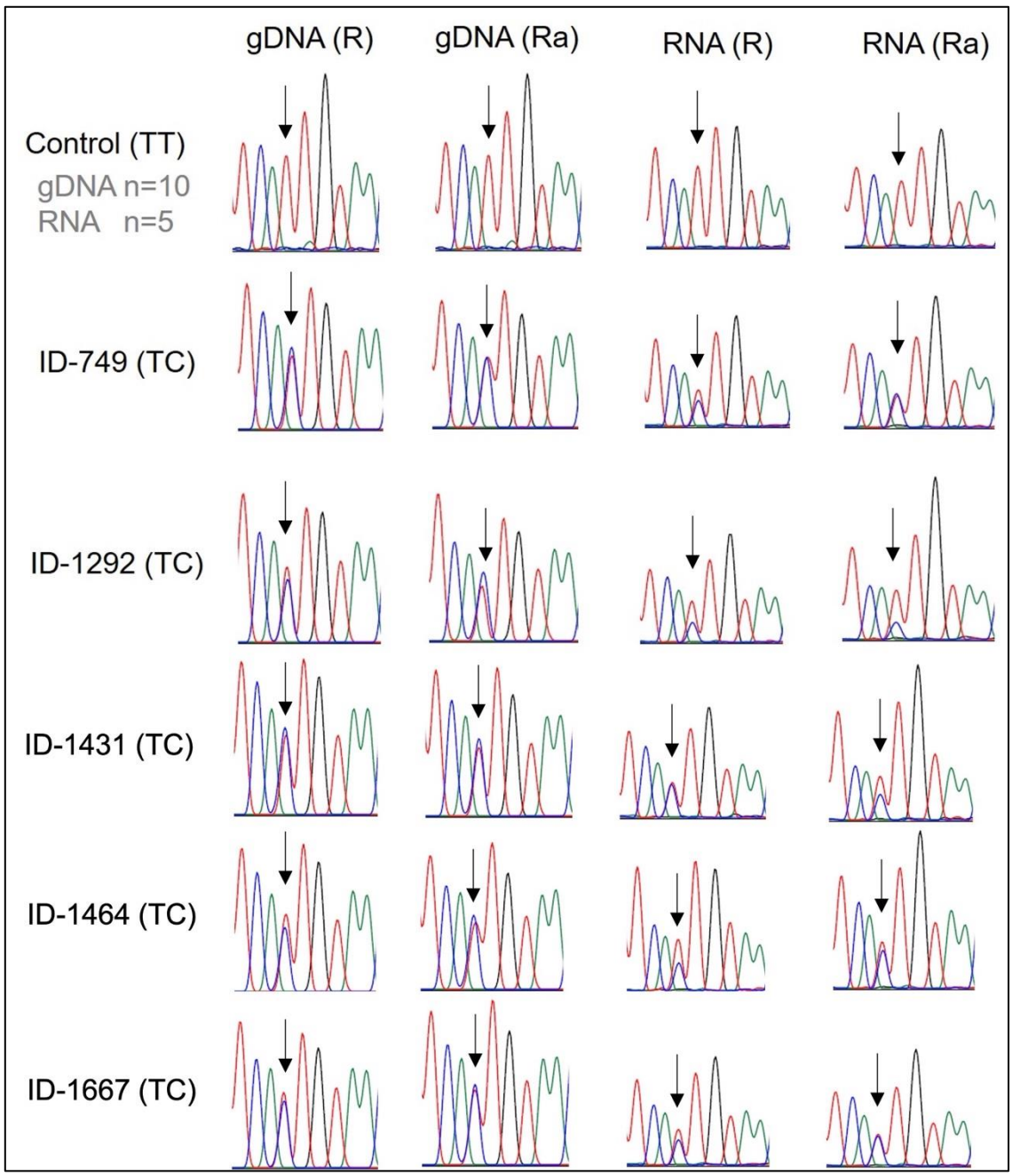


Figure 4 Electropherograms of gDNA and intestinal mRNA samples of SPINK1 p.N34S heterozygotes.

Genomic DNA (gDNA) and RNA of intestinal biopsies of heterozygotes (TC) and wild type (TT) were Sanger sequenced with two different reverse primers (SPINK1-N34S-Seq-Exon3R (R) and SPINK1-N34S-Seq-Exon3Ra, (Ra)) located downstream of p.N34S (c.101 T>C). The reverse complementary sequence is shown. For the homozygous wild type controls, representative electropherograms of gDNA (n=10) and RNA (n=5) are shown. For the p.N34S heterozygous carriers (n=5), the electropherograms are listed by the individual (ID). Color code of electropherogram: red = thymine, blue = cytosine, green = adenine, black = guanine.

4.1.2 Pre-selection of *SPINK1* p.N34S linked regulatory variants

In silico and experimental work selected *rs17107316*, *rs148276928*, *rs142703147* and *rs148911734* as regulatory variants of the *SPINK1* p.N34S haplotype for further analysis. Variants in high linkage disequilibrium (LD) to p.N34S were identified by Dr. Helmut Laumen by *SNiPa* (https://snipa.helmholtz-muenchen.de/snipa3/; 1000 Genomes Project Consortium, Abecasis *et al*., 2012). LD acquisition first focused on variants in the European population using the 1,000 Genomes data set (r^2 > 0.8, 1,000 Genomes Phase3v5, EUR). Under those settings, 26 variants were identified. Additionally, sequencing data from German pancreatitis patients was cross-referenced with the obtained data. Phylogenetic Module Complexity Analysis (PMCA) narrowed the selection down to 10 variants. PMCA is a bioinformatics approach to detect genetic regions with evolutionary conserved transcription factor binding site modules. This approach identifies *cis*-regulatory genomic regions, which could potentially influence the regulation of gene expression and thereby disease development. The method is described in detail by Claussnitzer *et al*. (2014). Furthermore, we assessed the overlap of the candidate variants to recently published human pancreas specific epigenomic marks of regulatory regions (Roadmap Epigenomics Consortium, Kundaje *et al*., 2015) and confirmed six of the PMCA identified variants. LD acquisition in the South Asian population (r^2 > 0.8, 1,000 Genomes Phase3v5, SAS) identified five different variants selected from the European population. The data is summarized in Table 27 below.

Table 27 In silico selected SPINK1 p.N34S linked variants

Variant	Position Chr5 (GRCh37)	Minor allele	r^2 European	r^2 South Asian
rs77658106	147189421	T	0.8	0.14
rs149882377	**147204875**	**T**	**1.0**	**0.86**
rs17107316	**147208894**	**C**	**1.0**	**1.0**
rs148276928	**147213230**	**C**	**1.0**	**0.93**
rs142703147	147215281	A	1.0	0.57
rs148911734	**147218371**	**T**	**1.0**	**0.86**
rs142684029	**147219066**	**A**	**1.0**	**0.86**
rs142120065	147224635	A	0.87	0.57

Variants in high LD to *SPINK1* p.N34S (*rs17107315)* in European (r^2 > 0.8, 1,000 Genomes Phase3v5, EUR) and South Asian (1,000 Genomes Phase3v5, SAS) super-populations in **bold**; variants identified by both, bioinformatics PMCA and epigenomic marks are highlighted in gray.

DNA sequencing of Indian and German patients and controls confirmed *rs148911734*, *rs148276928* and *rs17107316*. *rs142703147* was included into further investigations due to a recent publication of Boulling *et al* (2017). Functional analyses by reporter gene assay and EMSA further strengthened the selection of those variants. While *rs142703147* and *rs17107316* led to an increase in reporter gene activity under the influence of the non-risk allele, *rs148276928* showed increased promoter activity at the risk allele. In EMSA experiments, we observed an increased nuclear protein binding at the non-risk allele of *rs148276928* and the risk allele of *rs142703147* and *rs148911734.* In addition, for *rs148911734* we also found a reduced binding at the risk allele for another, lower running band. An overview of the variant selection process is provided in Figure 5. Please note, that the *in silico* findings resulted from the work of Prof. Dr. Heiko Witt and Dr. Helmut Laumen. The experimental work was conducted by Christina Schwalm (Sanger sequencing) and Lena Hoerbuger (reporter gene assay, initial EMSA experiments) as part of their Bachelor and Master thesis, respectively.

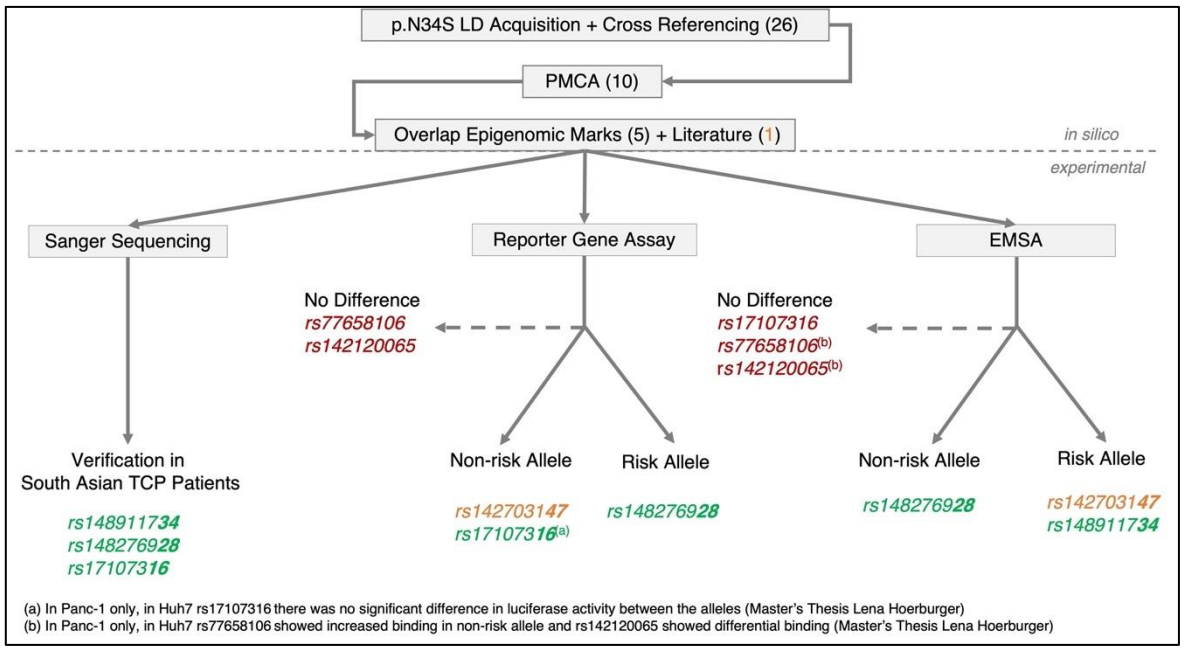


Figure 5 Summary of the preliminary work and selection of the candidate SNP of the SPINK1 p.N34S haplotype under investigation in this thesis.

The *SPINK1* p.N34S variant is in high LD with 26 SNPs (European population, r^2=0.8, 1,000 Genomes Phase3v5). The numbers in brackets indicate the number of variants obtained after each in silico analysis step. Variants, which were excluded from further analysis due

to experimental evidence are depicted in red. Variants highlighted in green were experimentally verified for further in-depth analysis in this thesis. *rs142703147* (orange) was included due to a recent publication of Boulling *et al.* (2017). SNP: Single Nucleotide Polymorphisms LD: Linkage Disequilibrium; PMCA: Phylogenetic Module Complexity Analysis; EMSA: Electron Mobility Shift Assay; TCP: Tropical Calcifying Pancreatitis

4.1.3 Haplotype and haplotype frequency estimation

We estimated the p.N34S haplotype frequencies for *rs17107315* (T>C, p.N34S), *rs142703147C>A* and *rs148911734C>T* of 253 cases and 256 controls of Indian ancestry. The most frequent haplotype was TCC (all non-risk allele, estimated frequency in cases 0.80, estimated frequency in controls 0.98). This haplotype was significantly associated with the control group (p-value 5.65×10^{-18}; score -8.64). The second most frequent haplotype was CAT (all risk-alle, estimated frequency in cases 0.15, estimated frequency in controls 0.01). This haplotype was significantly associated with the cases (p-value 9.21×10^{-14}; score 7.45). The third most frequent haplotype was CCT (risk allele: *rs17107315*, *rs148911734* non-risk: *rs142703147,* estimated frequency in cases 0.02, estimated frequency in controls 0.005). This haplotype was also significantly associated with the cases (p-value 2.81×10^{-2}; score 2.19)

The two haplotypes CAT (all risk alleles) and CCT (risk allele: *rs17107315*, *rs148911734)* were thus significantly associated with chronic pancreatitis status. Haplotype blocks exclusively containing the non-risk allele (TCC) clearly associated with the control status. Table 28 provides an overview of the estimated haplotype frequencies in cases and controls. Table 29 summarized the p-value and score of the association statistics.

Please note, the very rare haplotypes TAC, CCC, CAC and TAT were not included in association analysis, as they did not pass the *haplo.score default min.count* parameter of 5, which removes haplotypes based on the number of minimum expected counts in the sample. According to the *haplo.stats* manual (V. 1.7.7), including haplotypes with lower expected counts could result in unstable variance estimates and unreliable test statistics.

Table 28 Estimated haplotype frequencies

rs17107315	*rs142703147*	*rs148911734*	Frequency
Cases			
T	C	C	0.804
C	A	T	0.148
C	C	T	0.022
T	C	T	0.012
T	A	C	0.004
C	C	C	0.004
C	A	C	0.004
T	A	T	0.002
Controls			
T	C	C	0.984
C	A	T	0.010
C	C	T	0.006

Application of *haplo.em function*, default parameter. The analysis was performed by Andreas Schmidt

Table 29 Haplotype association statistics

rs17107315	*rs142703147*	*rs148911734*	p-value	Score	Frequency
T	C	C	5.65E-18	-8.64E+00	8.95E-01
C	A	T	9.21E-14	7.45E+00	7.86E-02
T	C	T	1.30E-02	2.48E+00	5.92E-03
C	C	T	2.81E-02	2.20E+00	1.38E-02

Application of *haplo.score function*. Negative scores indicate association with controls, positive scores indicate association with cases. The analysis was performed by Andreas Schmidt

4.1.4 Pre-selection of *SPINK1* expressing cell model

For the investigation of *SPINK1* regulatory variants of the p.N34S haplotype, we wanted to choose a cell model that had sufficient *SPINK1* expression both at the mRNA and at the protein level. The underlying rationale was as follows: (1) We expected cells with an endogenous *SPINK1* expression to possess the relevant regulatory factors which might be affected by the SNPs under investigation. First

insights into SNP-dependent DNA binding would be obtained during EMSA experiments. In a second step, affinity chromatography and subsequent proteomics analysis would identify the relevant factors. (2) Once we identified the transcription factors, subsequent analyses should establish a direct link between the transcription factor and *SPINK1* expression. Since SPINK1 is considered as a protective factor against CP under physiological conditions, we expected that regulatory variants of the CP associated p.N34S haplotype would reduce *SPINK1* expression. Thus, solid baseline levels of *SPINK1* mRNA and protein would facilitate the demonstration of a decreased expression *e.g*., during knock down experiments of the relevant candidates.

Caco2 and HepG2 cell qualified as study model for the functional analysis of the selected SNP. We collected mRNA from nine different cell lines of intestinal (Caco2, HT29), hepatic (HepG2, Huh7), pancreatic (Panc1, Paca44, COLO357) and renal (HEK293) origin. We compared *SPINK1* expression levels of the cell lines to those of the *SPINK1* expression in the human pancreas. Independent of the cell confluency, Caco2, HT29, HepG2 and Huh7 showed the highest expression levels, which even exceeded the expression levels of the corresponding human tissue (Figure 6A). *SPINK1* expression of the remaining cell lines fell far below the human expression levels, which is why these cell lines were excluded at this step. In a next filtering step, we evaluated the SPINK1 protein abundance by ELISA in cell culture conditioned media (supernatant) and total cell lysate after culturing the cells for 48 h, 72 h and 120 h in a dividing (1x) or confluent (2x) state (Figure 6B). Despite high mRNA expression levels in HT29, we excluded this cell line from the experiments due to a mycoplasma infection. Instead, we included the renal cell line RT4, which was listed with the highest *SPINK1* mRNA expression on https://www.proteinatlas.org/. SPINK1 abundance increased along with the duration of culturing and the cell density for Caco2, Huh7 and RT4. For HepG2 only, the confluent culture (2x) appeared to accumulate less SPINK1 than the dividing culture. The amount of SPINK1 in the total lysate reflected the findings from the supernatant. Overall, Huh7 cells displayed by far the highest levels, followed by Caco2, RT4 and HepG2. Based on these results, all four cell lines were included in the subsequent EMSA experiments.

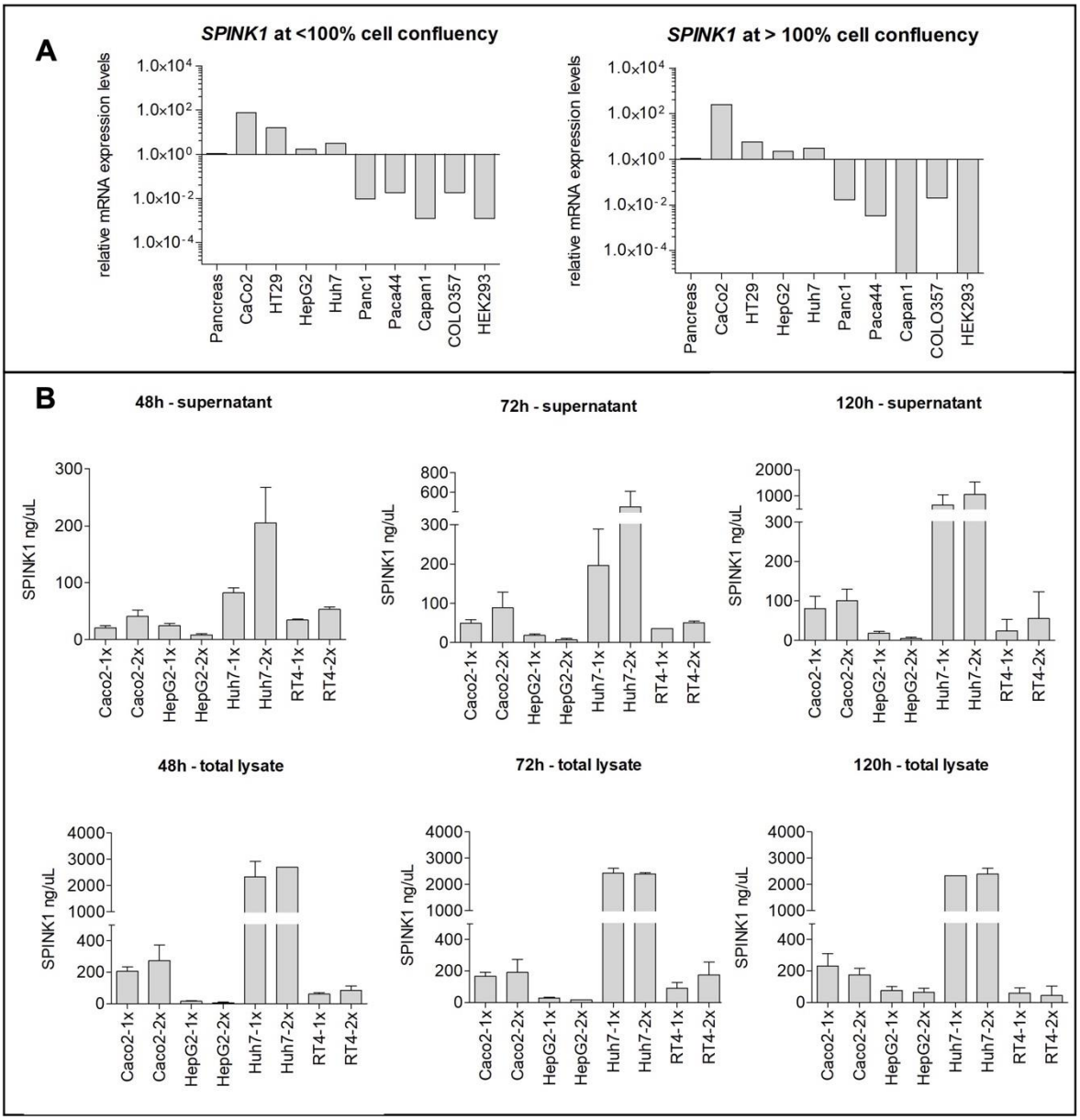


Figure 6 SPINK1 expression and protein abundance in human cell lines.

(A) Total RNA was extracted from cells below (<) or above (>) full confluency (100%) and SPINK1 mRNA expression levels were determined by qPCR (SYBR Green) in relation to SPINK1 expression levels in the human pancreas (baseline), n=1. (B) SPINK1 protein abundance was determined by ELSIA in supernatant and total lysate after 48 h, 72 h and 120 h incubation of cells in the dividing (1x) or confluent (2x) state. Measurements were performed in technical and biological replicates (n=2), standard deviation is shown. qPCR: quantitative-Real-Time-PCR, ELSIA: Enzyme-linked Immunosorbent Assay. SPINK1: serine protease inhibitor Kazal type 1.

4.1.5 Identification of differential transcription factor binding

Nuclear protein of Caco2 and HepG2 cells showed differential binding patterns at the risk and non-risk allele of *rs148276928* and *rs148911734*. To sequester further the most promising SNP and most suitable cell model for affinity chromatography experiments, we conducted EMSAs. We exposed nuclear extracts

of Caco2, HepG2, Huh7 and RT4 cells to Cy5-labeled synthetic oligonucleotides harboring either the non-risk or risk allele of *rs17107316*, *rs148276928*, *rs142703147* or *rs148911734* at mid-position (Figure 7). We found almost no protein binding at *rs17107316* for nuclear proteins in all cell lines. Therefore, we excluded this SNP from further analysis. For *rs142703147*, nuclear protein binding was observed for all cell lines. The binding pattern differed between the cell lines, however, no difference in protein binding depending on the non-risk or risk allele was detectable. Consequently, *rs142703147* disqualified for further analysis. However, the recent publication of Boulling *et al.* (2017) suggested a functional role of this SNP in *SPINK1* expression, which is why we decided to investigate the SNP further. In the case of *rs148276928*, an increased protein binding at the risk allele (C) was consistent in the lower half of the gel across all cell lines. In the upper half of the gel, nuclear protein binding of Caco2 extracts was increased at the non-risk allele (T). We observed this effect to a lesser extent for nuclear proteins of HepG2 and RT4, while the binding pattern in Huh7 was very faint. Similarly, bands from Huh7 nuclear extract were comparatively weak in EMSA with *rs148911734*. A binding pattern at mid height of the gel was again consistent through all cell lines and showed increased binding to the non-risk allele (C). A second pattern slightly above showed a stronger signal for the risk allele (T). This effect was best visible in binding reactions with Caco2 and HepG2. The same was true for a third and fourth binding pattern at the top of the gel. Conclusively, we selected *rs148276928* and *rs148911734* as candidate SNPs due to the differential binding pattern at the non-risk and risk allele. This pattern was most prominent in nuclear extracts from Caco2 and HepG2. We therefore selected Caco2 und HepG2 cells as cell models for the identification of differentially binding transcription factors.

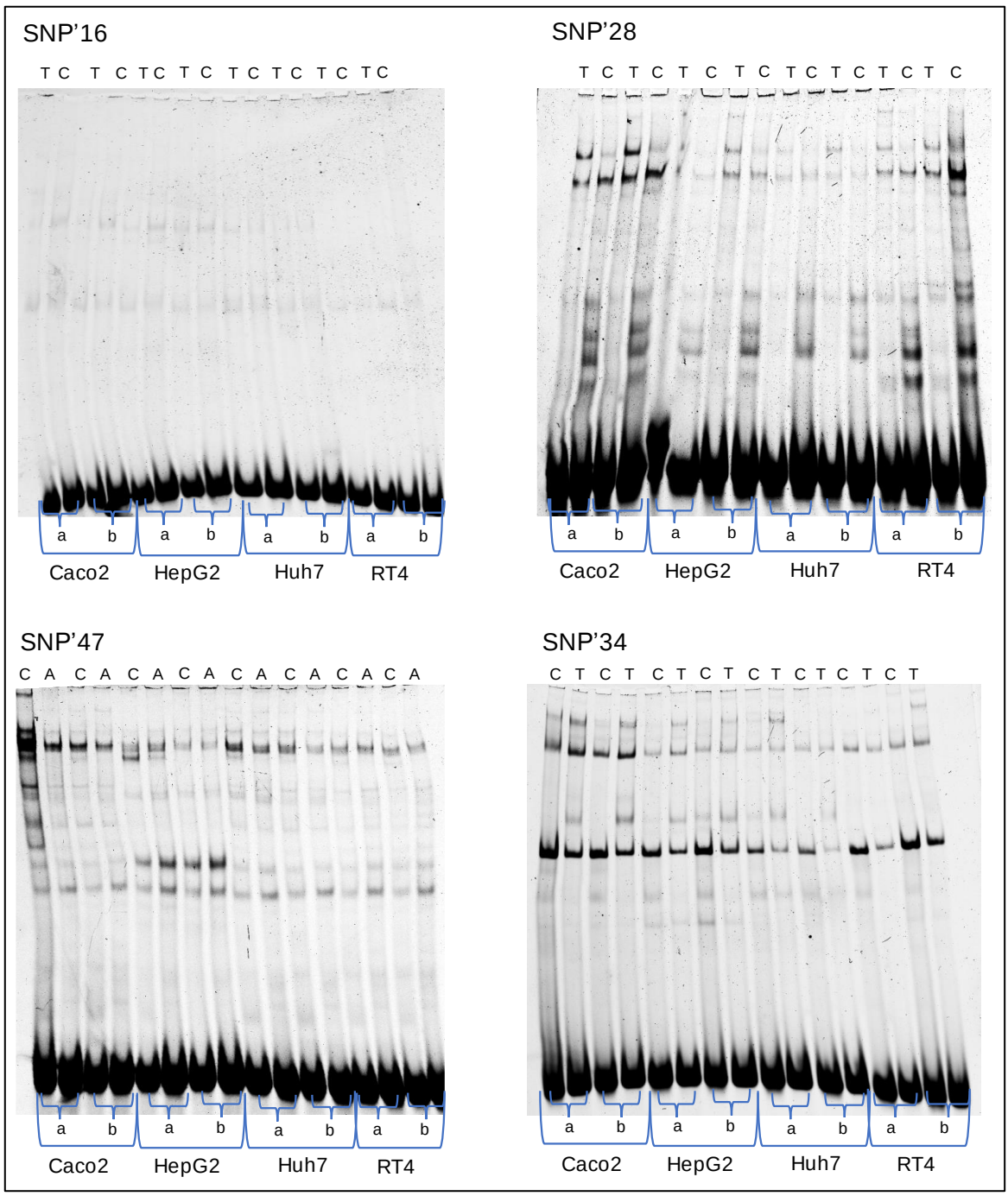


Figure 7 Nuclear protein binding of selected cell lines to the four candidate SNP.

EMSA was performed with nuclear protein from the intestinal cell line Caco2, the two liver derived cell lines HepG2 and Huh7 and the renal cell line RT4. Nuclear protein was used from two independently prepared nuclear extracts (a, b). The binding reaction with the non-risk allele was applied first, followed by the risk allele. Nuclear extract: 3 µg, polydIdC: 350 ng, Cy5-oligonucleotide: 1 ng. EMSA: Electron Mobility Shift Assay; SNP'16: rs17107316 T>C; SNP'28: rs148276928 T>C; SNP'47 *rs142703147* C>A; SNP'34: rs148911734C>T.

Table 30 Proteomics data of rs148911734 after filtering

Gene	Cell line	Unique Peptides	Fold change	P-value	Padj	TF	CoF	Strand	Data Source	pctRef	pctAlt	TFBS disruption
ZNF658	**CaCo2**	**1**	**3.16**	**1.00E-17**	**4.42E-16**	**TRUE**	**FALSE**	**NA**	**NA**	**NA**	**NA**	**NA**
YY1	**CaCo2**	**16**	**3.02**	**1.00E-17**	**4.42E-16**	**TRUE**	**FALSE**	**+**	**jaspar2018**	**0.78**	**0.53**	**strong**
ZNF714	CaCo2	1	2.9	1.00E-17	4.42E-16	TRUE	FALSE	NA	NA	NA	NA	NA
YY1	**HepG2**	**18**	**2.66**	**1.00E-17**	**5.01E-16**	**TRUE**	**FALSE**	**+**	**jaspar2018**	0.78	0.53	**strong**
ZNF714	HepG2	1	2.47	1.00E-17	5.01E-16	TRUE	FALSE	NA	NA	NA	NA	NA
ZNF658	**HepG2**	**1**	**1.95**	**1.00E-17**	**5.01E-16**	**TRUE**	**FALSE**	**NA**	**NA**	NA	NA	**NA**
NFRKB	CaCo2	22	1.25	8.04E-07	2.44E-05	TRUE	FALSE	NA	NA	NA	NA	NA
WIZ	CaCo2	18	1.19	8.66E-05	0.00181	TRUE	FALSE	NA	NA	NA	NA	NA
WIZ	HepG2	20	1.15	8.99E-05	0.00169	TRUE	FALSE	NA	NA	NA	NA	NA
NFRKB	HepG2	17	1.14	0.000595	0.00916	TRUE	FALSE	NA	NA	NA	NA	NA
POU2F1	HepG2	13	0.77	3.96E-11	1.56E-09	TRUE	FALSE	+	jolma2013	0.77	0.69	weak
POU2F1	CaCo2	12	0.68	5.39E-10	2.14E-08	TRUE	FALSE	+	jolma2013	0.77	0.69	weak
SATB2	**CaCo2**	**2**	**0.51**	**4.27E-06**	**0.000111**	**TRUE**	**FALSE**	**NA**	**NA**	NA	NA	**NA**
GATA6	**HepG2**	**1**	**0.43**	**1.00E-17**	**5.01E-16**	**TRUE**	**FALSE**	**-**	**jaspar2018**	0.70	0.83	**strong**
SATB2	**HepG2**	**8**	**0.33**	**1.00E-17**	**5.01E-16**	**TRUE**	**FALSE**	**NA**	**NA**	**NA**	**NA**	**NA**
GATA6	**CaCo2**	**12**	**0.33**	**1.00E-17**	**4.42E-16**	**TRUE**	**FALSE**	**-**	**jaspar2018**	**0.70**	**0.83**	**strong**

For detailed explanation of the table parameters please refer to the comment below. The analysis was performed by Andreas Schmidt.

Table 31 Proteomics data of rs142703147 after filtering

Gene	Cell line	Unique Peptides	Fold change	P-value	Padj	TF	CoF	Strand	Data Source	pctRef	pctAlt	TFBS disruption
ENO1	HepG2	3	0.86	0,000709	0,00965	FALSE	TRUE	NA	NA	NA	NA	NA
ENO1	CaCo2	3	0.61	2,30E-11	9,39E-10	FALSE	TRUE	NA	NA	NA	NA	NA
TFCP2	HepG2	1	0.59	1,88E-05	0,000344	TRUE	FALSE	-	jaspar2018	0.75	0.85	weak
UBP1	HepG2	1	0.56	7,81E-10	2,26E-08	TRUE	FALSE	NA	NA	NA	NA	NA
TFCP2	CaCo2	7	0.5	1,00E-17	4,81E-16	TRUE	FALSE	-	jaspar2018	0.75	0.85	weak
UBP1	CaCo2	2	0.48	5,76E-10	2,17E-08	TRUE	FALSE	NA	NA	NA	NA	NA

For detailed explanation of the table parameters please refer to the comment below. The analysis was performed by Andreas Schmidt.

Explanation of the table parameters of Table 30 and Table 31

Proteins were filtered for differential binding (FDR adjusted p-value < 0.01 (Padj)) in both cell lines and for annotation as a transcription factor / cofactor by *HumanDB*. Disruption of the transcription factor binding site (TFBS) by the SNP was predicted with *motifbreakR*. Gene = Gene Symbol, Cell line = origin of nuclear extract for affinity chromatography, Unique Peptides = Number of unique peptides to identify the protein, FC = Proteomics fold-change of the identified protein between the non-risk and risk allele, TF = transcription factor annotation by *HumanDB*, CoF = cofactor annotation by *HumanDB,* Strand = strand annotation of a potential overlapping TFBS predicted by *motifbreakR*, Data source = source database of TFBS motif, pctRef: Scaled binding scores [0,1] for the match between a TFBS and the reference sequence (non-risk allele). pctAlt: Scaled binding scores [0,1] for the match between a TFBS and the alternative sequence (risk allele). TFBS-disruption = predicted strength of the TFBS-disruption by *motifbreakR*. NA= Not applicable

4.2 Characterization of differential transcription factor binding

4.2.1 Affinity chromatography and proteomics

YY1, ZNF658, GATA6 and SATB2 are promising candidates for an allele dependent *SPINK1* regulation at *rs148911734.* We obtained data from the proteomic analysis of *rs148276928* (c.1-2090T>C), *rs142703147* (c.1-4141G>T) and *rs148911734* (c.1-7321C>T). After stringent filtering, eight candidate regulatory proteins remained for *rs148911734 (*Table 30*)*. For *rs142703147,* three candidate regulatory proteins remained *(Table 31)*. Each of the factors was replicated in both cell lines (HepG2, Caco2). *rs148276928* did not sustain these filtering criteria and no candidate regulatory proteins remained. A transcription factor with an increased binding to the non-risk allele (fold-change (FC) >1) was considered as potential activator while increased binding to the risk allele (FC <1) suggested a potential repressor. Importantly, we detected strong transcription factor binding site (TFBS)-disruption predictions from *motifbreakR*, only for the *rs148911734* candidate regulatory proteins GATA6 and YY1 (Figure 8). The scaled GATA6 TFBS binding score was higher in case of the alternative (risk) allele (pctRef = 0.70; pctAlt = 0.83), whereas the YY1 TFBS matched better to the reference (non-risk) sequence (pctRef = 0.78; pctAlt = 0.53). The directions of these predictions were in line with the differential binding fold changes from affinity chromatography mass spectrometry ACMS, indicating GATA6 (FC HepG2 = 0.43; FC CaCo2 = 0.33) to be a potential repressor and YY1 (FC HepG2 = 2.66; FC CaCo2 = 3.02) to be a potential activator. We decided to consider those two proteins for further experimental validation. Furthermore, we selected ZNF658 (FC HepG2 = 1.95, FC CaCo2 = 3.16) as potential activator and SATB2 (FC HepG2 = 0.33, CaCo2 = 0.51) as potential repressor at this site because of their strong fold changes.

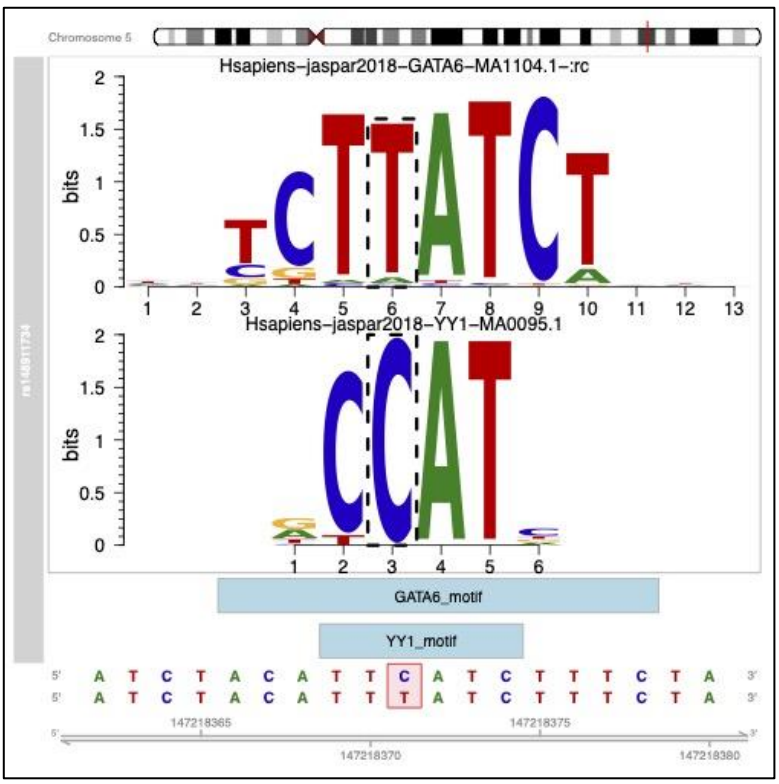


Figure 8 SNP-TFBS-disruption plot for YY1 and GATA6 at rs148911734.

Output from the *motifbreakR plotMB* function for the SNP *rs148911734* after filtering affinity chromatography mass spectrometry (ACMS) results for differentially binding candidate transcription factors of "strong"-predicted *motifbreakR* effect. At the bottom of the plot the genomic sequence (hg19) around the highlighted SNP is shown. The top sequence represents the non-risk allele. The bottom sequence represents the risk allele. The respective disrupted position in a transcription factor binding site (TFBS) motif is highlighted by a dashed frame. The size of each nucleotide correlates with its information content (bits). Blue bars indicate the positions of the sequence-TFBS matches. The analysis was conducted by Andreas Schmidt.

4.2.2 Super-shift EMSA

Super-shift EMSAs confirm an increased binding of YY1 and GATA6/4 from nuclear extracts of Caco2 and HepG2 to either the non-risk (C) or risk (T) allele of *rs148911734 in vitro*. To validate the proteomics findings, we added an antibody against the protein of interest to the mixture of the nuclear proteins and the synthetic oligonucleotide. We included a corresponding isotype control antibody into each run to rule out any unspecific effects through the addition of the antibody itself (Figure 9). In Caco2 cells, the addition of anti-GATA6 or anti-SATB2 antibody clearly led to a shift of the band that showed an increased binding to the risk (T) allele before. The addition of the anti-YY1 antibody shifted the band, which had shown stronger binding to the non-risk allele. For ZNF658, we observed no change in binding patter compared to the isotype control. In HepG2 cells, the anti-GATA6 and anti-SATB2

antibodies did not change the binding pattern. After ruling out technical issues, we reappraised the proteomics data and figured out that GATA4 but not GATA6 might be the relevant factor in the hepatic cell line. After adding an anti-GATA4 antibody to the binding reaction with HepG2 nuclear proteins, we observed the same change in the binding pattern as with Caco2 and the anti-GATA6 antibody. Otherwise, we obtained the same results for YY1 and ZNF658 for the HepG2 cells. A shift of the band, which was stronger for the non-risk allele (C), was apparent after addition of the anti-YY1 antibody, while we did not observe any change in the binding pattern for ZNF658. Taken together, we validated an increased binding of GATA6 (Caco2) and GATA4 (HepG2) to the risk allele and confirmed a stronger binding of YY1 to the non-risk allele in both cell lines. Therefore, we hypothesized that GATA6/GATA4 might suppress *SPINK1* expression while YY1 might enhance it.

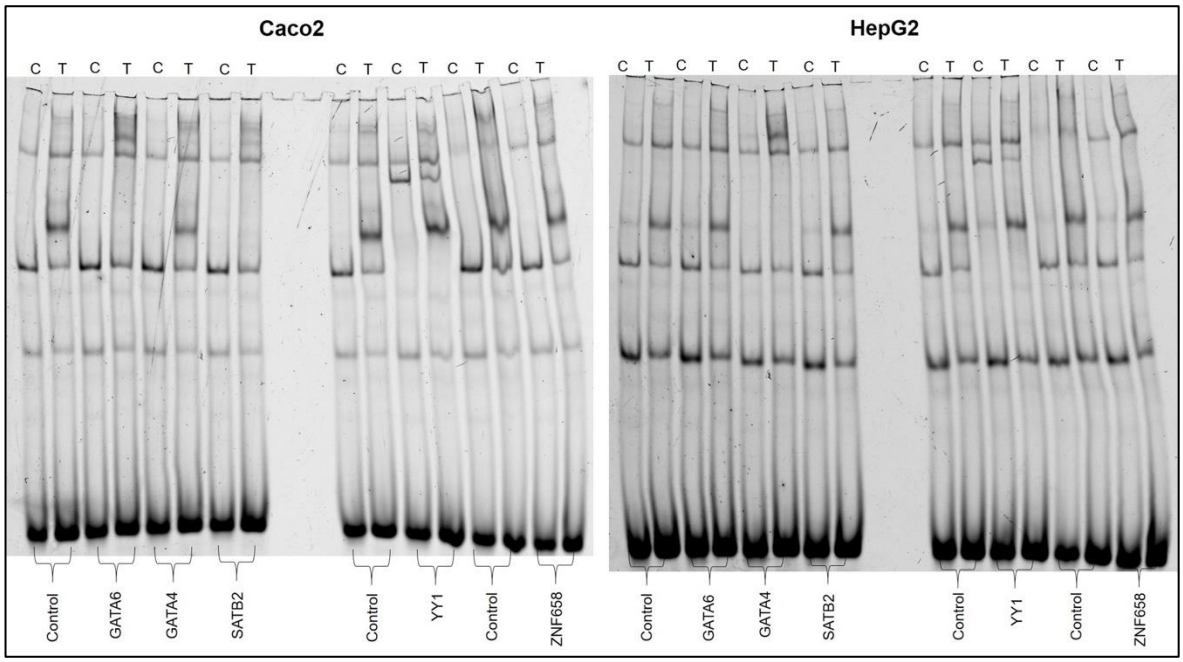


Figure 9 Super-shift EMSA with nuclear protein from Caco2 and HepG2, rs148911734 synthetic oligonucleotides and antibodies against the transcription factors of interest.

Cy5-labelled synthetic oligonucleotides containing *rs148911734* were combined with nuclear proteins from Caco2 and HepG2 cells. Antibodies directed against GATA6, GATA4 and SATB2 (increased binding to the risk allele) or YY1 and ZNF658 (increased binding to the non-risk allele) were added. Unspecific IgG antibodies from the same species were used as control. Nuclear proteins from three independent nuclear extract preparations (n=3) were used of which one representative is shown. The binding reaction with the non-risk (C) allele was applied first, followed by the risk allele (T). Nuclear extract: 3 µg, polydIdC: 350 ng, Cy5-oligonucleotide: 1 ng. EMSA: Electron Mobility Shift As, rs148911734C>T. GATA: GATA binding protein, SATB2: Special AT-rich sequence binding protein 2, YY1: Ying Yang 1, ZNF658: Zinc finger protein 658.

4.2.3 siRNA knock down: effects on mRNA level (qPCR)

siRNA knock down of *GATA6*, *GATA4* and *YY1* do not affect total *SPINK1* expression. After the *in vitro* validation of GATA6, GATA4 and YY1 by super-shift EMSA, we sought to confirm the functional consequences on the *SPINK1* expression *in vivo*. Upon the knock down of the potential repressors, GATA6 and GATA4, we expected an increased *SPINK1* expression, while the knock down of the potential enhancer YY1 should decrease the expression. We reduced *GATA6* expression to 61% and *YY1* expression to 33% in Caco2 cells. The *SPINK1* expression, however, rose to 1.5-fold (*GATA6* knock down.) and 1.4-fold (*YY1* knock down) of the control non-target siRNA (Figure 10 A). We made a similar observation for the knock down experiments in HepG2 cells (Figure 10 B). siRNAs targeting *GATA6*, *GATA4* and *YY1* reduced the corresponding gene expression to 43%, 37% and 27%, respectively. We decided to include a combined knock down of *GATA6* and *GATA4* in the HepG2 cells after we measured the endogenous expression levels of the factors by qPCR (appendix). Even though the expression of *GATA4* was higher, we also observed mRNA expression of *GATA6*. The combined effect of siRNA targeting *GATA6* and *GATA4* at the same time was less effective and reduced *GATA6* expression to only 81% and GATA4 expression to 53%. The *SPINK1* expression, strikingly, also increased to 1.5-fold upon the combined introduction of *GATA6* and *GATA4* siRNA. The *YY1* siRNA led to a 1.6-fold increase of *SPINK1* expression. Single knock down of *GATA6* and *GATA4* left the *SPINK1* expression unchanged compared to the non-target siRNA control. All effects should be interpreted with caution, as the standard deviation in each of the experiments was quite high. Moreover, the following should be of note. First, Caco2 and HepG2 cells both have high endogenous expression of *SPINK1*. This places a hurdle on observing an increase in *SPINK1* expression since the cells might already run at a very high capacity. Second, Caco2 and HepG2 cells both carry the wild type (non-risk) genotype of the *SPINK1* p.N34S variant. However, GATA6 and GATA4 bind stronger to the risk allele. Therefore, they cannot be expected to bind to the region of interest in the same extent as we observed in the *in vitro* experiments. In the case of YY1, no apparent technical explanation lies at hand to why we did not observe a decrease in *SPINK1* expression upon a *YY1* knock down.

Based on later experiments, we also assessed the SPINK1 expression in Paca44 cells after the knock down of *YY1.* Despite a decrease to 11% of YY1 expression, the *SPINK1* expression rose to 1.8-fold of the non-targeting control (Figure 10 C).

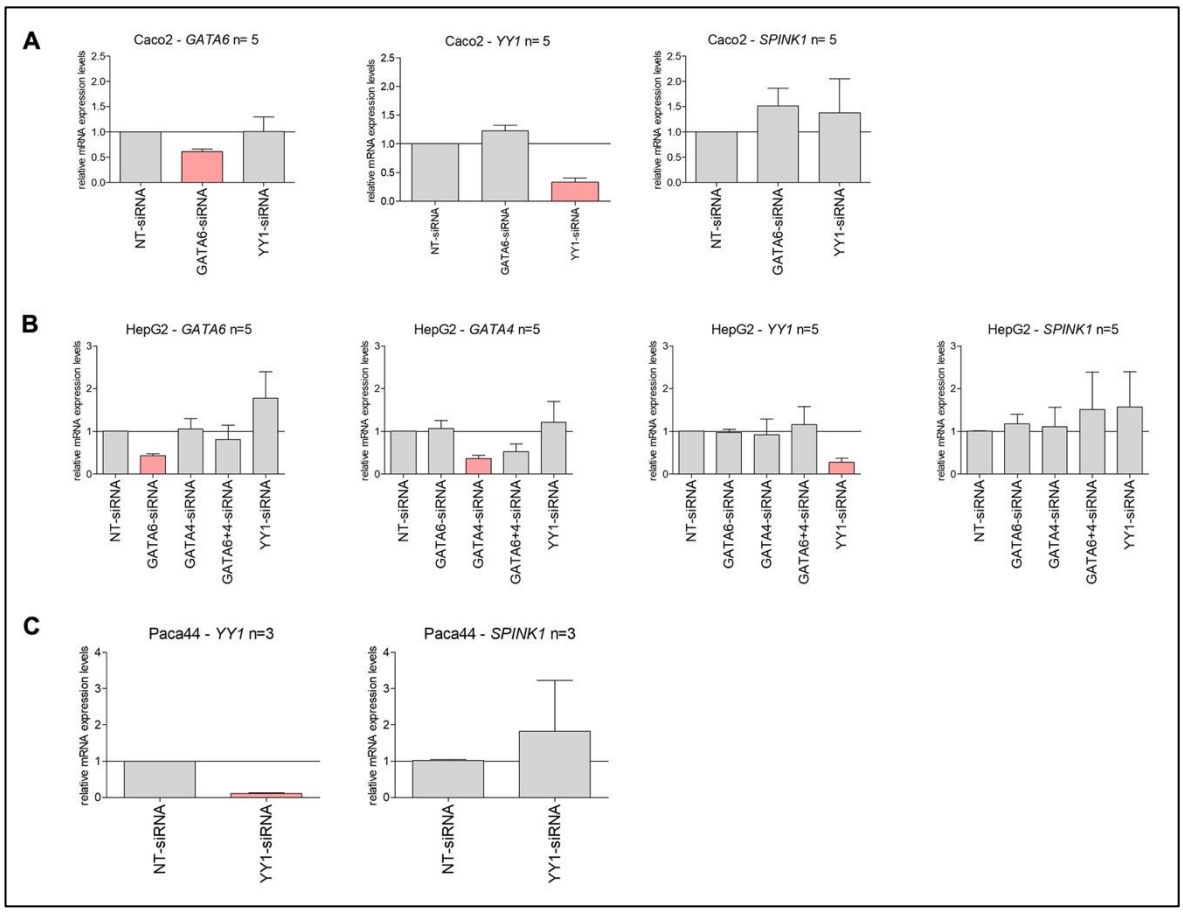


Figure 10 siRNA knock down of GATA6, GATA4 and YY1 in (A) Caco2, (B) HepG2 and (C) Paca44 cells.

ON-TARGET plus siRNAs were transfected in duplicates into Caco2 (n=5), HepG2 (n=5) and Paca44 (n=3) cells. The effect on gene expression was measured by qPCR (TaqMan assays, triplicates). Target gene expression in cell transfected with non-targeting (NT) siRNAs served as baseline control (horizontal line, relative mRNA expression= 1). Bars highlighted in red indicate the knock down effect of the corresponding factor. Standard deviation is shown. GATA: GATA binding protein, YY1: Ying Yang 1, SPINK1: serine protease inhibitor Kazal type 1. The experiments were conducted by Johanna Wallner as part of her Master thesis.

4.2.4 siRNA knock down: effects on promoter activity (reporter gene assay)

***SPINK1* promoter activity after the siRNA knock down of *GATA6*, *GATA4* and *YY1* does not reflect the expected effects.** The missing effect sizes on total *SPINK1* mRNA expression after siRNA knock down of *GATA6*, *GATA4* and *YY1* prompted us to turn to a different approach. Instead of total *SPINK1* expression, we decided to assess the *SPINK1* promoter activity by reporter gene assay as an

alternative measure for GATA6, GATA4 and YY1 function. We combined the siRNA knock down with the overexpression of firefly luciferase vectors. The vectors harbored the firefly luciferase coding sequence, which was placed under the control of the *SPINK1* promoter. Additionally, a 500bp nucleotide fragment framing either the non-risk (C) or risk allele (T) of the *rs148911734* site was cloned upstream of the *SPINK1* promoter (vector cloning by Andreas Schmidt). Conceptually, the vector constructs would allow us to assess the transcription factor binding (activity) on the non-risk and risk allele independent of the genotype of the cell line. At the same time, we would use the full endogenous repertoire of *SPINK1* expressing cell lines and selectively (knock down) investigate the contribution of GATA6, GATA4 and YY1 within this context. Based on our hypothesis, we expected increased binding of GATA6 and GATA4 to the vector construct carrying the risk allele (C) compared to the construct carrying the non-risk allele (T). An increased binding would translate to higher promoter activation, increased firefly luciferase expression and higher light emission in those samples. The knock down of *GATA6* and *GATA4* would lead to an opposite effect, in which the vector constructs with the risk allele would be less activated. The activation of the construct with the non-risk allele should be equally high in the control and knock down setting. The same concept applies to YY1 and the activation of the non-risk allele while the risk allele should remain unaffected.

However, the experiments did not back up our assumptions. In Caco2 cells, the promoter activation under the control of the risk allele increased upon the *GATA6* knock down compared to the non-target siRNA transfected control. The activation of the non-risk allele was even stronger. Even with the *YY1* knock down, the result showed the opposite of our expectations. As with the *GATA6* knock down, the non-risk as well as the risk allele construct showed higher activation after the *YY1* knock down. The risk allele showed an even higher activation than the non-risk allele.

In HepG2 cells, the experimental results complied with our expectations in the case of the *GATA4* knock down. The activation of the risk allele construct decreased while the activation of the non-risk allele remained in the same range as the control. The combined knock down of *GATA4* and *GATA6*, however, caused an increase in the promoter activity of the risk allele construct, while the non-risk allele remained rather unaffected. The knock down of *YY1* led to an equally high activation of the non-risk

and risk allele construct, with the non-risk allele showing a slightly higher increase compared to the non-target siRNA control.

It is worth mentioning that the promoter activity without the transfection of siRNAs was generally higher than in the knock down experiments. The construct harboring the risk allele showed the highest activation in both Caco2 and HepG2 cells (2.8 - and 2.5-fold, respectively). The activation of the non-risk allele was lower (2.2-fold in Caco2 and 1.8-fold in HepG2). The activation of the construct which harbored the *SPINK1* promoter exclusively (no *rs148911734* fragment) showed the lowest activation and the activity accounted for 1.8-fold in Caco2 and 1.1-fold in HepG2. Please note further that all values were referenced to the activation of the *SPINK1*-promoter construct without the *rs148911734* flanking fragment in the cells which were transfected with the non-targeting siRNA. Taken together, those experiments barely allow to draw any conclusions. It is possible that the experimental setup created a highly artificial system, in which the cells suffered from unusually high stress due to the transfection of siRNAs and luciferase vectors at the same time. In addition, the experimental set up did not allow to control for knock down efficiency within each round of experiment. Both factors might contribute to the low effect sizes and high standard deviations of the present results (Figure 11).

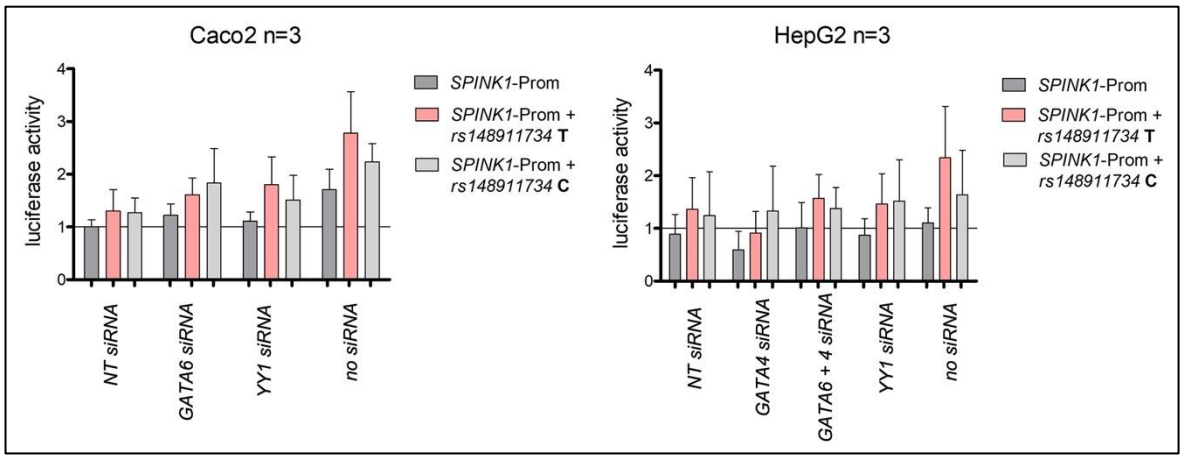


Figure 11 SPINK1 promoter activity after the siRNA knock down of GATA6, GATA4 and YY1 in Caco2 and HepG2 cells.

Luciferase activity was measured in Caco2 and HepG2 cells (technical triplicates) after the combined transfection of siRNAs for knock down of GATA6, GATA4 and YY1 and luciferase expression vectors (biological replicates of n=3). Firefly luciferase expression was placed under the control of the SPINK1 promoter (SPINK1-Prom). To analyze the effect of the non-risk (C) or risk allele (T) of the *rs148911734*, a 500bp nucleotide fragment framing either allele was cloned upstream of the SPINK1 promoter (SPINK1-Prom + *rs148911734* C; SPINK1-Prom + *rs148911734* T). All measurements are referenced to the luciferase activity

under the control of the SPINK1 promoter in cells transfected with the non-targeting (NT) siRNA. Standard deviation is shown. GATA: GATA binding protein, YY1: Ying Yang 1, Prom: Promoter. The experiments were conducted by Johanna Wallner as part of her Master thesis.

4.3 Interim conclusion

The current models are insufficient for the underlying research question - a new model approach is needed. The results described above demonstrated that we reached an impasse in conducting the necessary experiments to prove a functional dependency of *rs148911734*, the identified transcription factors GATA6, GATA4, YY1 and their effect on *SPINK1* expression and regulation. A major hurdle imposes the genotype of the two cell models Caco2 and HepG2, which both carry the *SPINK1* wild type. This hinders the investigation of the risk allele associated factors GATA6 and GATA4. Moreover, Caco2 and HepG2 are cells of intestinal and hepatic origin while we work on elucidating the pathomechanism of a pancreatic disease. The cell models from pancreatic origin did not qualify for our initial experiments because of their low *SPINK1* expression. To overcome the presented problems regarding the genotype, the origin of the cell line and the levels of *SPINK1* expression, we set out to develop a cell model that better reflects the necessary conditions.

4.4 Development of an acinar-like cell model

To develop an acinar model, we first conducted an extensive literature search on the topic of acinar cell identity and the key factors that determine it. The most important findings are summarized in the following sections.

4.4.1 Pre-considerations and incentive

Acinar to ductal metaplasia (ADM) shifts acinar cells towards a duct-like-phenotype and thereby disqualifies pancreatic cancer cell lines as a suitable model. Acinar cell plasticity and overexpression of acinar-specific transcription factors, however, open the gates for a targeted reprogramming of pancreatic ductal adenocarcinoma (PDA) cells. The use of pancreas derived cell lines for the study of acinar physiology and pathophysiology is seriously hampered for two reasons. First, primary acinar cells are difficult to obtain and culture due to a high auto-proteolytic activity of the cells once they have been

removed from the physiological context. Secondly, pancreatic cancer cells, which robustly sustain in cell culture and are easy to handle, have lost most of their original phenotype. In particular, the study of acinar genes becomes nearly impossible due to a reconditioning of the cells during cancerogenesis. Several studies established a direct link between the loss of acinar cell differentiation and the progression to a highly proliferative and duct-like phenotype (*e.g.*, Hoskins *et al*., 2014 (HNF1a), Luo *et al*., 2015 (HNF1a), Kim *et al*., 2015 (E47), Jakubison *et al*., 2018 (PTF1a), Kalisz *et al*., 2020 (HNF1a and KDM6)). The process is overall summarized under the term acinar to ductal metaplasia (ADM).

Even in their differentiated state, acinar cells seem to maintain a certain plasticity which allows adaptations in their phenotype based on internal and external challenges (reviewed by Storz, 2017). The overexpression of acinar specific transcription factors such as PTF1a and MIST1 in PDAC tumors of transgenic mice re-established the expression of digestive enzymes and acinar specific transcription factors as well as genes important to vesicle transport and the unfolded protein response. Hence, cellular programs vital to the physiological function of an acinar cell, namely the high production and trafficking of digestive enzymes, were turned back on upon the introduction of key acinar transcription regulators. Importantly, tumor growth was inhibited at the same time, pointing towards a physiological balance of increased differentiation and decreased proliferation (Jakubison *et al*., 2018). Similar findings were obtained from cell culture studies in the human PDA cell line Panc1. Stable overexpression of E47, a class A basic helix-loop-helix (bHLH) transcription factor, decreased cell proliferation and increased cell homeostasis as well as expression of acinar specific genes (Kim *et al*., 2015). Those studies laid the ground for our incentive to (1) identify the key regulatory acinar transcription factors from literature and (2) place these results in an experimental context.

4.4.2 Literature findings

Cockell *et al.* (1989) first identified the pancreas transcription factor 1 (PTF1) complex by DNA binding assays of nuclear extracts from the rat AR47J cell and synthetic oligonucleotides designed based on various digestive enzyme promoter sequences. Subsequent studies of Roux *et al.* (1989), Sommer *et al*. (1991), Krapp *et al*. (1996 and 1998), Beres *et al*. (2006) and Masui *et al*. (2007) provided detailed understanding of the composition and functioning of the complex and its target genes: The trimeric complex (PTF1-L) forms upon the dimerization of a tissues unspecific class A bHLH factor (e.g., E47), the pancreas specific class B bHLH factor PTF1a (p48) and RBPJL (p64). PTF1a binds to the E-box and RBPJL to the TC-box of the bipartite recognition site of PTF1. By RNA-Seq experiments in *Rbpjl* knock-out mice Masui *et al*. (2010) uncovered the substantial contribution of RBPJL to sufficient energy supply and secretory capacities in the adult acinar cells. Chromatin immune precipitation (ChiP) revealed diminished PTF1A binding to digestive enzyme promoters upon the loss of RBPJL. Holmstrom *et al*. (2011) compared ChiP-and RNA-sequencing data of *Rbpjl* and *Nr5a2* knock-out mice and identified co-regulation and -localization of the two transcription factors. Reporter gene assays further confirmed activation of the *Rbpjl* promoter by NR5A2. By using NR5A2 over-expression and luciferase reporter assays with *CEL* promoter constructs Fayard *et al*. (2003) verified *CEL* regulation by NR5A2 after Northern Blot analysis in situ hybridization of human and mouse pancreas. Boulling *et al*. (2017) demonstrated the cooperative action of PTF1-L and HNF1a on target gene activation by co-expression reporter gene assays in HEK293T on two CP-predisposing *SPINK1* promoter variants. Experiments conducted by Molero *et al*. (2012) displayed significantly reduced *NR5A2* transcript levels in *HNF1A* knock-out mice compared to wild type controls. HNF1A ChiP revealed binding of HNF1A to the promoter of *Nr5a2* and reporter gene assays verified *Nr5a2* promoter activation by HNF1A. Taken together, we concluded that the overexpression of the PTF1 complex, consisting of the E47, PTF1a and RBPJL subunit, HNF1a and NR5A2 had the potential to reconstitute an acinar phenotype in PDA cells. Figure 12 provides an overview of the literature findings on acinar target gene regulation.

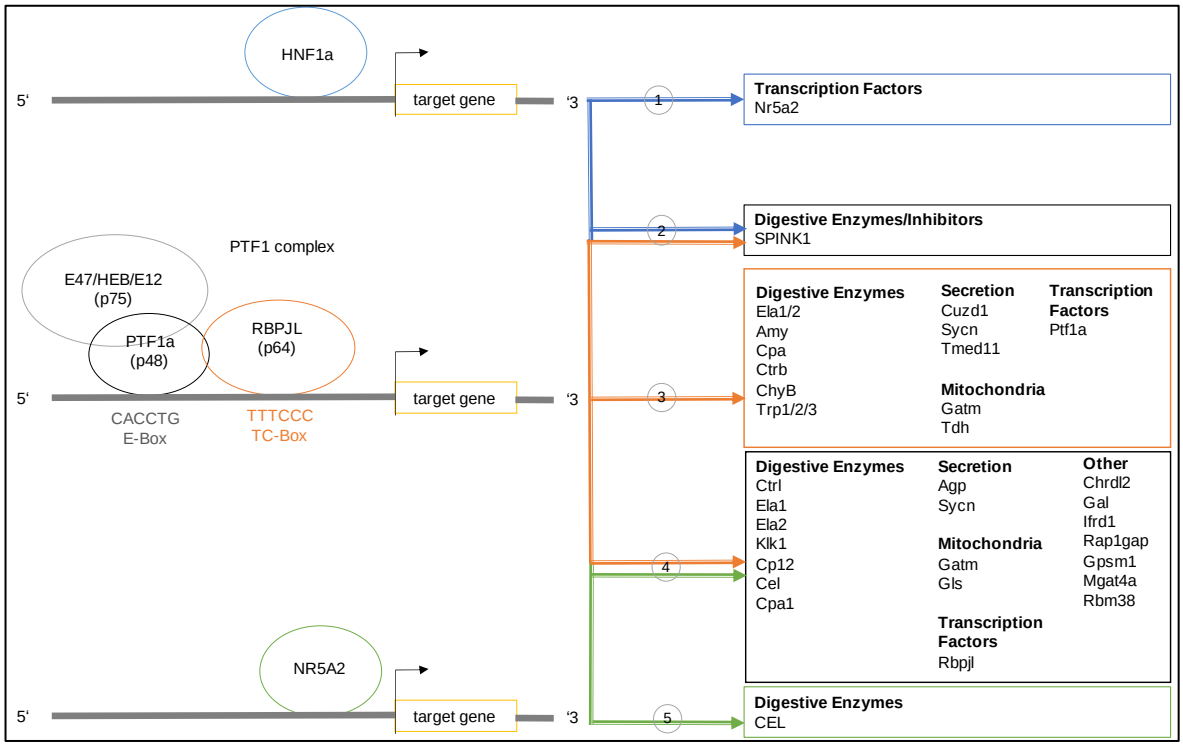


Figure 12 Summary of acinar target gene activation and regulation by the key transcription factors identified in literature.

(1) Molero *et al.* (2011), (2) Boulling *et al*. (2017), (3) Cockell *et al.* (1989), Roux *et al*. (1998), Sommer *et al*. (1991), Krapp *et al.* (1998), Beres *et al*. (2006), Masui *et al*. (2007) (4) Holmstrom *et al*. (2011) (5) Fayard *et al*. (2003). Target genes in capital letters indicate evidence based on human data. Target genes in small letters indicate evidence based on mouse experiments. PTF1: Pancreas Transcription Factor 1, E12/E47: E2A immunoglobulin enhancer-binding factors E12/E47 (Transcription factor 3), HEB: HeLa E-box binding protein, PTF1a: Pancreas associated transcription factor 1a, RBPJL: recombination signal binding protein for immunoglobulin kappa J region like, HNF1a: Hepatocyte nuclear factor 1 homeobox A, NR5A2: nuclear receptor subfamily 5 group A member 2, SPINK1: serine protease inhibitor Kazal type 1, Ela: Elastase, Amy: Amylase, Cpa/Cp: Carboxypeptidase, Ctrb: Chymotrypsin B, Trp: Trypsin, Cuzd1: CUB and zona pellucida like domains, Sync: Syncollin, Tmed11: transmembrane p24 trafficking protein 11, Gatm: glycine amidinotransferase, Tdh: L-threonine dehydrogenase, Ctrl: Chymotrypsin like, Klk1: Kallikrein 1, Cel: Carboxyl ester lipase, Agp: Aquaporin, Gls: Glutaminase, Chrdl2: Chordin like 2, Gal: galanin and GMAP prepropeptide, Ifrd1: interferon related developmental regulator 1, Rap1gap: Rap1 GTPase-activating protein, Gpsm1: G-protein signaling modulator 1, Mgat4a: mannoside acetylglucosaminyltransferase 4, isoenzyme A, Rbm38: RNA binding motif protein 38

4.4.3 Experimental application of literature findings in Panc1 and Paca44

We transfected Panc1 cells with different combinations of the identified transcriptions factors (E47, PTF1a, RBPJL, HNF1a, NR5A2) and assessed their impact on a panel of target genes. To control the successful overexpression of the transcription factors, we analyzed their expression on mRNA and protein level. To this end, we measured the expression of *TCF3* (E47), *PTF1a*, *RBPJL*, *HNF1a* and

NR5A2 by qPCR (TaqMan assays, Figure 14) and analyzed total protein lysates by Western blot (Figure13). We obtained clear signals for all factors at the expected protein size. Only for NR5A2 we could not detect the expected signal. As a possible explanation, we suspected disruption of the NR5A2 synthesis and stability by the FLAG-tag. Therefore, we adopted the vector construct and removed the FLAG-tag. With this new construct, we detected NR5A2 on Western blots with an antibody directed against the protein itself, but the signal never reached a comparable intensity to that of the other factors (work of Katharina Eiseler, results not shown).

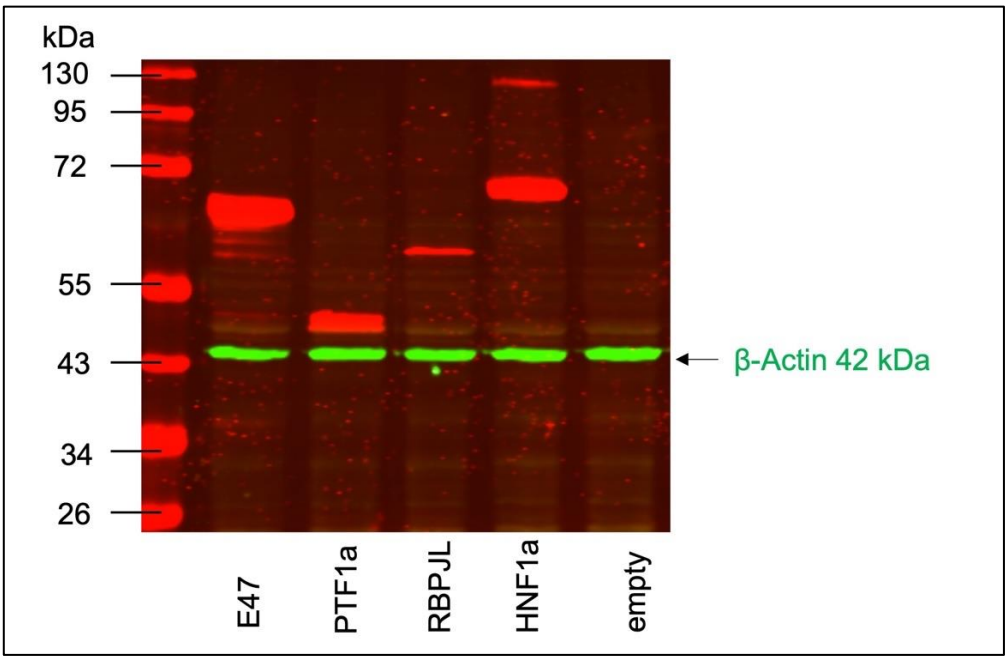


Figure 13 Western blot of total lysate from transfected Panc1 cells

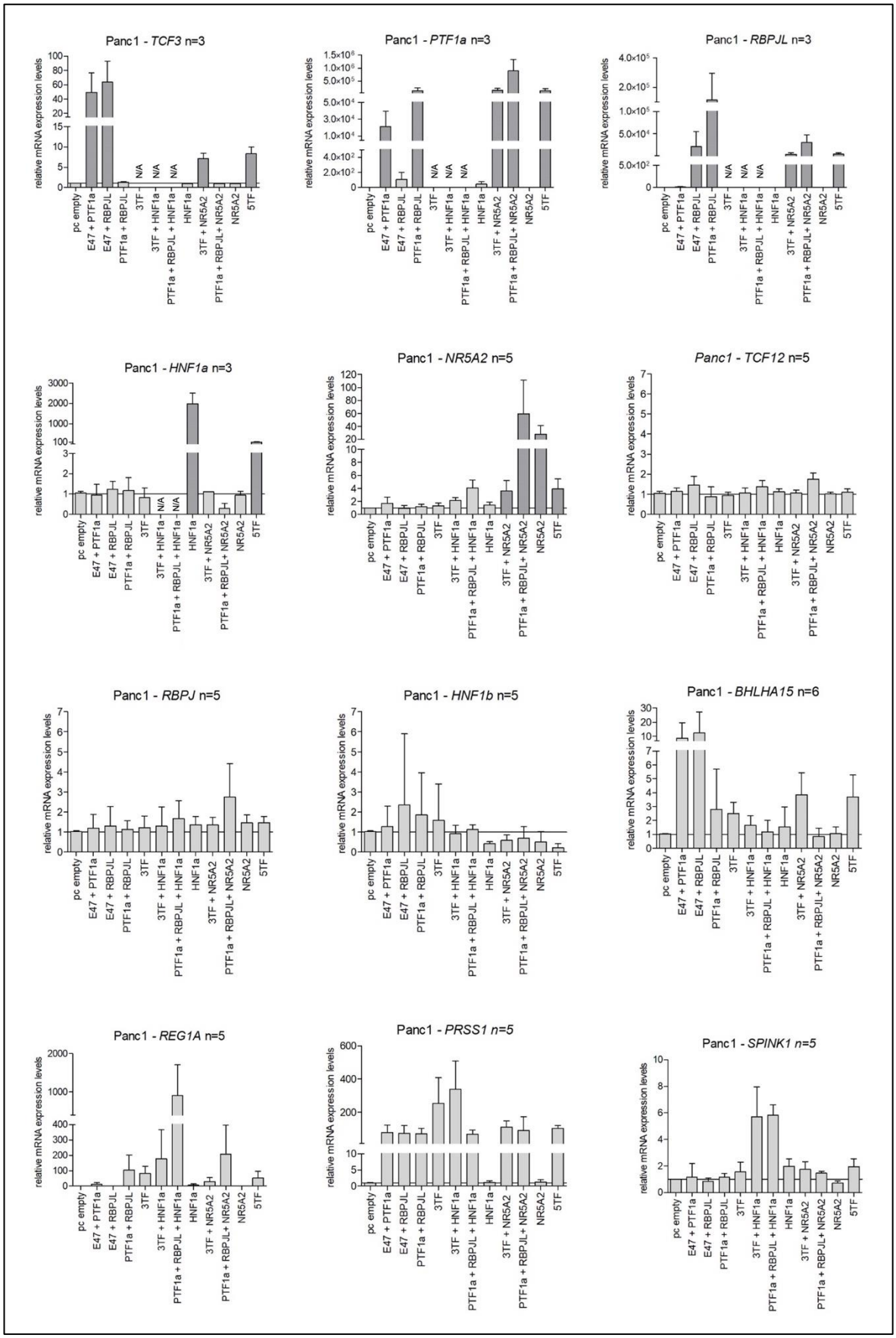
Overexpressed transcription factors were detected by an antibody directed against the C-terminal FLAG-tag (in red: E47: 64 kDa; PTF1a: 35 kDa; RBPJL 57 kDa; HNF1a: 67 kDa, FLAG-tag + 1 kDa). β-Actin (green, 42 kDa) was used as internal loading control. Transfection with the pcDNA3.1 vectors without insert (empty) served as negative control. Representative of a minimum of n=2 is shown. 50 µg of total lysate were used. E47: E2A immunoglobulin enhancer-binding factors E12/E47 (Transcription factor 3), PTF1a: Pancreas associated transcription factor 1a, RBPJL: recombination signal binding protein for immunoglobulin kappa J region like, HNF1a: Hepatocyte nuclear factor 1 homeobox A.

The panel of target genes consisted of different genes that were supposed to reflect the characteristics of an acinar phenotype. The acinar marker genes were selected from the following categories: proteases and inhibitor (*PRSS1, SPINK1, CPA1, CELA3AB, CTRC, CTRL*), lipases and related proteins (*CEL, CLPS, PNLIP, PNLIPRP1, PLA2G1B*), zymogen granule and secreted proteins (*CUZD1, GP2,*

REG1A, SYTL1, SYCN), transcription factors (*TCF12, RBPJ, HNF1B, BHLHA15*). Furthermore, we measured ductal genes (*TRPV6, CFTR*) and the ER stress marker BiP (*HSPA5*).

The induction of an acinar-like phenotype is possible in Panc1. The overexpression of E47, PTF1a, RBPJL, HNF1a and NR5A2 led to an increase in the corresponding mRNA expression level. The expression level varied based on the amount and the combination of transcription factors. The fold change differed greatly between the individual factors and covered a wide range from an increase around a few hundred up to a few million-fold. We suspected mutual induction of the factors that would boost gene expression beyond the exogenous expression from the transfected vectors. On the one hand, we included the measurement of *TCF12, RBPJ* and *HNF1b* to control for the expression of similar and related transcription factors. Those genes merely responded to the overexpression. On the other hand, *BHLHA15* (MIST1) was measured as potential target. Indeed, an increase in mRNA expression of *BHLHA15* was most related to the overexpression of E47. We further assessed the expression of *HSPA5* (BiP), a common marker for ER stress. The simultaneous overexpression of several transcription factors and the expected induction of multiple target genes were likely to impose an intolerable load of mRNA expression and protein synthesis for the cells. Since we were aiming for an improved cell model, we wanted to avoid the induction of ER stress. Fortunately, *HSPA5* were unaffected by the different transfection conditions (appendix). With respect to the induction of an acinar-like phenotype, we investigated the expression of various categories of acinar marker gene (see above). The induction of mRNA expression of *PRSS1*, *CELA3AB* (proteases), *CEL* (lipase), *CUZD1* and *SYCN* (zymogen granule proteins) was most pronounced upon the overexpression of the PTF1 complex (E47, PTF1a, RBPJL) and HNF1a as an optional fourth factor. E47 seemed to contribute especially to the gene expression of *CELA3AB*. We observed the same effect for the induction of *TRPV6* (ductal) and *CTRL* (protease), for which the combination of E47 and RBPJL and to a lesser extent E47 and PTF1a revealed the highest expression levels. The increase of *SPINK1* and *CTRC* (protease inhibitors and protease) and *REG1A* (secretory protein) mRNA expression was mostly due to PTF1a, RBPJL and HNF1a. Induction of *SPINK1* and *CTRC*, however, fell short of our expectations. E47 further appeared to play a minor role in

CLPS induction, since we observed the strongest increase in mRNA expression with PTF1a and RBPJL. For the *CPA1* (protease) expression, NR5A2 appeared to be particularly effective. When NR5A2 was included in either combination with the other factors, the levels in *CPA1* expression were boosted far beyond the effect of the PTF1 complex and HNF1a. This latter combination was most effective for the proteases *PRSS1* and *CELA3AB*. The full panel of the experiments and target gene effects is summarized in Figure 14. No response to the overexpression of either transcription factor combination was observed in the case of CFTR (ductal) as well as PNLIP, PNLIPRP1 and PLA2G1B (lipases). The effects for SYTL1 and for GP2 were minor or ambiguous, which is why neither are shown in the summary figure.


Panc1 - TCF3 n=3
Panc1 - PTF1a n=3
Panc1 - RBPJL n=3
Panc1 - HNF1a n=3
Panc1 - NR5A2 n=5
Panc1 - TCF12 n=5
Panc1 - RBPJ n=5
Panc1 - HNF1b n=5
Panc1 - BHLHA15 n=6
Panc1 - REG1A n=5
Panc1 - PRSS1 n=5
Panc1 - SPINK1 n=5
relative mRNA expression levels
pc empty
E47 + PTF1a
E47 + RBPJL
PTF1a + RBPJL
3TF
3TF + HNF1a
PTF1a + RBPJL + HNF1a
HNF1a
3TF + NR5A2
PTF1a + RBPJL+ NR5A2
NR5A2
5TF

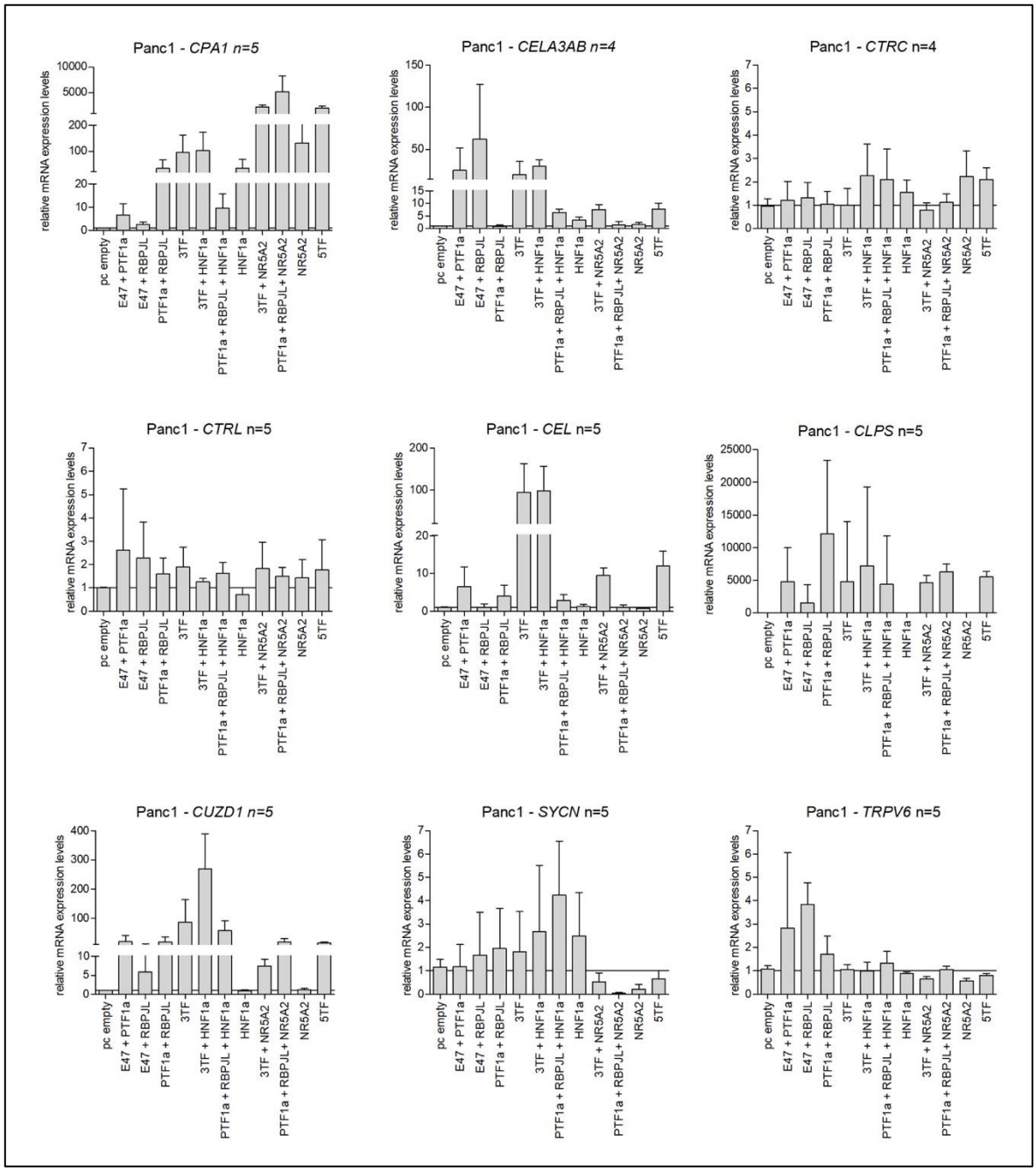


Figure 14 Expression of pancreatic genes upon the overexpression of acinar key transcription factors.

Total RNA was collected 72 h post transfection and transcribed to cDNA. Relative mRNA expression of the indicated target genes was measured by qPCR with TaqMan assays (technical triplicates). Expression of β-Actin and GAPDH served as internal controls. Expression levels were normalized to control transfected Panc1 cells (pc empty). The horizontal line illustrates the baseline relative mRNA expression (=1). 3TF represents the trimeric PTF1 complex consisting of E47 (TCF3), PTF1a, and RBPJL. 5TF represents the combined overexpression of E47 (TCF3), PTF1a, RBPJL, HNF1a and NR5A2. Cell line, target gene and number of biological replicates are given above each graph. Bars in dark grey indicate the measurement of a transfected factor. Standard deviation is shown. N/A: not available. TCF3: Transcription factor 3 (E47), PTF1a: Pancreas associated transcription factor 1a, RBPJ: recombination signal binding protein for immunoglobulin kappa J region, RBPJL: RBPJ-like, HNF1a: Hepatocyte nuclear factor 1 homeobox A, NR5A2: nuclear

receptor subfamily 5 group A member 2, TCF12: Transcription factor 12, HNF1b: Hepatocyte nuclear factor 1 homeobox B, PRSS1: Serine protease 1 (trypsinogen), SPINK1: serine protease inhibitor Kazal type 1, CPA1: Carboxypeptidase A1, CELA3AB: Chymotrypsin like elastase 3A and 3B, CTRC: Chymotrypsin C, CTRL: Chymotrypsin like, CEL: Carboxyl ester lipase, CLPS: Colipase, CUZD1: CUB and zona pellucida like domains 1, SYTL1: Synaptotagmin like 1, SYCN: Syncollin, TRPV6: Transient receptor potential cation channel subfamily V member 6, BHLHA15: Basic helix-loop-helix family member a15 (MIST1), REG1A: Regenerating family member 1 alpha

Paca44 cells are less susceptible to *SPINK1* induction. The promising induction of acinar genes in Panc1 cells encouraged us to apply the same to the Paca44 cell line, which was of particular interest because of its *SPINK1* p.N34S heterozygosity. For the transfection, we used a reduced set of transcription factors combinations that effectively induced *SPINK1* in the previous experiment, namely E47 (optional), PTF1a, RBPJL and HNF1a. We checked the expression of these factors again by qPCR (TaqMan assays). The panel of target genes was reduced to only a subset of the previously assessed marker genes (*PRSS1, SPINK1, CPA1, CELA3AB, CUZD1, CEL, CLPS, REG1A, NR5A2, BHLHA15*). *GATA6* and *YY1* were included as additional markers. The results are summarized in (Figure 15). The highest induction of *SPINK1* in Paca44 after the overexpression of PTF1a, RBPJL and HNF1a was significantly lower (2-fold) than in Panc1 (6-fold). Additionally, the fluctuation of the observed effects was high as reflected in the high standard deviation. Given that Paca44 cells may only have one functional allele for *SPINK1* expression, the highest expected effect would have been half the fold change (3-fold) we observed in Panc1. The expression of GATA6 and YY1 was not affected by any of the combinations of the transcription factors.

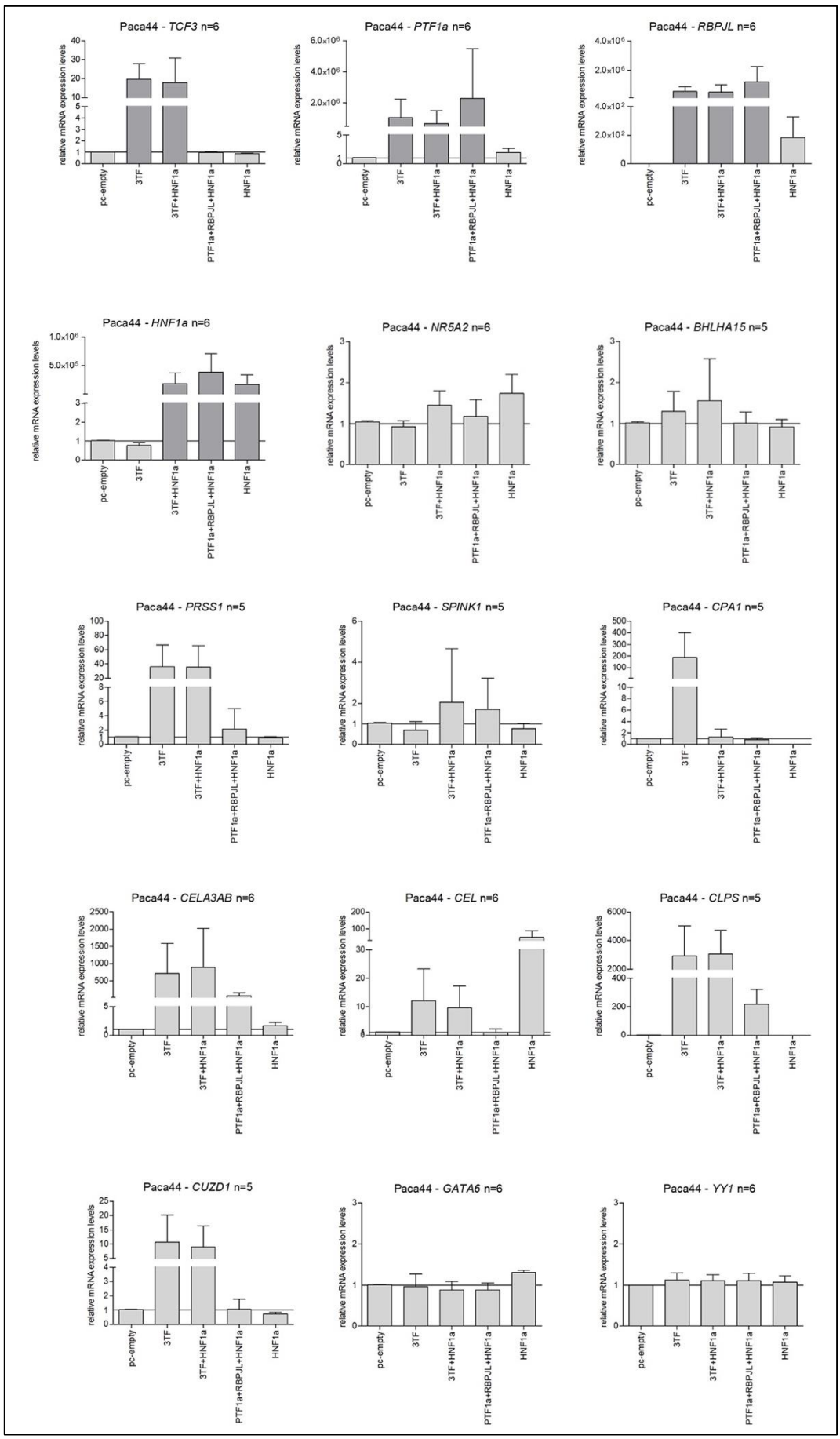

Paca44 - TCF3 n=6
Paca44 - PTF1a n=6
Paca44 - RBPJL n=6
Paca44 - HNF1a n=6
Paca44 - NR5A2 n=6
Paca44 - BHLHA15 n=5
Paca44 - PRSS1 n=5
Paca44 - SPINK1 n=5
Paca44 - CPA1 n=5
Paca44 - CELA3AB n=6
Paca44 - CEL n=6
Paca44 - CLPS n=5
Paca44 - CUZD1 n=5
Paca44 - GATA6 n=6
Paca44 - YY1 n=6
relative mRNA expression levels
pc-empty
3TF
3TF+HNF1a
PTF1a+RBPJL+HNF1a
HNF1a

Figure 15 Expression of acinar genes and transcription factors upon the overexpression of acinar key transcription factors in Paca44

Total RNA was collected 72 h post transfection and transcribed to cDNA. Relative mRNA expression of the indicated target genes was measured by qPCR with TaqMan assays (technical triplicates). Expression of β-Actin and GAPDH served as internal controls. Expression levels were normalized to control transfected Paca44 cells (pc empty). The horizontal line illustrates the baseline relative mRNA expression (=1). 3TF represents the trimeric PTF1 complex consisting of E47 (TCF3), PTF1a, and RBPJL. Bars in dark grey indicate the measurement of a transfected factor. Standard deviation is shown. Cell line, target gene and number of biological replicates are given above each graph. For full gene names please refer to legend of Figure 14. GATA6: GATA binding protein 6, YY1: Ying Yang 1

4.5 Implementation of the previous findings within the adopted model

After the successful induction of *SPINK1* expression in Panc1 (6-fold) and in Paca44 (2-fold) we set out to apply the previous experiments on our adopted models. Our aim was not only to reproduce the findings that we obtained from the studies in Caco2 and HepG2 cells, but also to go beyond the previous impasse and to deepen the functional analysis.

4.5.1 Sanger sequencing in Paca44

Paca44 cells show a decreased expression of the risk allele. The risk allele becomes detectable upon the combined expression of E47, PTF1a, RBPJL and HNF1a. Based on the *SPINK1* p.N34S heterozygosity of the Paca44 cells, we examined the allelic expression by Sanger sequencing. We applied the same procedure as for the human intestinal RNA samples of *SPINK1* p.N34S heterozygotes (section 4.1). Paca44 cells were transfected with *E47*, *PTF1a*, *RBPJL* and *HNF1a* in different combinations. The transfection with the pcDNA3.1 vectors without insert (empty) served as baseline control. In addition to mRNA, gDNA of Paca44 cells (genotype TC, heterozygous control) as well as gDNA and mRNA of Caco2 cells (genotype TT, homozygous non-risk allele control) were included (Figure 16). As expected, the electropherograms of the gDNA samples showed a clear double peak (TC) for the Paca44 samples. We observed a single peak (T) for the Caco2 gDNA as well as the mRNA sample. The electropherograms of the Paca44 mRNA samples presented distinctly: a single peak for the non-risk allele (T) appeared in the control and *E47*, *PTF1a*, and *RBPJL* transfected samples. The risk allele (C) only showed along with the non-risk allele upon the combined transfection of *E47* (optional), *PTF1a*, *RBPJL* and *HNF1a*. The risk allele (C)

displayed a weaker signal than the non-risk allele (T) in both transfections with and without *E47*. A decreased expression of the risk allele was thus apparent in all samples. We assumed that the overexpression of the *E47* (optional), *PTF1a*, *RBPJL* and *HNF1a* would boost the *SPINK1* expression independently of the *SPINK1* p.N34S haplotype associated variants to an extent, that the risk allele would become detectable. This assumption was based on the previous finding that total *SPINK1* expression was induced in Panc1 and Paca44 cells by the same factors.

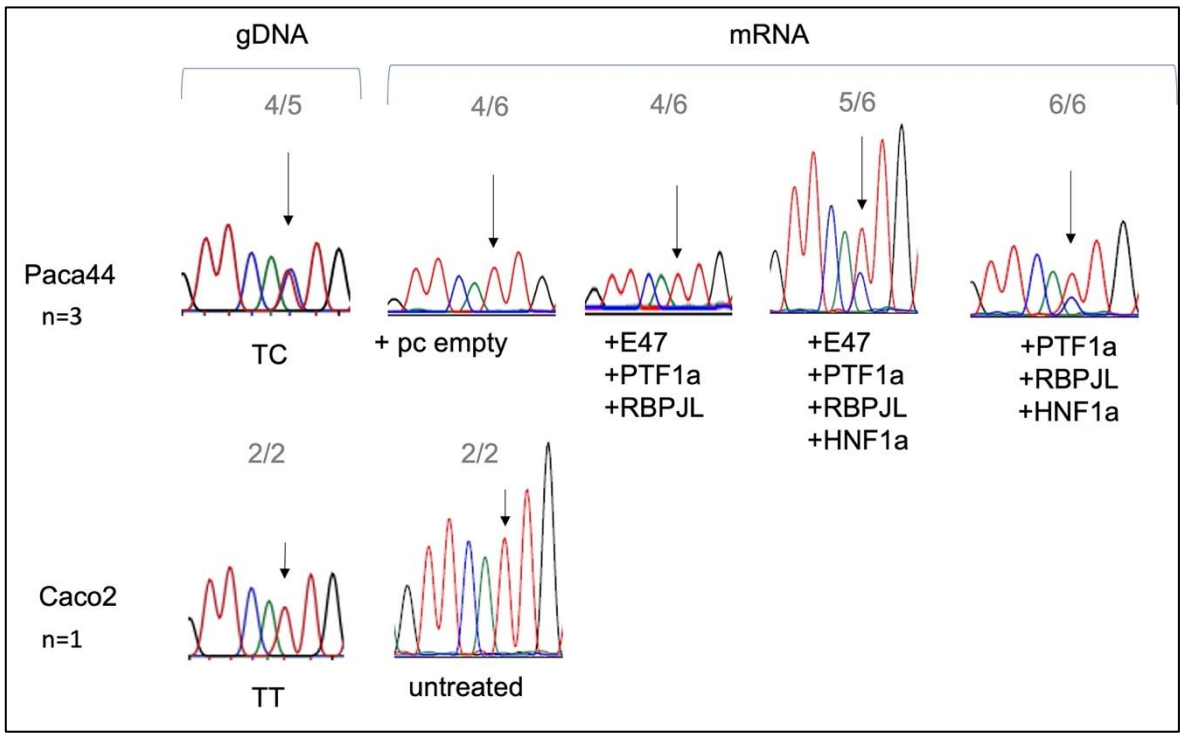


Figure 16 Electropherograms of gDNA and mRNA sequencing of Paca44 and Caco2 cells at SPINK1 p.N34S.

Genomic DNA (gDNA) and mRNA of Paca44 and Caco2 were Sanger sequenced with two different reverse primers (technical duplicates) located downstream of *SPINK1* p.N34S (c.101 T>C). The reverse complementary sequence is shown. Paca44 cells were transfected with E47, PTF1a, RBPJL and HNF1a. Transfection with the pcDNA3.1 vector without insert (empty) was used as baseline control. RNA was collected and measured from three independent experiments (n=3). The Caco2 cells were left untreated. Representative electropherograms were selected. The numbers above each graph indicate the number of the presented result out of the total number of measurements. Total measurements <6 are due to excluded electropherograms with poor sequencing quality. Heterozygous: TC, homozygous: TT. Color code of electropherogram: red = thymine (T), blue = cytosine (C), green = adenine (A), black = guanine (G). E47: E2A immunoglobulin enhancer-binding factors E12/E47 (Transcription factor 3), PTF1a: Pancreas associated transcription factor 1a, RBPJL: recombination signal binding protein for immunoglobulin kappa J region like, HNF1a: Hepatocyte nuclear factor 1 homeobox A.

Modulation of the risk allele expression by *GATA6* and *GATA4* overexpression in Paca44 cells remains inconclusive. After the risk allele (C) became detectable in the Sanger Sequencing electropherograms due to the overexpression of *E47* (optional), *PTF1a*, *RBPJL* and *HNF1a*, the question arose whether we would detect an *in vivo* repression of the risk allele upon the overexpression of *GATA6* and *GATA4*. Consequently, we expected a decrease or even complete loss of the risk allele (C) signal. We expanded the previous overexpression set of *PTF1a*, *RBPJL* and *HNF1a by GATA6* and *GATA4* either separately or in combination. As control samples, we used the pcDNA3.1 vectors without insert (empty) and again added *GATA6* and *GATA4* separately or combined (Figure 17). The transfection with pcDNA3.1 vector without insert (empty) led to the same observation as in the previous experiments and only the non-risk allele (T) was displayed. Due to the proclaimed repressive action of GATA6 and GATA4, we did not expect any changes in the electropherogram upon their overexpression. Strikingly, the non-risk allele (C) appeared after the overexpression of *GATA6* and *GATA4* and to a lesser extent when we combined both factors. The non-risk allele (T) showed a stronger signal than the risk allele (C) in all cases.

The combined transfection of *PTF1a*, *RBPJL* and *HNF1a* increased the expression of the risk allele as observed before. Contrary to what was expected, the overexpression of *GATA6* and *GATA4* did not lead to a decrease or even loss of the risk allele (C) signal. Methodologically, it should be added that the evaluation of the electropherogram signals only allows a rough estimate, but no actual quantification. The results should therefore be interpreted with caution. For quantification, an allele-specific qPCR or RNA sequencing would be the method of choice.

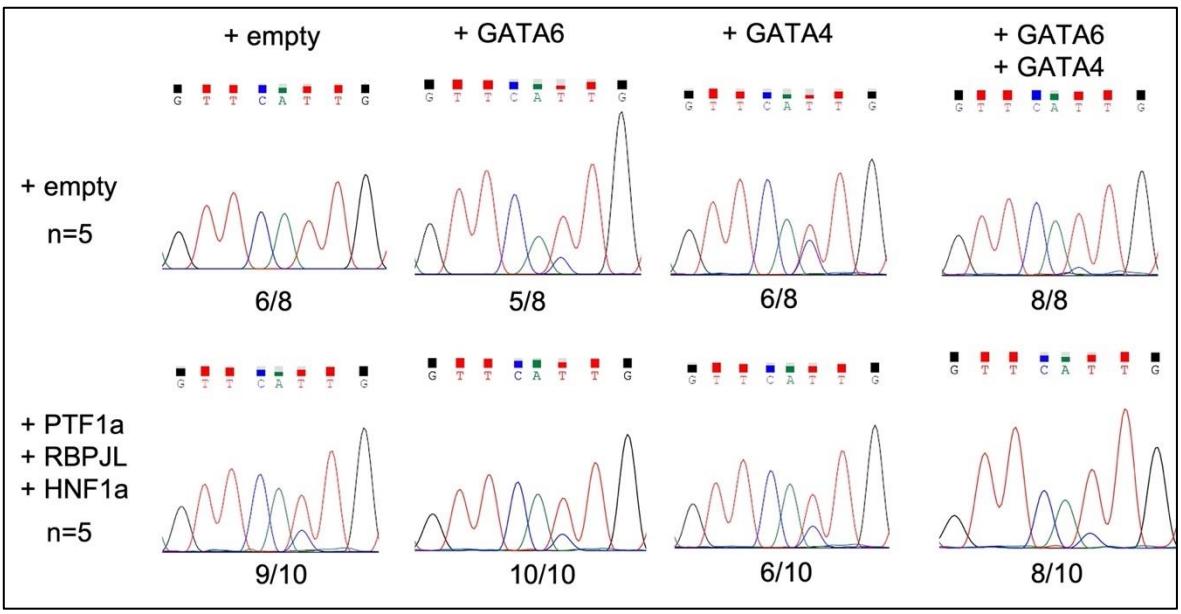


Figure 17 Electropherograms of mRNA sequencing of Paca44 cells after transcription factor overexpression at SPINK1 p.N34S.

mRNA of Paca44 was Sanger sequenced with two different reverse primers (technical duplicates) located downstream of *SPINK1* p.N34S (c.101 T>C). Paca44 cells were transfected with different transcription factor combinations. The reverse complementary sequence is shown. Transfection with the pcDNA3.1 vector without insert (empty) and the combination of PTF1a, RBPJL and HNF1a was used as baseline control. The additional transfection with GATA6 and GATA4 aimed to assess their effect on the risk allele (C, blue). RNA was collected and measured from five independent experiments (n=5). Representative graphs are shown. The numbers below each graph indicate the number of the presented result out of the total number of measurements. Total measurements <10 are due to excluded electropherograms with poor sequencing quality. E47: E2A immunoglobulin enhancer-binding factors E12/E47 (Transcription factor 3), PTF1a: Pancreas associated transcription factor 1a, RBPJL: recombination signal binding protein for immunoglobulin kappa J region like, HNF1a: Hepatocyte nuclear factor 1 homeobox A, GATA: GATA binding protein. The experiments were conducted by Johanna Wallner as part of her Master thesis.

The influence of *YY1* on the expression of the non-risk allele by siRNA knock down in Paca44 cells remains unclear. To assess the *in vivo* effect of the potential enhancer YY1 on the non-risk allele, we decided to combine the overexpression of *PTF1a*, *RBPJL* and *HNF1a* with the knock down of *YY1*. As control, we overexpressed the pcDNA3.1 vector without insert (empty) and used non-targeting (NT) siRNA (Figure18). According to our expectations, the knock down of YY1 would lead to a reduced expression of the non-risk allele (T), which would lead to a decreased signal in the electropherogram. Likewise, a reduced activation would converge the signal of the non-risk allele (T) and the risk allele (C) to a similar signal strength. This effect was in part observed when we looked at the pcDNA3.1 vector without insert (empty) transfection and the effect of the non-targeting siRNA

compared to the *YY1* siRNA. While the risk allele (C) was barely detectable in case of the non-targeting siRNA, the signal seemed to intensify under the addition of the *YY1* siRNA as a consequence of the decrease in the non-risk allele (T) signal strength. However, when we compared the same knock down conditions during the overexpression of *PTF1a*, *RBPJL* and *HNF1a, we* could not reproduce this effect. Instead, the signal strength of the non-risk (T) and risk allele (C) seemed to diverge. We further conducted the same experiments with the knock down of *GATA6* and *GATA4*. Interestingly, the knock down of *GATA6* in combination with the overexpression of *PTF1a*, *RBPJL* and *HNF1a* converged the signal of the non-risk (T) and risk allele (C) to similar peak height. However, endogenous *GATA6* expression in Paca44 became highly questionable based on later experiments. Measurements of mRNA expression level by qPCR and protein expression by super-shift EMSA experiments rather argued in favor of a lack of GATA6 in those cells. Hence, the effects of a *GATA6* siRNA knock down are questionable. In addition to the critical assessment in the section above, the combination of overexpression and knock down experiments are likely to create a highly stressful setting for the cells and increase the risk for experimental artefacts.

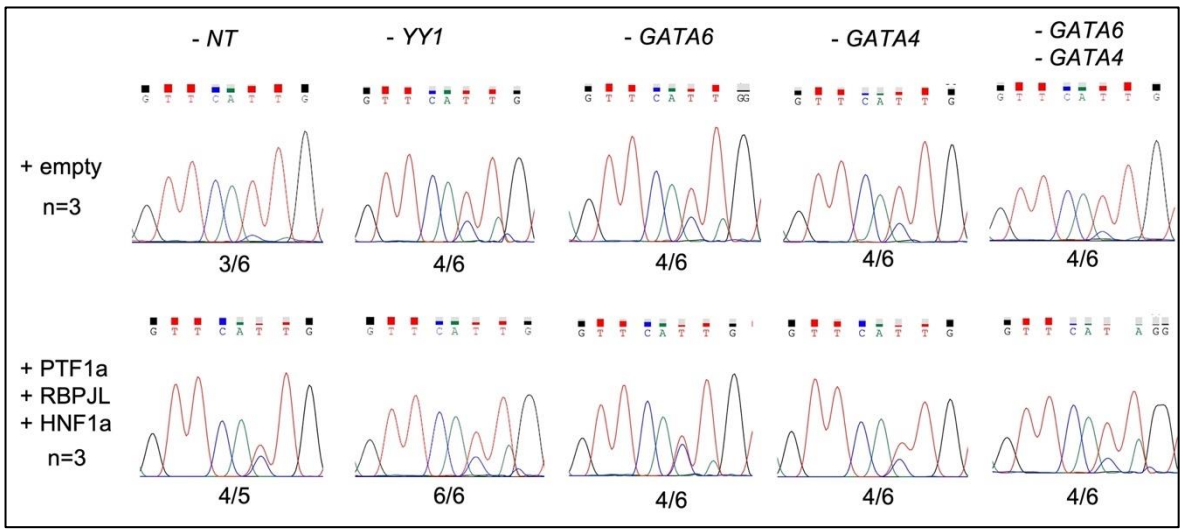


Figure 18 Electropherograms of mRNA sequencing of Paca44 cells at SPINK1 p.N34S after transcription factor overexpression and simultaneous knock down.

mRNA of Paca44 was Sanger sequenced with two different reverse primers (technical duplicates) located downstream of *SPINK1* p.N34S (c.101 T>C). The reverse complementary sequence is shown. Paca44 cells were transfected with the pcDNA3.1 vector without insert (empty) and the combination of PTF1a, RBPJL and HNF1a as baseline condition. The additional transfection with siRNA targeting YY1, GATA6 and GATA4 aimed to assess their effect on the risk allele (C, blue) and non-risk (T, red) allele. Transfection

with non-targeting (NT) siRNA served as control (3/6 showed exclusively the T allele, 3/6 showed a minor peak for the C allele). RNA was collected and measured from three independent experiments (n=3). Representative graphs are shown. The numbers below each graph indicate the number of the presented result out of the total number of measurements. Total measurements <6 are due to excluded electropherograms with poor sequencing quality. Color code of electropherogram: red = thymine (T), blue = cytosine (C), green = adenine (A), black = guanine (G). E47: E2A immunoglobulin enhancer-binding factors E12/E47 (Transcription factor 3), PTF1a: Pancreas associated transcription factor 1a, RBPJL: recombination signal binding protein for immunoglobulin kappa J region like, HNF1a: Hepatocyte nuclear factor 1 homeobox A, GATA: GATA binding protein, YY1: Ying Yang 1. The experiments were conducted by Johanna Wallner as part of her Master thesis.

4.5.2 EMSA in Panc1 and Paca44

After the successful induction of *SPINK1* expression upon the overexpression of *E47*, *PTF1a*, *RBPJL* and *HNF1a* in Panc1 (*SPINK1* wild type) and to a lesser extent in Paca44 (*SPINK1* p.N34S heterozygous), we wondered whether the binding of those factors to the *SPINK1* regulatory region at *rs142703147 and rs148911734* might be affected. We conducted EMSA experiments with nuclear extracts of Panc1 in which we had overexpressed the corresponding factors either separately or in combination.

Overexpression experiments in Panc1 argue against PTF1a and HNF1a as relevant transcriptional regulators at *rs142703147*. Overexpression experiments in Panc1 argue against PTF1a and HNF1a as relevant transcriptional regulators at *rs142703147*. A recent publication of Boulling *et al.* suggested a disruption of a PTF1-HNF1a binding site by the risk allele (A) of *rs142703147* (Boulling *et al*., 2017). We therefore expected an increased binding of PTF1a and HNF1a at the non-risk allele (C) in the EMSA. Surprisingly, we did not observe any differential binding pattern for nuclear extracts from the control transfected cells (pcDNA3.1 vector without insert (empty)) nor from the single transfections with *PTF1a*, *RBPJL* and *HNF1a*. Only in the case of *E47*, an additional band appeared at the non-risk allele (C), which was lost at the risk allele (A). We also observed this band in the combined transfections of *E47*, *PTF1a*, *RBPJL* and *HNF1a* (Figure 19 A), but not in the combined overexpression of *PTF1a, RBPJL and HNF1a* without E47 (data not shown). Since the induction of *SPINK1* expression was independent of *E47* (qPCR measurements), this effect was not considered relevant for the underlying research question and no follow up experiments were conducted. However, the experiments show that disturbed DNA binding of PTF1a

and HNF1a to *rs142703147* is probably not the pathogenic cause of the *SPINK1* p.N34S haplotype.

Overexpression experiments in Panc1 rule out a direct interaction of E47, PTF1a, RBPJL and HNF1a and *rs148911734*. When we conducted the above described EMSA experiment with *rs148911734*, we did not overserve any differential binding patterns which could be assigned to the single or combined overexpression of *E47*, *PTF1a*, *RBPJL* and *HNF1a* (Figure 19 B). Importantly, regardless of the overexpressed factors, the nuclear extracts from Panc1 showed the same binding pattern that we previously observed in Caco2 and HepG2. The bands appeared at the same height and pattern as the bands we identified by super-shift EMSA as binding of GATA6, GATA4 and YY1. Taken together, we were able to rule out any allele dependent binding of E47, PTF1a, RBPJL and HNF1a at *rs148911734*. The experiments further suggest that the binding of GATA6, GATA4 and YY1 at this site is independent of those factors.

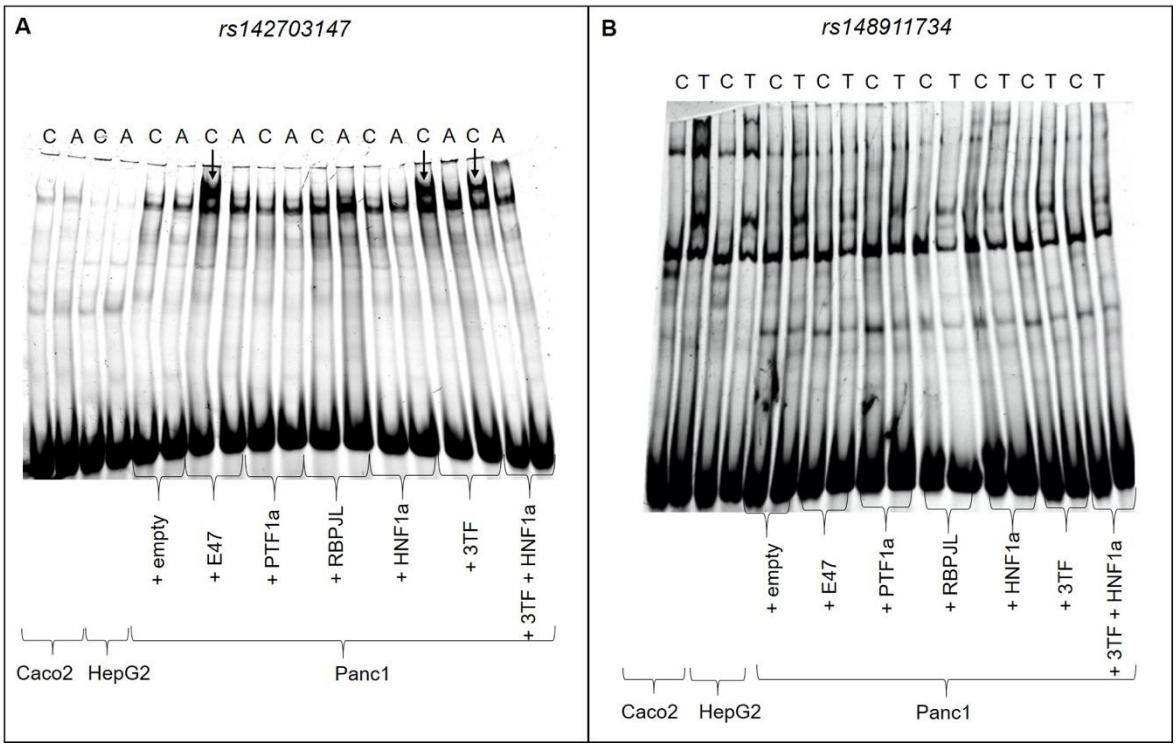


Figure 19 Nuclear protein binding of Panc1 nuclear extracts after transcription factor overexpression.

Panc1 cells were transfected with E47, PTF1a, RBPJL and HNF1a either as single factors or in combination. Transfection with the pcDNA3.1 vector without insert (empty) was used as baseline control. Nuclear extracts were prepared from three independent experiments (n=3, representative shown). Nuclear protein from Caco2 and HepG2 from the affinity chromatography experiments were included as controls for the previously observed binding

patterns. The binding reaction with the non-risk allele was applied first, followed by the risk allele. (A) EMSA experiments with nuclear extracts of Panc1 and Cy5-labelled synthetic oligonucleotides harboring the non-risk (C) or risk (A) allele of rs142703147. Increased binding of E47 to the non-risk allele is highlighted by the black arrows. (B) EMSA experiments with nuclear extracts of Panc1 and Cy5-labelled synthetic oligonucleotides harboring the non-risk (C) or risk (T) allele of rs148911734. Nuclear extract: 3 µg, polydIdC: 350 ng, Cy5-oligonucleotide: 1 ng. EMSA: Electron Mobility Shift Assay; *rs142703147* C>A, rs148911734C>T, 3TF: represents the trimeric PTF1 complex consisting of E47, PTF1a, and RBPJL.

Nuclear proteins of Panc1 and Paca44 show diminished to no binding of GATA6 at *rs148911734*. We conducted super-shift EMSA experiments with nuclear extracts from Panc1 and Paca44 at the non-risk (C) and risk all (T) allele of *rs148911734* (Figure 20) to assess further the transcription factor binding we observed in the previous experiments. Especially anti-GATA4 and to a minor extent anti-GATA6 antibodies caused a shift at the risk allele (T) with Panc1 extracts. The addition of anti-SATB2 or ZNF658 antibodies did not affect the binding pattern, while the anti-YY1 antibody clearly shifted the binding pattern at the non-risk allele (C). In the case of Paca44, only the addition of the anti-YY1 antibody caused the expected shift at the non-risk allele. The addition of any other antibodies remained without effect. In summary, differential binding of Panc1 nuclear protein to the risk allele (T) is mostly attributed to GATA4 and to a minor extent to GATA6 while YY1 binding is apparent at the non-risk allele (C). Nuclear extracts of Paca44 show a reduced profile of differential protein binding in which we could only confirm the binding of YY1 to the non-risk allele (C). The absent binding at the risk allele and the missing shift of GATA6 and GATA4 suggest no to very low expression of those factors in Paca44 cells.

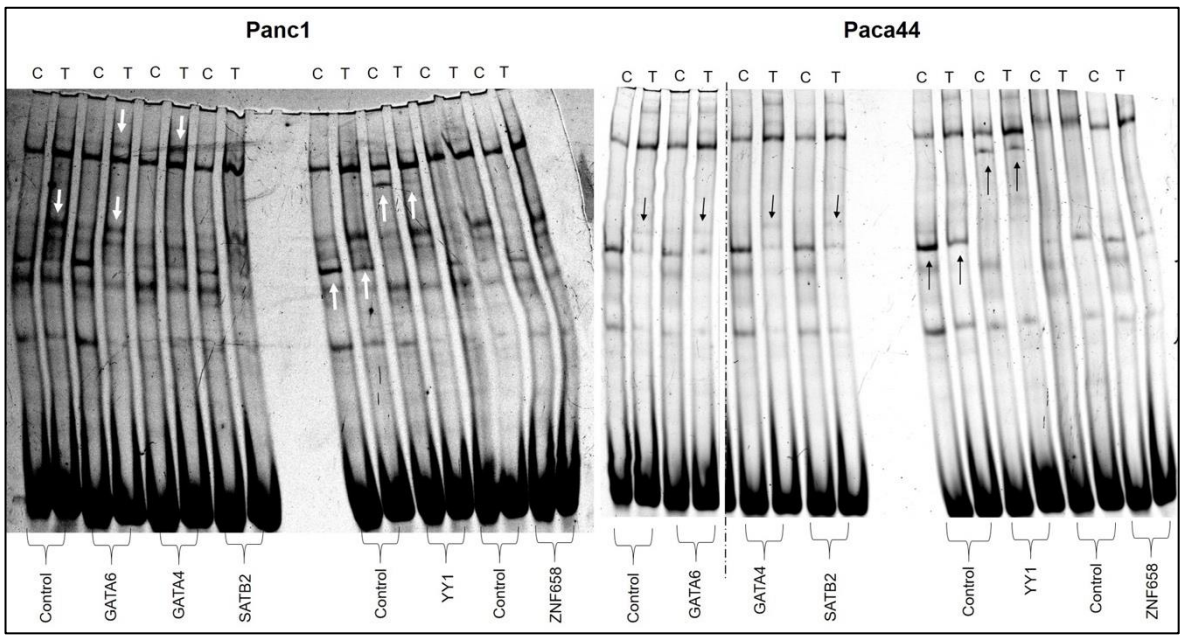


Figure 20 Super-shift EMSA with nuclear protein from Panc1 and Paca44, rs148911734 synthetic oligonucleotides and antibodies against the transcription factors of interest.

Cy5-labelled synthetic oligonucleotides of *rs148911734* were combined with nuclear protein from Panc1 and Paca44 cells. Antibodies directed against GATA6, GATA4 and SATB2 (factors with increased binding affinity to the risk allele) or YY1 and ZNF658 (factors with increased binding affinity to the non-risk allele) were added to the reaction. Unspecific IgG antibody from the same species was used for the control reaction of each antibody, respectively. Nuclear protein was used from three independent nuclear extract preparations (n=3) of which one representative is shown. The binding reaction with the non-risk (C) allele was applied first, followed by the risk allele (T). White arrows highlight the differential binding and related shifts in Panc1. Black arrows highlight the relevant observations in Paca44 cells. Nuclear extract: 3 µg, polydIdC: 350 ng, Cy5-oligonucleotide: 1 ng, EMSA: Electron Mobility Shift Assay, rs148911734C>T. GATA: GATA binding protein, SATB2: Special AT-rich sequence binding protein 2, YY1: Ying Yang 1, ZNF658: Zinc finger protein 658.

Overexpression of GATA4 but not GATA6 in Paca44 reconstitutes differential binding pattern at *rs148911734*. The missing binding at the *rs148911734* risk allele of Paca44 nuclear proteins prompted us to reconstitute the expression of GATA6 and GATA4 by exogenous transfection. We overexpressed *GATA6* and *GATA4* in Paca44 cells either as single factors or in combination with *RBPJL*, *PTF1a* and *HNF1a* and repeated the EMSA experiments (Figure 21 A). We observed an increased binding to the non-risk allele (T) after overexpression of *GATA4* when transfected alone or in combination. Binding of *GATA6,* in contrast, remained absent. Since missing protein expression was no longer a plausible explanation for the lack of GATA6 binding, we looked for other causes. Interestingly, the size of overexpressed GATA6 in the Western blots differed significantly from the expected

size (Figure 21 B). We therefore suspected a protein modification that prevents GATA6 from binding to the DNA fragments in EMSA. If this proves to be true, this would be a reasonable explanation for the lack of inhibitory effects in previous GATA6 overexpression experiments (Sanger sequencing, reporter gene assay).

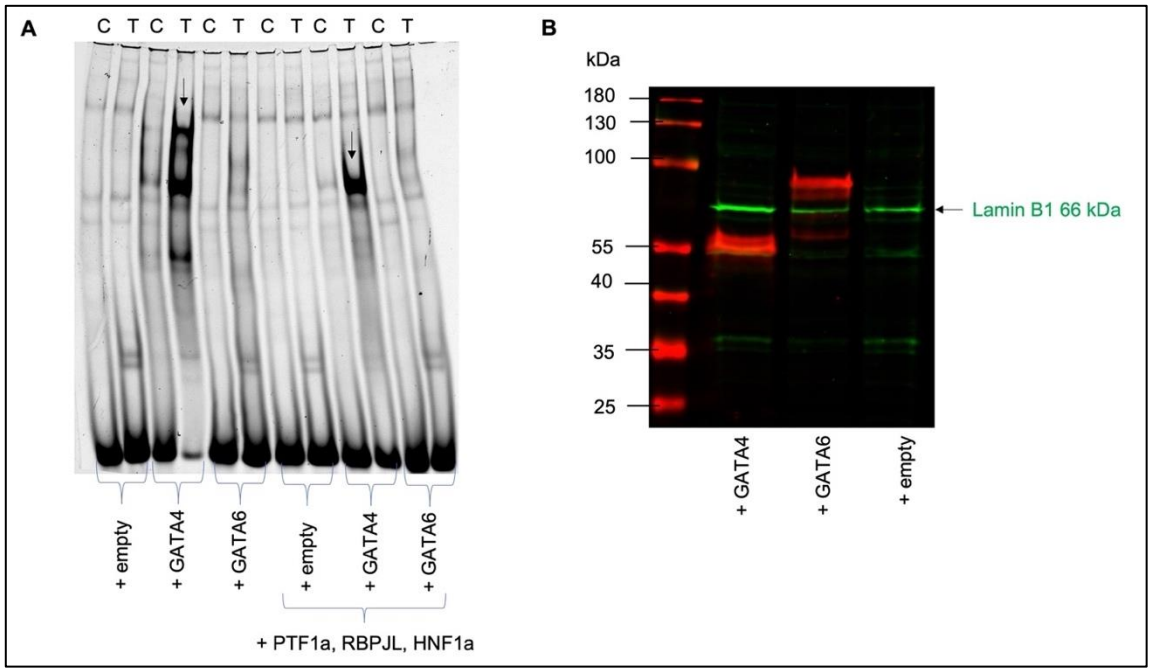


Figure 21 Overexpression of GATA6 and GATA4 in Paca44 displayed in EMSA and Western blot.

(A) Nuclear Protein binding of Paca44 nuclear extracts after transcription factors overexpression. EMSA experiments were performed with nuclear extracts of Paca44 and Cy5-labelled synthetic oligonucleotides harboring the non-risk (C) or risk (T) allele of rs148911734. Paca44 cells were transfected with GATA6 and GATA4 either as single factors or in combination with PTF1a, RBPJL and HNF1a. Transfection with the pcDNA3.1 vector without insert (empty) was used as baseline control. Nuclear extracts were prepared from three independent experiments (n=3, representative shown). The binding reaction with the non-risk allele was applied first, followed by the risk allele. Nuclear extract: 5µg, polydIdC: 350 ng, Cy5-oligonucleotide: 1 ng, EMSA: Electron Mobility Shift Assay; rs148911734C>T. (B) Western blot for nuclear extracts after GATA6 and GATA4 overexpression. Transfection with the pcDNA3.1 vector without insert (empty) was used as baseline control. Overexpressed transcription factors were detected by an antibody directed against the C-terminal FLAG-tag (in red: GATA6: 60 kDa; GATA4: 44.5 kDa, FLAG-tag + 1 kDa). Lamin B1 (green, 60 kDa) was used as internal loading control. 15 µg of nuclear extract were used. GATA: GATA binding protein. The experiments were conducted by Johanna Wallner as part of her Master thesis.

4.5.3 Western blot of GATA6 and GATA4 in different cell lines

The molecular size of GATA6 but not GATA4 differs depending on the cell line and the endo- or exogenous expression. The expected molecular weight of GATA6 and GATA4 were 60 kDa and 44.5 kDa, respectively. The C-terminal FLAG-

tag of the overexpressed proteins accounted for an additional 1 kDa. Unexpectedly, GATA6 showed a strong signal at 90 kDa and a weak signal at 60 kDa. The signal for GATA4 was at 55 kDa. Samples transfected with the pc.DNA3.1 vector without insert (empty) led to no signal, as expected (Figure 21 B). We therefore investigated whether the high molecular weight is a consequence of exogenous protein expression or whether it is also inherent to endogenously expressed GATA6 and GATA4. To this end, we analyzed untreated nuclear extracts of Paca44, Panc1, Caco2, HepG2, HEK293 and HEK293T cells with anti-GATA6 and anti-GATA4 antibodies (Figure 22) as wells as nuclear extracts of the same cell lines after overexpression of *GATA6* and *GATA4* with the anti-FLAG antibody (Figure 23).

We detected strong endogenous GATA6 expression in Caco2 and to a much lesser extent in HEK293 and HEK293T cells. In Caco2, the signal at 60 kDa was about three times as strong as the signal at 90 kDa. At an overall lower signal intensity, we observed the same in HEK293 and HEK293T cells. Those results comply with the observations we made in the previous EMSA experiments. GATA6 was identified from nuclear extracts of Caco2 cells and binding of GATA6 could not be reproduced in EMSA experiments with nuclear extracts from Paca44 and Panc1 cells. Similarly, endogenous expression of GATA4 was most strongly detected in HepG2 cells. A faint signal was also visible in HEK293 and HEK293T and to an even lesser extent in Paca44. In all cell lines the signal appeared only at the corresponding size of 55 kDa. It remains enigmatic, though, why we did not observe any signal in Panc1 since super-shift EMSAs effectively demonstrated GATA4 binding in those cells. In contrast, GATA4 binding from nuclear extracts of Paca44 could not be proven in the super-shift EMSA.

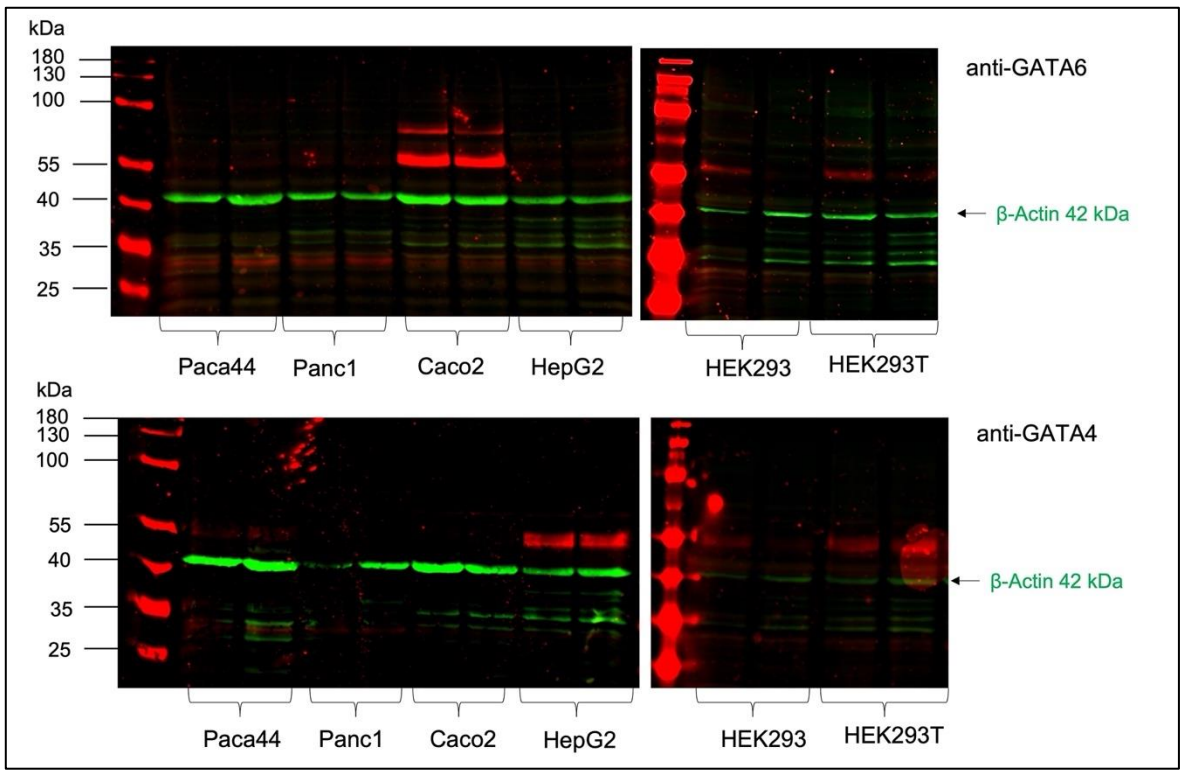


Figure 22 Western blot of nuclear extracts from different cell lines with anti-GATA6 or anti-GATA4 antibodies.

Nuclear extracts of Paca44, Panc1, Caco2, HepG2. HEK293 and HEK293T were prepared from two independent batches (n=2). Antibodies against GATA6 and GATA4 (red) were used to detect the endogenous expression of the protein (expected size GATA6: 60 kDa, GATA4: 44.5 kDa). β-Actin (green, 42 kDa) was used as internal loading control. All cells: 50 µg nuclear extract. The experiments were conducted by Johanna Wallner as part of her Master thesis.

When GATA6 was overexpressed in the same cell lines, the picture turned, and the signal intensity was always highest at the 90 kDa and a weaker signal appeared at 60 kDa. Panc1, HEK293 and HEK293T cells showed the strongest expression, while the expression in HepG2 and Paca44 was less intense. Interestingly, Paca44 exclusively showed the signal for the 90 kDa sized form. No signal was obtained in Caco2 cells. A similar picture emerged with GATA4 overexpression. Panc1, HEK293 and HEK293T cells showed the strongest band intensities. Compared to the signal obtained from the endogenous samples, we observed an additional, even yet weaker band at 45 kDa, which was consistent with the expected GATA4 size of 44.5 kDa. The signals in Paca44 cells were similar but overall weaker. Overexpression of GATA4 in Caco2 and HepG2 led to a very weak signal at 55 kDa in Caco2 and barely to any signal in HepG2 cells (Figure 23).

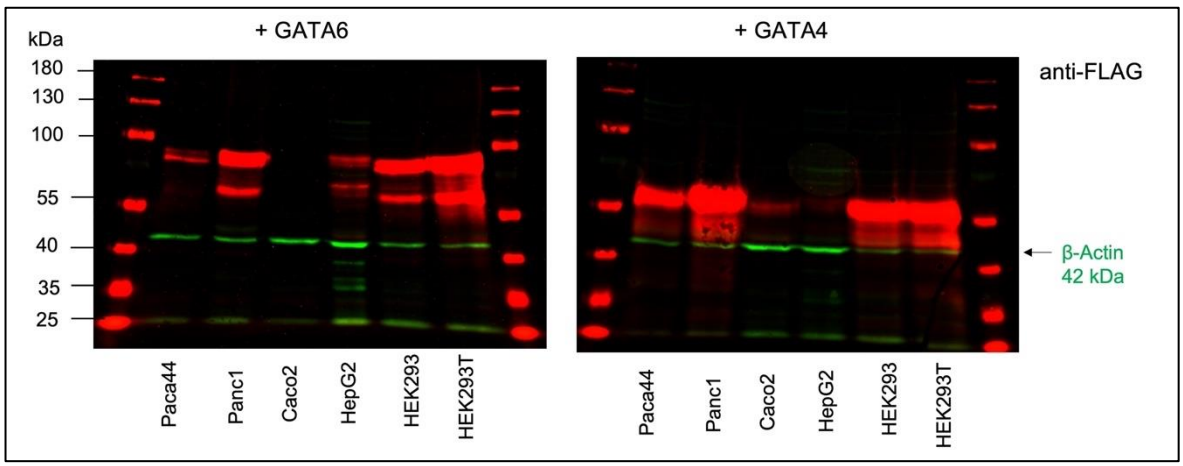


Figure 23 Western blot of nuclear extracts from different cell lines after GATA6 and GATA4 overexpression.

GATA6 and GATA4 were overexpressed in Paca44, Panc1, Caco2, HepG2, HEK293 and HEK293T and nuclear extracts were prepared. A representative of two independent experiments (n=2) is shown. An antibody directed against the C-terminal FLAG-tag (red) was used to detect the overexpressed protein (expected size GATA6: 60 kDa, GATA4: 44.5 kDa, FLAG-tag + 1 kDa). β-Actin (green, 42 kDa) was used as internal loading control. For all cells15 µg nuclear extract were used. GATA: GATA binding protein. The experiments were conducted by Johanna Wallner as part of her Master thesis.

Taken together, endogenous expression of GATA6 and GATA4 is most dominant in Caco2 and HepG2 cells, respectively. GATA6 appears at two molecular sizes, namely the expected 60 kDa and an unexpected 90 kDa. The lower molecular weight form is the predominant form in nuclear extracts of untreated Caco2 cells. GATA4 from untreated HepG2 nuclear extracts only presents at an unexpected higher molecular size of 55 kDa. The exogenous overexpression of GATA6 inverses the signal intensities of the 60 kDa and 90 kDa in favor of a stronger signal for the higher molecular size. The overexpression of GATA4 leads to the appearance of the originally expected 44.5 kDa band.

In the context of the missing binding in EMSA experiments with GATA6 overexpressed protein from Paca44 cells, the results suggest a loss of GATA6 DNA binding ability of the form with higher molecular weight, which after overexpression was the only GATA6 form in Paca44. In contrast, GATA6 from endogenous expression of Caco2 predominantly appeared at the lower molecular size and showed DNA binding ability as judged from super-shift EMSA. Interestingly, GATA4 DNA binding ability seemed independent of its molecular size. We observed DNA

binding in EMSA experiments with GATA4 from endogenous expression of HepG2 cells (55 kDa) and after overexpression in Paca44 cells (44.5 kDa).

We analyzed GATA6 and GATA4 in human and mouse pancreatic extracts to find out which molecular weights are physiologically predominant (Figure 24). Unfortunately, we could not draw any conclusions: While the human tissue sample did not show any signal at all, the mouse sample quality was impaired by protein degradation.

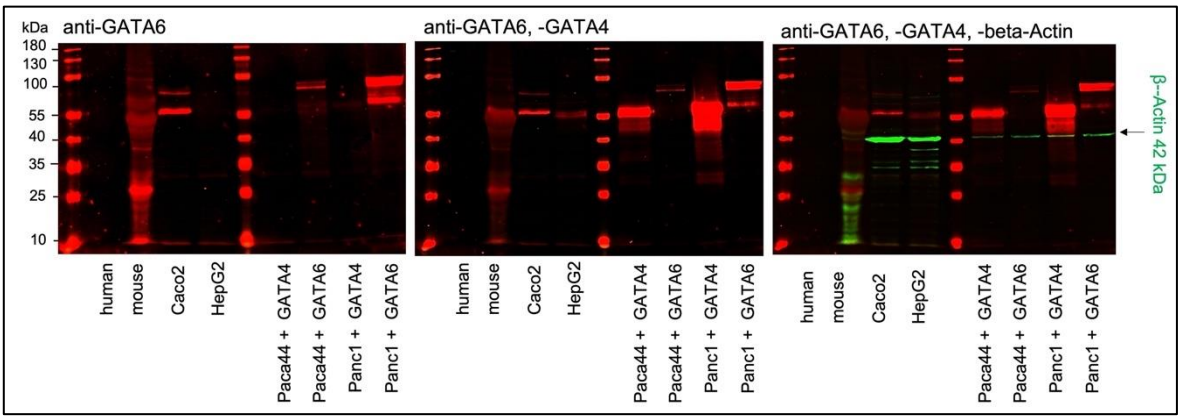


Figure 24 Western blot of whole human and mouse pancreas lysates incubated with anti-GATA6 and anti-GATA4 antibodies.

Protein extracts from human (n=1) and mouse (best out of n=4) were analyzed. Nuclear extracts from untreated Caco2 and HepG2 were used as controls for the endogenous expression of GATA6 (60 kDa) and GATA4 (55 kDa), respectively. Nuclear extracts from Panc1 and Paca44 after GATA6 (90 kDa) and GATA4 (44.5 kDa) overexpression were included as controls for exogenous protein expression. Antibodies against GATA6, GATA4 and β-Actin were applied sequentially. The blot was scanned after the addition of each antibody (in red: GATA6, GATA4). β-Actin (42 kDa, green) was used as internal loading control. For Human/mouse and Caco2/HepG2 50 µg total protein or nuclear extract were used, respectively. For Paca44/Panc110 µg nuclear extract were used. GATA: GATA binding protein. The experiments were conducted by Johanna Wallner as part of her Master thesis.

4.6 Characterization of GATA6-p.K484R variant

To explain the deviation of the high-molecular weight band of GATA6 in Western blots with overexpressed protein (90 kDa) compared to the expected band of endogenously expressed GATA6 (60 kDa) we looked for potential post-translational modification. Human GATA6 possesses 14 sites for phosphorylation, 1 site for methylation and 3 reported sites for SUMOylation (p.K484, p.K429, p.K473; Hendriks *et al.*, 2017). SUMO proteins (SUMO1-4) have an estimated molecular

size of 10-12 kDa. SUMOylation therefore caught our particular interest since this is reported for GATA factors (Chen *et al.,* 2020) and the high molecular weight of GATA6 and GATA4 in our Western blots differed about 10-30 kDa from the expected size. Based on the *JASSA* database, SUMOylation at amino acid K484 appeared to be the most likely modification of GATA6.

Consequently, we introduced the p.K484R mutation into the pcDNA3.1 GATA6 wild type construct to remove potential SUMO residues at this site. The subsequent analysis indented to investigated whether the p.K484R variant would display an enhanced activity compared to its wild type counterpart. To this end, we overexpressed GATA6 and the p.K484R variant either as single factors or in combination with *PTF1a*, *RBPJL*, *HNF1a* and the empty vector construct (pcDNA3.1 vector without insert) in Paca44 and Panc1 cells.

4.6.1 Western blot and EMSA after overexpression

We prepared total protein lysates and nuclear extracts form Paca44 and Panc1 after overexpression of GATA6 and mutated GATA6 (p.K484R) and analyzed the molecular size and DNA binding ability at *rs148911734* by Western blot (Figure 25) and EMSA (Figure 26), respectively. Other than expected no difference in protein size of wild type or p.K484R GATA6 was detectable in Paca44 or Panc1. In Paca44 cells, the bands were equally strong for both overexpressed forms. In Panc1, we observed a tendency towards a stronger signal of GATA6 wild type compared the K484R variant. In line with our findings from the Western blot, DNA binding in EMSA experiments did not differ depending on the GATA wild type or the p.K484R overexpressed protein. Both proteins showed an increased binding at the risk allele as we observed in previous EMSA experiments. The DNA binding with Paca44 nuclear extracts was overall very weak and no further conclusion should be drawn from these experiments. Nuclear protein binding from Panc1 extracts was stronger and confirmed the trend we observed in the Paca44 experiment. It should be of note, that compared to Paca44, Panc1 have some endogenous GATA6 expression and already showed a differential binding patter in previous experiments. It is possible, that the observed binding is a consequence of the endogenous rather than the exogenous GATA6 expression.

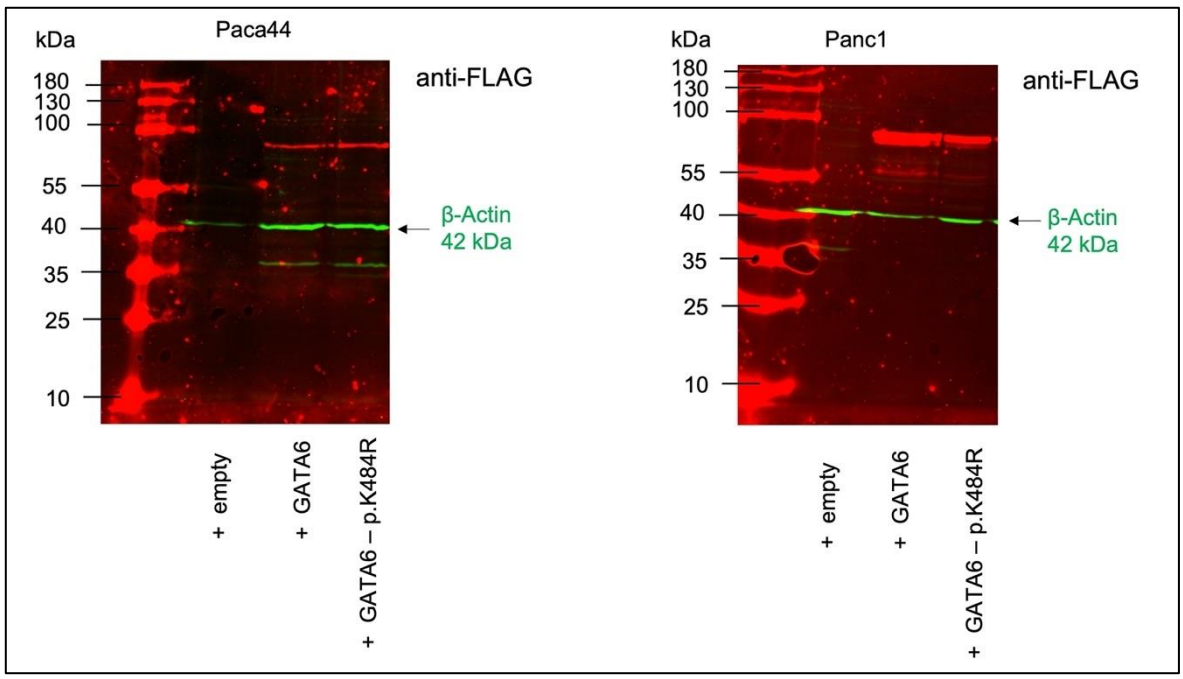


Figure 25 Western blot of GATA6 and GATA6-p.K484R from total lysate.

GATA6 and GATA6-p.K484R were overexpressed in Paca44 and Panc1 and total lysates were prepared. Transfection with the pcDNA3.1 vectors without insert (empty) served as control. A representative of two independent experiments (n=2) is shown. An antibody directed against the C-terminal FLAG-tag (red) was used to detect the overexpressed protein (expected size GATA6: 60 kDa, FLAG-tag + 1 kDa). β-Actin (green, 42 kDa) was used as internal loading control. For Panc1 10 µg total lysate and for Paca44 20 µg total lysate were used. GATA: GATA binding protein.

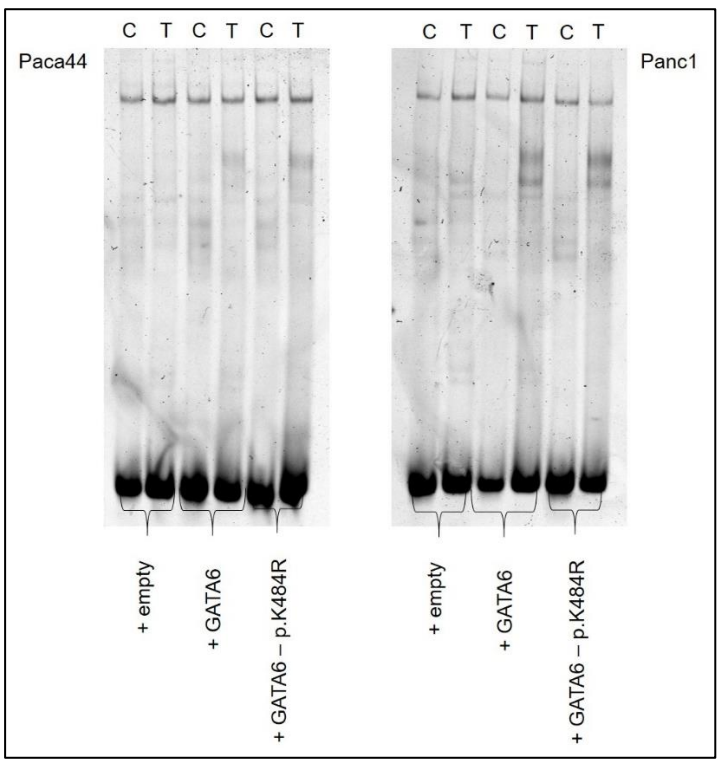


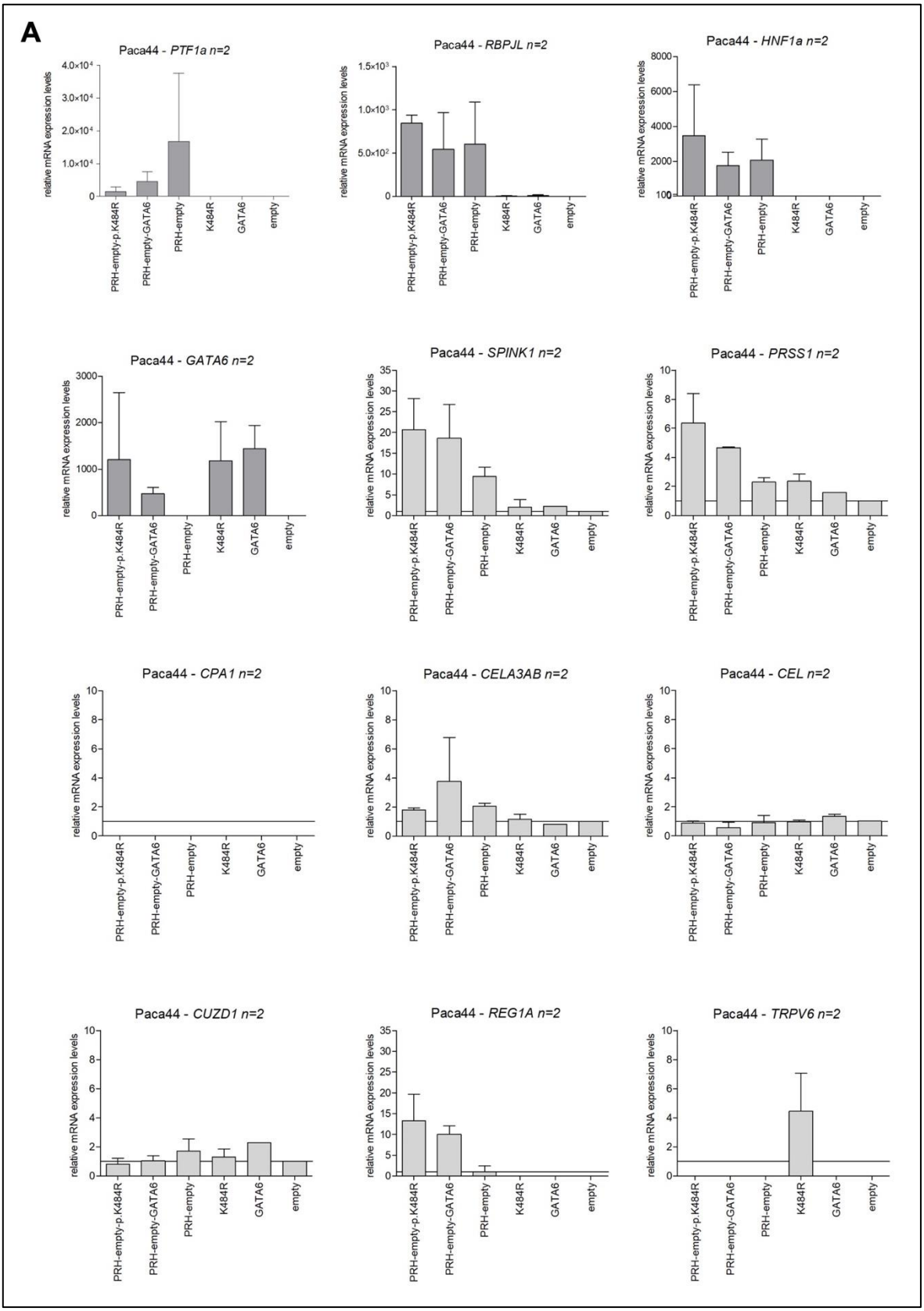
Figure 26 Nuclear protein binding of GATA6 and GATA6-p.K484R at rs148911734.

GATA6 and GATA6-p.K484R were overexpressed in Paca44 and Panc1.Transfection with the pcDNA3.1 vectors without insert (empty) served as control. Nuclear extracts were prepared from two independent experiments (n=2, representative is shown). The binding reaction with the non-risk allele was applied first, followed by the risk allele. Nuclear extract: 6µg, polydIdC: 350 ng, Cy5-oligonucleotide: 1 ng, EMSA: Electron Mobility Shift Assay; *rs148911734*C>T*.* GATA: GATA binding protein

4.6.2 Activation of target genes

GATA6-p.K484R shows a tendency towards increased target gene activation compared to the wild type in Panc1 but not in Paca44 cells. When we assessed potential effects of GATA6 wild type and p.K484R transcriptional activity on selected target genes, their effect as single factors was negligible and barely differed between the wild type and the mutant. Interestingly, when GATA6 wild type and p.K484R were expressed in combination with RBPJL, PTF1a and HNF1a, the mutant showed a trend towards stronger target gene activation than the wild type. In Panc1 (Figure 27 B), this trend was consistent across the digestive enzymes (*PRSS1, CPA1, CELA3AB, CEL*) and *CUZD1* as a representative of the zymogen granule proteins. The effect was especially pronounced in *SPINK1,* however, the high standard deviation questions the robustness of this result. It should be of note, that target gene activation was in general higher with either GATA6 wild type or the p.K484R variant in combination with RBPJL, PTF1a and HNF1a than when the latter three factors were expressed on their own.

In Paca44 (Figure 27 A), the picture of target gene expression was inconsistent. While *PRSS1* followed the trend of a higher activation with the p.K484R variant, yet at much lower levels than in Panc1, *CELA3AB* showed the opposite trend. *CPA1* did not show any response in target genes expression at all and the effects on *CEL* and *CUZD1* were neglectable. Surprisingly, *TRPV6* expression was boosted the most upon the single expression of the p.*K484R* variant. The expression levels of *SPINK1*, *PRSS1*and *REG1A* were higher for both GATA6 forms in combination with PTF1a, RBPJL and HNF1a than without them. Overall, *SPINK1* expression levels increased beyond any effects we observed in previous experiments. Taken together, the results suggest no relevant effect of the GATA6 p.K484R variant on target genes activation compared to the wild type. It should be noted that these results are based on only two biological replicates and would benefit from additional replicates to substantiate the observed effects.


A
Paca44 - PTF1a n=2
Paca44 - RBPJL n=2
Paca44 - HNF1a n=2
Paca44 - GATA6 n=2
Paca44 - SPINK1 n=2
Paca44 - PRSS1 n=2
Paca44 - CPA1 n=2
Paca44 - CELA3AB n=2
Paca44 - CEL n=2
Paca44 - CUZD1 n=2
Paca44 - REG1A n=2
Paca44 - TRPV6 n=2
relative mRNA expression levels
PRH-empty-p.K484R
PRH-empty-GATA6
PRH-empty
K484R
GATA6
empty

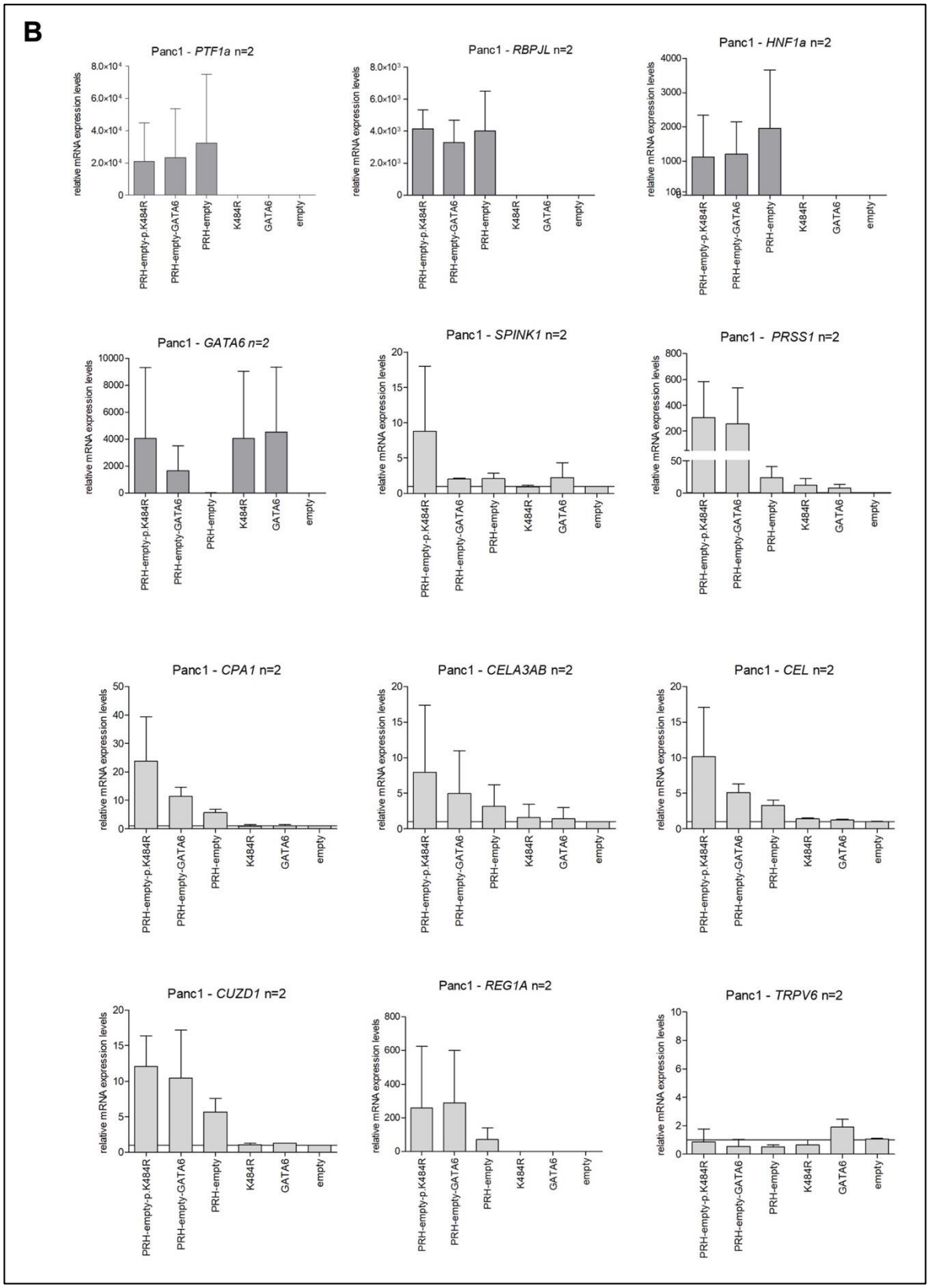


Figure 27 Comparison of relative mRNA expression in Paca44 (A) and Panc1 (B) cells after transfection with GATA6 wild type and GATA6-p.K484R.

Cells were transfected with the GATA6 p.K484R or the wild type either in co-transfection with PTF1a, RBPJL and HNF1a (PRH) or as single factors. The pcDNA3.1 vector with no insert (empty) was used as transfection control and to fill up the total vector amount to keep

the co-expressed factors at an equal fraction throughout the different conditions. Total RNA was collected 72 h post transfection and transcribed to cDNA. Relative mRNA expression of the indicated target genes was measured by qPCR with TaqMan assays. Experiments were carried out in two independent biological replicates (n=2), TaqMan assays were performed in technical triplicates. The horizontal line illustrates the baseline relative mRNA expression (=1). Bars in dark grey indicate the measurement of a transfected factor. Standard deviation is shown. For full gene names please refer to legend of Figure 14.

5 Discussion

Chronic Pancreatitis (CP) is relapsing or continuing inflammatory disease of the pancreas (Witt & Becker, 2002). In the year 2000, Witt *et al*. identified *SPINK1* p.N34S as the strongest and clinically most relevant genetic risk factor for idiopathic CP (ICP) (Witt *et al*., 2000). The genetic association (Pfützer *et al*., 2000, Chandak *et al*., 2002, Truninger *et al.*, 2002, Bhatia *et al*., 2002) and physiological relevance of *SPINK1 (*Pubols *et al*., 1974**,** Ohmuraya *et al*., 2006**,** Nathan *et al*., 2005**,** Romac *et al*., 2010**)** was elaborated in several studies. The biochemical and cellular consequences of the *SPINK1* p.N34S haplotype, however, remain enigmatic (Kuwata *et al*., 2001**,** Kuwata *et al*., 2002**,** Király *et al*., 2007**)**. After altered inhibitor function was excluded, the focus shifted onto *SPINK1* p.N34S linked non-coding variants. But none of the linked intronic variants c.56-37T>C (intron 1), c.87+268A>G (intron 2), c.195-606G>A and c.195-66_65insTTTT (intron 3) showed a functional defect (Kereszturi *et al*., 2009). Preliminary research in our group narrowed *SPINK1* p.N34S linked regulatory variants down to three final candidates. While functional analysis of one candidate, *rs17107316* (c.87+268T>C) was already published with negative results on *SPINK1* expression, the other two candidates, *rs148276928* (c.1-2090T>C) and *rs148911734* (c.1-7321C>T), were not investigated so far. We built on the hypothesis that those variants might affect transcription factor binding and thereby diminish *SPINK1* expression. Decreased levels of SPINK1 could shift the fine balance between proteolytic and inhibitory activity within the pancreas through increased trypsin activity and contribute to CP initiation and progression.

5.1 *SPINK1* p.N34S haplotype decreases the expression of the risk allele

Assessment of *SPINK1* mRNA expression levels from the *SPINK1* p.N34S heterozygous pancreatic cancer cell lines Paca44 and PancTu-I revealed an up to 6.7-fold decreased expression of the variant allele. This observation fostered the assumption that *SPINK1* p.N34S linked variants located in non-coding regions suppress its expression (Kereszturi *et al*., 2017**)**. We were able to confirm the previous observation in Paca44 cells when we analysed their mRNA expression by Sanger sequencing. The mRNA samples showed a decreased signal intensity in the electropherograms for the risk allele compared to the non-risk allele, while the

gDNA samples displayed equally strong signals for both alleles. Further, we analysed the allelic *SPINK1* mRNA expression in intestinal biopsies of five *SPINK1* p.N34S heterozygous (TC) and five wild type (TT) subjects by Sanger sequencing and RNA-sequencing. Indeed, the heterozygous carriers showed decreased expression of the risk alle. We thus confirmed the inhibitory effect of *SPINK1* p.N34S haplotype on the risk allele expression in Paca44 cells and additionally in primary human samples and strengthened the role of regulatory variants as potential disease cause.

5.2 Not *rs142703147* but *rs148911734* as the underlying pathological variant

Based on *in-silico* and *in-vitro* investigations, Boulling *et al* proposed a functional relevance of *rs142703147C>A* (c.1-4141G>T). The variant was in full linkage with *SPINK1* p.N34S in a population of 548 French ICP patients and 562 matched controls (r^2=1). When we performed LD acquisition in the European population (*SNiPa* 1,000 Genomes Phase3v5, EUR), *rs142703147* was also found to be in full linkage with p.N34S (r^2=1). However, once we extended our analysis to the South Asian population (*SNiPa*,1,000 Genomes Phase3v5, SAS), the high association was lost (r^2=0.57). Sanger sequencing in a cohort of TCP patients of Indian ancestry, German patients and matching controls further argued against *rs142703147* as a relevant candidate.

In a next step, we estimated haplotype frequencies of genotype data from 253 cases and 256 controls of Indian ancestry for *rs17107315T>C* (p.N34S), *rs142703147C>A* and *rs148911734C>T*. In line with our expectations, haplotype blocks exclusively containing the non-risk alleles (TCC) were clearly enriched in controls. In contrast, the haplotype blocks CAT (all risk alleles) and CCT (risk allele: *rs17107315*, *rs148911734)* significantly associated with CP. Importantly, haplotype blocks containing the *rs142703147* risk allele, but not the *rs148911734* risk allele (CAC and TAC) were not associated with CP. Those findings reinforce the assumption that *rs148911734* is the potential pathological variant of the CP associated *SPINK1* p.N34S haplotype while weakening the relevance of *rs142703147.*

Moreover, we took a closer look into the supplementary genetic data of the Indian population investigated by Boulling *et al*. (Table 32). The heterozygous allele state

of both *rs142703147* (CA) and *rs17107315* (CT) was significantly enriched in the patient group (OR=14.4, *P*<0.001, CI=5.2-40.2). Nonetheless, it became evident that the proclaimed pathological relevance of *rs142703147* is highly questionable. The association of a heterozygous allele state of *rs17107315* (CT) in combination with the wild type *rs142703147* (CC) was highly significant (OR=6.6, *P*<0.001, CI=2.31-18.9). Vice versa, wild type *rs17107315* (TT) in combination with *rs142703147* (CA) heterozygosity lost all significance (OR=1.53, *P*=0.73, CI=0.28-8.4). If both variants were to exert an effect on the disease risk, the association of *rs142703147* and *rs17107315* (p.N34S) should be in line with one another. The present data, however, strongly suggest a *rs142703147*- independent effect on the disease risk.

Table 32 Allele status in Indian population investigated by Boulling et al. (2017)

rs142703147 *C>A*	*rs17107315* T>C	Patients	Controls	P-value	OR	95% CI
CC	TT	228/347	254/264	0.000	0.08	0.04-0.15
CC	**TC**	**32/347**	**4/264**	**0.000**	**6.6**	**2.31-18.9**
CC	CC	3/347	0/264	0.26	Inf	NA
CA	**TT**	**4/347**	**2/264**	**0.73**	**1.53**	**0.28-8.4**
CA	TC	63/347	4/264	0.000	14.4	5.2-40.2
AA	TT	2/347	0/264	0.51	Inf	NA
AA	CC	15/347	0/264	0.000	Inf	NA

The table was composed by Prof. Dr. Heiko Witt out of the data provided in the supplement of Boulling *et al*., 2017. OR=Odds Ratio, CI=Confidence Interval, Inf=infinite, P-value=0.000 (<0.001), NA=not applicable

Even though genetic analysis made *rs142703147* unlikely, experimental evidence was lacking. So, we included the variant for further functional analysis. We conducted EMSA experiments with nuclear extracts of Caco2, HepG2, Huh7 and RT4 cells, but none of the cell lines showed a differential binding pattern depending on the non-risk or risk allele. Affinity chromatography with nuclear extracts of Caco2 and HepG2 cells, proteomics analysis and stringent data filtering, however, revealed three potential repressors (ENO1, UBP1 TFCP2; FC <1) which were replicated in both cell lines. ENO1, which showed the strongest FC, was only annotated as cofactor but not as transcription factor and had no predicted effect. Although UBP1

and TFCP2 were annotated as transcription factors, their predicted effect was not available or predicted to be weak. In contrast, the results we obtained for the *rs148911734* were robust. Not only showed the FC a stronger effect size, ZNF658, YY1, SATB2 and GATA6 were all annotated transcription factors and predicted to have a strong effect at this site. Importantly, GATA6 and YY1, had a strong transcription factor binding site (TFBS)-disruption prediction, which strengthened their selection as promising candidates. All these findings led us to focus on *rs148911734* and the transcription factors that bound to it in the subsequent functional studies.

On a functional level, Boulling *et al.* suggested the disruption of a putative HNF1A-PTF1L binding site though the *rs142703147C>A* transversion. This transversion would alter the E-Box recognition site of the bipartite binding sequence (E-Box and TC-Box) to which the PTF1-L complex binds. In EMSA experiments with nuclear extracts from HEK293T with the transiently expressed PTF1-L complex, the authors indeed demonstrated the disruption of PTF1-L binding to the risk (A) allele (Boulling *et al*., 2017). We undertook a similar experimental approach with nuclear extracts from Panc1 cells after the overexpression of *E47*, *PTF1a*, *RBPJL* and *HNF1a* either in combination or as single factors. Except for those nuclear extracts, that contained E47, none of the conditions led to a differential binding pattern at *rs142703147*. We were therefore unable to reconstitute the previous observations made by Boulling *et al.* in our model.

The biological relevance of the increased binding of E47 to the non-risk allele in our experiments is questionable. First, we did not confirm E47 by super-shift EMSA and the observed band could also stem from an E47-induced factor rather than from E47 itself. Moreover, E47 is a basic helix-loop-helix (bHLH) class A transcription factor which is widely expressed in various tissues. Nonetheless, in none of the EMSA experiments with nuclear extracts from Caco2 (intestinal), HepG2 and Huh7 (hepatic) and RT4 (renal) cells, a differential binding patterns was detectable at the site of interest. While Boulling *et al.* did not include E47 in their experiments at all, we observed some redundancy of the factor in our overexpression experiments. Precisely, the induction of *SPINK1* expression was independent of E47 as long as PTF1a, RBPJL and HNF1a were present.

5.3 PTF1a, RBPJL and HNF1a drive *SPINK1* expression independent of *rs148911734*

The synergistic action of HNF1A and PTF1-L on *SPINK1* expression was effectively shown in *SPINK1* promoter transactivation assays in HEK293T cells. *SPINK1* luciferase reporter gene activity was increased by 6-fold upon the combined overexpression of HNF1a and PTF1-L compared to the empty vector transfected cells (Boulling *et al*., 2017**)**. The combined overexpression of E47, PTF1a, RBPJL and HNF1a in our experiments led to a successful induction of *SPINK1* mRNA expression in Panc1 (6-fold) and in Paca44 (2-fold). For both cell lines, the *SPINK1* induction was equally effective in the presence or absence of E47. Sufficiently high endogenous levels of E47 are unlikely. Kim *et al*. engineered pancreatic ductal adenocarcinoma (PDA) cells, among them Panc1, to express an inducible form of E47. Their model was based on previous observations in which acinar cells underwent acinar to ductal metaplasia after the loss of bHLH signaling. The loss of bHLH signaling was attributed to complexing with ID3 (inhibitor of differentiation 3). The complex formation of ID3 and E47 would remove E47 as dimerization partner for other bHLH transcription factors and restrain E47 from DNA binding. Through inducible E47 overexpression, E47 partially restored the acinar phenotype by the induction of digestive enzyme and MIST1 and diminished tumorigenesis (Kim *et al*., 2015). Consequently, while E47 seems important for the maintenance of acinar cell homeostasis and suppression of tumorigenic potential, it is redundant for *SPINK1* expression. In turn, PTF1a, RBPJL and HNF1a are vital, and our findings are consistent with those of Boulling *et al. (2017).*

After confirming the importance of PTF1a, RBPJL and HNF1a for *SPINK1* regulation, we wanted to assess whether the action of these factors is dependent on *rs148911734.* EMSA experiments with nuclear extracts from Panc1 after the overexpression of the corresponding factors, however, did not reveal any differential binding pattern based on the overexpression of either single factor or their combination. Instead, the same binding patterns showed as observed for nuclear extracts from Caco2 and HepG2. This reinforced our assumption that a pathological contribution of *rs148911734* was due to an altered binding of a different transcription factor, but not because of PTF1a, RBPJL and HNF1a. This EMSA experiment further suggested that the binding of this transcription factor would be independent

of the overexpressed factors (no change in binding pattern after transcription factor overexpression). The expression and assembly of the PTF1 complex is a highly acinar-specific event. Caco2 and HepG2 cells, on the other hand, have an intestinal or hepatic background. Indeed, the pancreas, intestine and liver share high levels of *SPINK1* expression (https://www.proteinatlas.org). Moreover, we effectively demonstrated that *SPINK1* p.N34S carriers show a diminished expression of the risk allele in the intestinal tissue. Conclusively, the effect of the *SPINK1* p.N34S haplotype is not restricted to the pancreas and is likely caused by a tissue-overlapping factor.

5.4 GATA6 and *rs148911734* – so close and yet the enigma continues

After the evaluation of the affinity chromatography results, we decided to focus on *rs148911734* as the favorite candidate SNP of the *SPINK1* p.N34S haplotype. Out of the top candidates, we selected ZNF658 and YY1 as potential activators and GATA6 and SATB2 as potential repressors. Each of the factors was replicated with the same trend in both cell lines (HepG2, Caco2). Moreover, we detected a strong TFBS- disruption predictions from *motifbreakR* for GATA6 and YY1 at this site.

To validate the selection of YY1 and ZNF658 as potential enhancers and GATA6 (Caco2), GATA4 (HepG2) and SATB2 as potential repressors, we conducted super-shift EMSAs. The addition of antibodies against GATA6 (Caco2), GATA4 (HepG2) and YY1 clearly assigned those factors to the differential binding patterns observed previously. The *in vitro* validation of the proteomic analysis prompted us to functionally investigate the direct impact of GATA6, GATA4 and YY1 on *SPINK1* expression.

Besides our experimental findings, GATA6 and GATA4 were especially interesting candidates based on their relevance for pancreatic development and tissue homeostasis (Villamayor *et al*., 2020). During early embryonic development (embryonic day 9.5) in mice, GATA4 and GATA6 have overlapping expression patterns at first, before they become restricted to the exocrine and endocrine compartment, respectively. In the adult pancreas, GATA4 is no longer detectable in the exocrine compartment, but was expressed in pancreatic islet cells. GATA6 was found to play a predominant role in the initial specification of the pancreas and

pancreatic cell type differentiation. Decker *et al*. expressed GATA4-Engrailed and GATA6-Engrailed dominant repressor fusion proteins in mouse embryos. Pancreatic tissue (at embryonic day 17.5) from transgenic mouse embryos expressing a GATA6 repressor did either not form at all or showed abnormal tissue organization. The repression of GATA4, in contrast, did not lead to a pancreatic phenotype (Decker *et al*., 2006). Interestingly, in a second approach Carrasco *et al*. investigated the effects of single and combined *Gata4* and *Gata6* inactivation in the developing mouse pancreas. Mice with a conditional (flox) allele of Gata4 (*Gata4*$^{flox/flox}$) or Gata6 (*Gata6*$^{flox/flox}$) were crossed to a transgenic mouse line that expressed Cre recombinase under the control of the pancreatic and duodenal homeobox gene 1 (*Pdx1*) promoter (*Pdx1-Cre mice*). The inactivation of either Gata4 (*Gata4*$^{flox/flox}$*; Pdx1-Cre*) or Gata6 (*Gata6*$^{flox/flox}$*; Pdx1-Cre*) on their own did not affect the pancreas formation. In contrast double Gata4/Gata6 mutant mice (*Gata4*$^{flox/flox}$*; Gata6*$^{flox/flox}$*; Pdx1-Cre*) did not develop any pancreas at all and died shortly after birth. The individual impact of GATA4 and GATA6 was further assessed when the expression of one allele of Gata4 or Gata6 was reconstructed on the conditional knock out of *Gata6* and *Gata4*, respectively. When *Gata6* was expressed from only one allele along with a full *Gata4* knock down (*Gata4*$^{flox/flox}$*; Gata6*$^{flox/+}$*; Pdx1-Cre*), the pancreas mass was severely reduced. When *Gata4* was expressed from one allele along with a full *Gata6* knock down (*Gata4*$^{flox/+}$*; Gata6*$^{flox/flox}$*; Pdx1-Cre)*, normal pancreas development was possible (Carrasco *et al*., 2012). Xuan *et al*. reported similar findings in which either loss of GATA4 or GATA6 resulted in only mild pancreatic defects, while the simultaneous loss caused severe pancreatic agenesis (Xuan *et al*., 2012). While the study by Decker *et al*. highlights the importance of GATA6 during mouse embryonic development, the study of Carrasco *et al*. assigns GATA4 the dominant role. Both studies, supported by the findings of Xuan *et al*., however, effectively demonstrated that the spatial and temporal expression of GATA4 and GATA6 is crucial for proper pancreatic development and that they complement each other rather than being mutually exclusive.

The relevance of GATA6 in the human context was highlighted after mutations in *GATA6* were identified as the most common cause of human pancreatic agenesis. In a study by Allen *et al.*, 56% (15/27) of the affected individuals carried a

heterozygous mutation in *GATA6* while none of the mutations was identified in the control population of 1,094 subjects of the 1,000 Genomes Project data base. In contrast to homozygous mutations in *PDX1* and *PTF1A,* which cause pancreatic agenesis through a complete loss of the protein, *GATA6* haploinsufficiency leads to the same phenotype. The functional investigation of selected *GATA6* mutations (p.R456C, p.N466D, p.A467T, p.K473Q) revealed a diminished promoter activity in reporter gene assays compared to the wild type while the expression levels were unchanged. Consequently, the deleterious effect of *GATA6* mutations appears to be lost protein function that affects DNA binding and promoter activation (Allen *et al*., 2011). Mouse studies revealed that GATA6 is especially central for cell homeostasis in the adult exocrine department. This effect was partially explained through a direct regulation of the acinar-specific transcription factors PTF1a, RBPJL and MIST1. In *Gata6*$^{P-/-}$ mice, in which pancreatic *Gata6* was deleted at embryonic day 10.5 using a *Ptf1a*-driven Cre recombinase, qPCR mRNA expression levels for those factors compared to the expression levels in control animals were significantly lower. Likewise, downstream targets of the PTF1 complex such as *Amy2a*, *CtrB*, *Cpa*, *Cel* and *Pnlip* were significantly reduced. ChiP experiments further confirmed the binding of GATA6 to the promoter of *Rbpjl*, *Mist1* and *Pnlip*. Moreover, GATA6 inactivation caused serious acinar tissue remodeling. A massive loss of acinar cells was accompanied by increased markers for cell apoptosis, fat accumulation, and acinar to ductal metaplasia. Increased cell proliferation was evident by significantly increased staining of Ki67-positive cells and was considered as a potential compensating mechanism (Martinelli *et al*., 2013).

Despite the promising findings from our proteomic analysis and the strongly suggestive evidence from literature, neither of our attempts to functionally pin down the expected enhancing effect of YY1 and the repressive effect of GATA6 and GATA4 on *SPINK1* expression succeeded. siRNA knock down of the factors in Caco2 and HepG2 aimed to assess their global effect on total *SPINK1* expression. While the missing effect on a decrease in *SPNK1* expression upon YY1 knock down remains unresolved, the missing effect on an increase in *SPINK1* expression after the knock down of GATA6 and GATA4 can be explained by the inherent characteristics of the two cell models Caco2 and HepG2. Both cell lines are homozygous for the non-risk allele, so they lack the corresponding binding site.

Moreover, *SPINK1* expression was high in those cells. An inhibitory effect of GATA6 and GATA4, which causes an increased *SPINK1* expression when removed, is therefore unlikely.

Based on the missing effect on total *SPINK1* expression and the difficulties that came with the inherent genotype and the high *SPINK1* expression of Caco2 and HepG2, we focused our investigations on allelic effects. At first, reporter gene assays appeared as a suitable way to assess the transcription factor binding on the non-risk and risk allele independent of the genotype of the cell line. Using firefly luciferase expression vectors under the control of the *SPINK1* promoter and a 500 bp nucleotide fragment framing either the non-risk (C) or risk allele (T) of *rs148911734*, we aimed to reconstitute the two genotypes in Caco2 and HepG2 and to elucidate the contribution of GATA6, GATA4 and YY1 on the *SPINK1* promoter activity within this context. However, in both cell lines, the knock down of GATA6, GATA4 and YY1 remained without effect on the luciferase expression of the risk and non-risk allele, respectively. When we tried to assess the transcription factor effects on the allelic *SPINK1* expression by knock down or overexpression in the *SPINK1* p.N34S heterozygous cell line Paca44, the result also remained inconclusive. We are aware that the models used to investigate GATA6, GATA4 and YY1functionally for *SPINK1* expression create a highly artificial setting. It remains enigmatic, though, why we could not observe any reasonable trend in the context of the previous findings.

5.5 The lacking effects of GATA6 – a matter of model and modification?

Our functional investigation of the effects of GATA6 and GATA4 on *SPINK1* expression had reached an impasse. First, siRNA knock down of endogenously expressed GATA6 and GATA4 from Caco2 and HepG2 cells did increase total *SPINK1* expression on the background of a homozygous non-risk genotype. Second, siRNA knock downs did not lead to an increase in *SPINK1* promoter activity in the cells transfected with the *rs148911734* risk (T) allele. Third, GATA6 and GATA4 exogenous overexpression in Paca44, did not repress the expression of the risk allele compared to the non-risk allele. The latter observations raised the question if overexpressed GATA6 and GATA4 in Paca44 would at all bind to the site of interest. In EMSA experiments, nuclear extracts from Paca44 showed an

increased binding to the risk allele after the overexpression of GATA4, but almost no binding was observed in response to GATA6 overexpression. Western blot analysis revealed an increased abundance of a higher molecular weight form of GATA6 (Paca44, Panc1, HEK293, HEK293T) after exogenous overexpression compared to the endogenously expressed form (Caco2). Interestingly, endogenous GATA4 from HepG2 cells presented at a higher molecular weight than expected and overexpression of GATA4 had a less dramatic effect. While an additional smaller molecular weight form appeared upon GATA4 overexpression, the higher molecular form still dominated. Conclusively, GATA4 DNA binding ability was independent of endo- or exogenous expression, while GATA6 lost its DNA binding ability upon exogenous expression. The high molecular weight suggested protein modifications of the overexpressed form. Human GATA6 possesses 14 sites for phosphorylation, 1 site for methylation and 3 reported sites for SUMOylation (K429, K473, K484, Hendriks *et al*., 2017). Small ubiquitin-related modifiers (SUMO) proteins (SUMO1-4) have an estimated molecular size of 10-12 kDa. Hence, SUMOylation caught our particular interest to explain the approximate 30 kDa difference in molecular size of endo- and exogenous GATA6 on the Western blots. SUMO proteins were first described in 1996 by Martunis *et al*. and belong to the ubiquitin and ubiquitin-like superfamily (Matunis *et al*., 1996). In contrast to ubiquitin, the cellular function of SUMO group goes beyond 26S proteasomal degradation and can modify proteins in their activity and cellular localization and even protect them from ubiquitination (Bernier-Villamor *et al*., 2002). Based on the *JASSA* database, SUMOylation at amino acid p.K484 appeared to be the most likely modification of GATA6. We thus introduced the p.K484R mutation into the pc.DNA3.1 GATA6 wild type construct in order to remove potential SUMO residues at this site. Subsequent overexpression of the wild type and p.K484R variant of GATA6 in Paca44 and Panc1 cells, however, did not lead to the expected decrease in the molecular size of the GATA6 p.K484R variant in Western blot analysis. The bands showed similar signal intensities, which suggested equally high expression levels. Also, the DNA binding ability of overexpressed GATA6 was not restored in the p.K484R variant. We further supported these observations by qPCR measurement, in which the overexpression of GATA6 wild type or p.K484R variant led to comparable mRNA expression levels. Likewise, no substantial trend with respect to target genes activation was deducible. While target gene expression in Panc1 appeared to

respond more effectively with GATA6 p.K484R than with the wild type, the differences were minor and merely represent a tendency. The picture presented very inconsistent in Paca44 cells, so that no conclusions could be drawn. It should be noted, however, that GATA6, regardless of its form as wild type or p.K484R variant, overall uttered an enhancing effect on the gene expression of typical acinar markers, which in combination with PTF1a, RBPJL and HNF1a exceeded the effects of the three factors alone. This effect was particularly pronounced in Panc1 (*SPINK1, PRSS1, CPA1, CEL3AB, CEL, CUZD1*) cells and to a lesser extent in Paca44 cells (*SPINK1, PRSS1*). Especially in Paca44, the combination of GATA6 and the *SPINK1* p.N34S heterozygous genotype did not show any repressive effect on the *SPINK1* expression. Instead, *SPINK1* induction was higher upon the presence of GATA6 than upon the action of PTF1a, RBPJL and HNF1a alone.

Taken together, the higher molecular weight of GATA6 when expressed exogenously cannot be explained by SUMOylation of p.K484. Mutation of the SUMOylation site to p.K484R did not lead to a change in protein size (Western blot), DNA binding ability (EMSA) or target gene activation (qPCR). Further experiments should therefore aim to identify the reason for the different molecular size of GATA6 and its effects on protein function. As a first step, the kind and site of post-translational GATA6 modification should be clarified and should take phosphorylation and methylation besides SUMOylation into considerations (Hendriks *et al*., 2017). Unfortunately, attempts to detect GATA6 SUMOylation by Western blot with or without upstream immunoprecipitation remained without success (work conducted by Johanna Wallner as part of her Master thesis). For future experiments, chemical or enzymatic removal of the potential modifications should be pursued as an alternative approach before further investigating the protein size or function.

While our experiments argue strongly against a p.K484 as relevant site for SUMO modification, the loss of DNA binding ability in EMSA experiments, however, reminds of the defects observed on human GATA6 mutations (p.R456C, p.N466D, p.A467T, p.K473Q). For these mutations, a diminished promoter activity was observed in reporter gene assays, suggesting impaired DNA binding and promoter activation. Importantly, the expression levels of GATA6 remained unaffected. (Allen

et al., 2011). The same observations apply for the expression levels of GATA6 wild type and p.K484R in our experiments. Besides EMSA, chromatin immune precipitation (ChiP) could provide an alternative approach for the investigation of protein and DNA interaction. ChiP experiments would bear the advantage of an *in vivo* approach which would assess the transcription factor binding within the cellular context.

Moreover, Chitforoushzadeh *et al.* provide valuable insights on the origin and function of the different molecular size GATA6 in the intestinal HT29 cell line. For instance, the authors distinguished a full-length ($GATA6_L$, 60 kDa) and a short form ($GATA6_S$, 45 kDa) which originate from a long N-terminal extension or an in-frame alternative start codon (Met^{147}), respectively. In terms of biological function, $GATA6_L$ was suggested as the more potent transcriptional regulator despite its lower abundance compared to $GATA6_S$. In addition, posttranslational modifications on either $GATA6_L$ or $GATA6_S$ were described to cause an apparent molecular weight of 69-75 kDa ($GATA6_L$) and 54 kDa ($GATA6_S$) depending on the electrophoresis conditions. The investigation of posttranslational modifications centered around the amino acid and phosphorylation site Ser^{37} in $GATA6_L$. Phosphorylation by KT–glycogen synthase kinase 3 (GSK3) accelerated $GATA6_L$ degradation, increased GATA6 turnover and thereby reduced its transcriptional repression (Chitforoushzadeh *et al*., 2016).

Without doubt, these observations add complexity to the topic. But more importantly, they provide plausible explanations to our so-far unresolved questions and suggest how to answer them. The evaluation of our present results and the design of further experiments should certainly be built onto these findings.

5.6 Chances and limitations of a new acinar cell model approach

The identification and functional characterization of the pathological variant underlying the *SPINK1* p.N34S haplotype was particularly complicated by the lack of an adequate study model. While *SPINK1* expressing cells (Caco2, HepG2) are not from a pancreatic background and do not carry the risk allele, cells of apparent pancreatic origin (Panc1, Paca44) showed low *SPINK1* expression levels on mRNA and protein level. We thus chose a middle path in which we would use the *SPINK1*-

expressing cells to identify relevant transcriptional regulators, while we intended to apply those findings in the setting of the pancreatic cell lines. We hoped to take advantage of the p.N34S heterozygosity of Paca44 cells. Original attempts to insert a heterozygous genotype into Caco2 cell line by CRISPR/Cas9 faced serious challenges and had to be put aside for the current investigations. A recent publication of Kim *et al*. fostered the idea to adopt and improve the existing pancreatic cell lines by a targeted re-programming through transcription factor overexpression (Kim *et al*., 2015). An extensive literature search led us to the conclusion that the overexpression of the PTF1 complex, consisting of the E47, PTF1a and RBPJL subunits, HNF1a and NR5A2 had the potential to reconstitute an acinar phenotype in PDA cells (Figure 12). When we applied those findings in overexpression experiments in Panc1, we were able to partially reconstitute the findings from the literature. While the expression of some genes (*PRSS1, CELA3AB, CEL, CUZD1, SYNC, TRPV6, CTRL, MIST1*) clearly benefited from the presence of E47 as part of the PTF1 complex, E47 was optional for the expression of *SPINK1* and *CTRC*; the latter genes were more dependent on the presence of PTF1a, RBPJL and HNF1a. Those findings mostly went in line with our expectations. However, we were unable to reconstitute the effect of NR5A2 on *CEL* expression (Holmstrom *et al*., 2011) and instead observed a major effect on the expression of *CPA1*. Further, the suggested effect of HNF1a on *NR5A2* expression (Molero *et al*., 2012) was missing in our experiments. Most importantly for the underlying project was the successful induction of *SPINK1* by the overexpression of PTF1a, RBPJL and HNF1a, which confirmed the findings of Boulling *et al.* (Boulling *et al*., 2017). Astonishingly and against all expectations from literature (Masui *et al*., 2010, Martinelli *et al*., 2013**)**, the expression of *PNLIP*, *PNLIPRP1* and *PLA2G1B* was completely irresponsive to any of the transcription factor combinations. Also, when we extended the panel of overexpressed transcription factors by GATA4, GATA6 and HEYL (Hairy/enhancer of split related with YRPW motif-like protein), the lipase genes did not show any response (data in appendix).

After the successful induction of *SPINK1* expression in Panc1 (6-fold), we thought to apply those findings in Paca44 cells. While the Panc1 cell line is homozygous for the non-risk allele of the *SPINK1* p.N34S variant, the induction of *SPINK1* expression in a heterozygous cell line such as Paca44 appeared to be a suitable

model. Unfortunately, the promising approach that resulted in an acinar-like phenotype in Panc1 cells was less effective in Paca44 cells. The combination of PTF1a, RBPJL and HNF1a (with or without E47) only led to a barely 2-fold increase in *SPINK1* expression with high standard deviation. This instability in *SPINK1* expression could in part also explain the inconsistent results we obtained from the investigation of the allelic expression in Paca44 after the overexpression of PTF1a, RBPJL and HNF1a and GATA4, GATA6 or the knock down of *YY1*, *GATA4* and *GATA6*. While we did not apply the full set of transcription factors, some trends were still confirmed. For instance, *PRSS1*, *CELA3AB*, *CEL*, *CUZD1*, *MIST1* and additionally *CLPS* showed increased expression when E47 was present compared to when the combination of PTF1a, RBPJL and HNF1a was used alone. Surprisingly, *CEL* responded most effectively to the overexpression of HNF1a as a single factor. Also, the effect size of gene induction differed between the two cell lines. For example, *PRSS1*, *SPINK1*, *CEL*, *CUZD1* and *BHLHA15* (MIST1) showed lower effect sizes, the fold change in gene expression of *CELA3AB* was much more pronounced and the induction of *CPA1* and *CLPs* was in a similar range.

We can summarize that while Panc1 clearly maintain a certain plasticity for the reprogramming towards an acinar-like phenotype, we were unable to reconstitute the full panel of marker genes in Paca44. Those results suggest the role of a yet unidentified factor that is crucial for acinar cell identity. Our experiments, though, lay a solid basis for the development of an acinar-like cell model. While our experiments were all based on transient overexpression, the final model should stably overexpress the selected transcription factors to reduce batch-to-batch variations. An inducible system (*e.g.*, by tamoxifen), would be recommendable since we effectively demonstrated the induction of many digestive enzymes which could complicate the maintenance of the cells. Paca44, on the other hand, seem less susceptible to *SPINK1* induction and the reprogramming to an acinar-like phenotype. Although they have the advantage of a *SPINK1* p.N34S heterozygous genotype, they do not provide a suitable model for the functional study of the underlying pathological mechanism even after the minor induction of *SPINK1* expression by PTF1a, RBPJL and HNF1a.

5.7 *SPINK1* p.N34S in perspective of known CP mechanisms and the present work

After the genetic nature of pancreatitis was first described in 1952 by Comfort and Steinberg (Comfort & Steinberg, 1952) and the *PRSS1* p.R122H gain-of-function mutation was identified to result in intra-pancreatic autoactivation of trypsinogen (Whitcomb *et al*., 1996) the search for CP associated risk genes and their related variants revealed several different genes and disease mechanisms. The categorization of risk genes follows distinct premises based on *i.e.,* cellular expression (acinar *vs*. ductal), functional consequences of variants (gain- *vs*. loss-of-function) and contribution to the underlying disease mechanism. Moreover, two major pathological models were established depending on the underlying cellular events: the trypsin-dependent and the misfolding-dependent pathway (Hegyi & Sahin-Tóth, 2017, Sahin-Tóth, 2017).

SPINK1 is highly expressed in the acinar compartment and forms a first line of defense against prematurely activated trypsin (Pubols *et al*., 1974). It is therefore considered a protective factor in the context of the trypsin-dependent CP initiation and progression. Even though genetic studies undoubtfully established the association of *SPINK1* p.N34S and CP, the underlying disease mechanism remains elusive. Cumulative evidence from functional investigations of the last years (Kuwata *et al*., 2002**,** Király *et al*., 2007**,** Kereszturi *et al*., 2009**,** Kereszturi *et al*., 2017**,** Boulling *et al*., 2017) point towards a mechanism in which instead of an altered protein function, impaired gene regulation might alter *SPINK1* expression levels and thus influence the fine balance of proteases and their specific inhibitors within the pancreas.

While many variants of digestive enzymes lead to a biochemical gain- or loss-of function effect (*PRSS1* (Whitcomb *et al*., 1996**)**, *PRSS2* (Jancsó *et al*., 2016), *CTRC* (Rosendahl *et al*., 2008), *CELA3B* (Moore *et al*., 2019, Masson *et al*., 2021), and *PNLIP* (Lasher *et al*., 2019). others affect the cellular stress levels based on protein misfolding and the induction of ER stress (*PRSS1* (Kereszturi *et al*., 2009) CPA1 (Witt *et al*., 2013), *CEL (*Fjeld *et al*., 2015). A less investigated pathological cause lays in the elevation of gene expression levels. For instance, a variant in the *PRSS1–PRSS2* locus (*rs10273639*) conferred the disease risk by altering

trypsinogen expression levels. The variant is located in the proximal promoter region of *PRSS1* (c.1-408 T>C) and analysis of human pancreatic tissue revealed a clear correlation between the allelic state and *PRSS1* transcripts. While homozygous risk allele carriers (CC) showed the highest *PRSS1* expression levels, intermediate expression levels were detected in the heterozygous (CT) carriers and the lowest levels in non-risk carriers (TT). In this case, increased CP susceptibility is rather attributed to increased trypsinogen expression levels than to an altered protein function (Whitcomb *et al*., 2012).

A second example is the inversion in the *CTRB1-CTRB2* locus, which affects the expression ratio of *CTRB1/CTRB2.* CTRB2 was shown to degrade PRSS2 and to attenuate trypsin activity more effectively than CTRB1. Analysis of human pancreatic mRNA, though, revealed a significantly higher *CTRB1/CTRB2* expression ratio in homozygous than in heterozygous risk allele carriers. The protective PRSS2 degradation is impaired upon decreased *CTRB2* expression and thus facilitates increased trypsinogen activation in CP pathogenesis (Rosendahl *et al*., 2018).

Conclusively, regulatory variants in the *SPINK1* locus that cause diminished *SPINK1* expression are likely the cause of the high genetic association in ICP and TCP patients. Our experiments clearly argue against the previously suggested *rs142703147C>A* (c.1-4141G>T) variant as the pathological variant. Instead, we narrowed the selection of p.N34S linked variants down to *rs148276928* (c.1-2090T>C) and *rs148911734* (c.1-7321C>T). Based on our proteomics findings, we turned to *rs148911734* for deeper investigation. However, our functional analysis was seriously hampered by the lack of an adequate study model. While we were unable to prove a repressive action of GATA6 at this site so far, the experimental evidence and importance of the factor within pancreas development and homeostasis strongly suggest a relevance. In this perspective, the clarification of the nature and consequence of the different molecular size GATA6 will be crucial.

References

1000 Genomes Project Consortium, **Abecasis** GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65. doi: 10.1038/nature11632. PMID: 23128226; PMCID: PMC3498066.

Allen HL, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R; International Pancreatic Agenesis Consortium, Ferrer J, Hattersley AT, Ellard S. GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat Genet. 2011 Dec 11;44(1):20-22. doi: 10.1038/ng.1035. PMID: 22158542; PMCID: PMC4062962.

Archer H, Jura N, Keller J, Jacobson M, Bar-Sagi D. A mouse model of hereditary pancreatitis generated by transgenic expression of R122H trypsinogen. Gastroenterology. 2006 Dec;131(6):1844-55. doi: 10.1053/j.gastro.2006.09.049. Epub 2006 Oct 1. PMID: 17087933.

Athwal T, Huang W, Mukherjee R, Latawiec D, Chvanov M, Clarke R, Smith K, Campbell F, Merriman C, Criddle D, Sutton R, Neoptolemos J, Vlatković N. Expression of human cationic trypsinogen (PRSS1) in murine acinar cells promotes pancreatitis and apoptotic cell death. Cell Death Dis. 2014 Apr 10;5(4):e1165. doi: 10.1038/cddis.2014.120. PMID: 24722290; PMCID: PMC5424103.

Bartelt DC, Shapanka R, Greene LJ. The primary structure of the human pancreatic secretory trypsin inhibitor. Amino acid sequence of the reduced S-aminoethylated protein. Arch Biochem Biophys. 1977 Feb;179(1):189-99. doi: 10.1016/0003-9861(77)90103-5. PMID: 843082.

Beres TM, Masui T, Swift GH, Shi L, Henke RM, MacDonald RJ. PTF1 is an organ-specific and Notch-independent basic helix-loop-helix complex containing the mammalian Suppressor of Hairless (RBP-J) or its paralogue, RBP-L. Mol Cell Biol. 2006 Jan;26(1):117-30. doi: 10.1128/MCB.26.1.117-130.2006. PMID: 16354684; PMCID: PMC1317634.

Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW 3rd, Bulyk ML. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol. 2006 Nov;24(11):1429-35. doi: 10.1038/nbt1246. Epub 2006 Sep 24. PMID: 16998473; PMCID: PMC4419707.

Bernier-Villamor V, Sampson DA, Matunis MJ, Lima CD. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. Cell. 2002 Feb 8;108(3):345-56. doi: 10.1016/s0092-8674(02)00630-x. PMID: 11853669.

Bhatia E, Choudhuri G, Sikora SS, Landt O, Kage A, Becker M, Witt H. Tropical calcific pancreatitis: strong association with SPINK1 trypsin inhibitor mutations. Gastroenterology. 2002 Oct;123(4):1020-5. doi: 10.1053/gast.2002.36028. PMID: 12360463.

Boulling A, Masson E, Zou WB, Paliwal S, Wu H, Issarapu P, Bhaskar S, Génin E, Cooper DN, Li ZS, Chandak GR, Liao Z, Chen JM, Férec C. Identification of a functional enhancer variant within the chronic pancreatitis-associated SPINK1 c.101A>G (p.Asn34Ser)-containing haplotype. Hum Mutat. 2017 Aug;38(8):1014-1024. doi: 10.1002/humu.23269. Epub 2017 Jun 15. PMID: 28556356.

Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976 May 7;72:248-54. doi: 10.1006/abio.1976.9999. PMID: 942051.

Carrasco M, Delgado I, Soria B, Martín F, Rojas A. GATA4 and GATA6 control mouse pancreas organogenesis. J Clin Invest. 2012 Oct;122(10):3504-15. doi: 10.1172/JCI63240. Epub 2012 Sep 24. PMID: 23006330; PMCID: PMC3461915.

Chandak GR, Idris MM, Reddy DN, Bhaskar S, Sriram PV, Singh L. Mutations in the pancreatic secretory trypsin inhibitor gene (PSTI/SPINK1) rather than the cationic trypsinogen gene (PRSS1) are significantly associated with tropical calcific pancreatitis. J Med Genet. 2002 May;39(5):347-51. doi: 10.1136/jmg.39.5.347. PMID: 12011155; PMCID: PMC1735106.

Chen H, Sun W, Zhu J, Yuan H, Chu M, Wen B. Modification of cardiac transcription factor Gata6 by SUMO. Biochimie. 2020 Mar;170:212-218. doi: 10.1016/j.biochi.2020.01.014. Epub 2020 Feb 1. PMID: 32017966.

Chiari H, Über Selbstverdauung des menschlichen Pankreas, Zeitschrift für Heilkunde 1896: 17; 69-96

Chitforoushzadeh Z, Ye Z, Sheng Z, LaRue S, Fry RC, Lauffenburger DA, Janes KA. TNF-insulin crosstalk at the transcription factor GATA6 is revealed by a model that links signaling and transcriptomic data tensors. Sci Signal. 2016 Jun 7;9(431):ra59. doi: 10.1126/scisignal.aad3373. PMID: 27273097; PMCID: PMC4914393.

Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, Lee H, Oskolkov N, Fadista J, Ehlers K, Wahl S, Hoffmann C, Qian K, Rönn T, Riess H, Müller-Nurasyid M, Bretschneider N, Schroeder T, Skurk T, Horsthemke B; DIAGRAM+Consortium, Spieler D, Klingenspor M, Seifert M, Kern MJ, Mejhert N, Dahlman I, Hansson O, Hauck SM, Blüher M, Arner P, Groop L, Illig T, Suhre K, Hsu YH, Mellgren G, Hauner H, Laumen H. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 2014 Jan 16;156(1-2):343-58. doi: 10.1016/j.cell.2013.10.058. PMID: 24439387; PMCID: PMC7116609.

Cockell M, Stevenson BJ, Strubin M, Hagenbüchle O, Wellauer PK. Identification of a cell-specific DNA-binding activity that interacts with a transcriptional activator of genes expressed in the acinar pancreas. Mol Cell Biol. 1989 Jun;9(6):2464-76. doi: 10.1128/mcb.9.6.2464-2476.1989. PMID: 2788241; PMCID: PMC362319.

Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics. 2015 Dec 1;31(23):3847-9. doi: 10.1093/bioinformatics/btv470. Epub 2015 Aug 12. PMID: 26272984; PMCID: PMC4653394.

Cohn JA, Friedman KJ, Noone PG, Knowles MR, Silverman LM, Jowell PS. Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. N Engl J Med. 1998 Sep 3;339(10):653-8. doi: 10.1056/NEJM199809033391002. PMID: 9725922.

Comfort MW, Steinberg AG. Pedigree of a family with hereditary chronic relapsing pancreatitis. Gastroenterology. 1952 May;21(1):54-63. PMID: 14926813.

Conwell DL, Lee LS, Yadav D, Longnecker DS, Miller FH, Mortele KJ, Levy MJ, Kwon R, Lieb JG, Stevens T, Toskes PP, Gardner TB, Gelrud A, Wu BU, Forsmark CE, Vege SS. American Pancreatic Association Practice Guidelines in Chronic Pancreatitis: evidence-based report on diagnostic guidelines. Pancreas. 2014 Nov;43(8):1143-62. doi: 10.1097/MPA.0000000000000237. PMID: 25333398; PMCID: PMC5434978.

Decker K, Goldman DC, Grasch CL, Sussel L. Gata6 is an important regulator of mouse pancreas development. Dev Biol. 2006 Oct 15;298(2):415-29. doi: 10.1016/j.ydbio.2006.06.046. Epub 2006 Jul 4. PMID: 16887115; PMCID: PMC2824170.

Derikx MH, Kovacs P, Scholz M, Masson E, Chen JM, Ruffert C, Lichtner P, Te Morsche RH, Cavestro GM, Férec C, Drenth JP, Witt H, Rosendahl J; PanEuropean Working group on Alcoholic Chronic Pancreatitis Members and Collaborators. Polymorphisms at PRSS1-PRSS2 and CLDN2-MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. Gut. 2015 Sep;64(9):1426-33. doi: 10.1136/gutjnl-2014-307453. Epub 2014 Sep 24. PMID: 25253127.

Durno C, Corey M, Zielenski J, Tullis E, Tsui LC, Durie P. Genotype and phenotype correlations in patients with cystic fibrosis and pancreatitis. Gastroenterology. 2002 Dec;123(6):1857-64. doi: 10.1053/gast.2002.37042. PMID: 12454843.

Fayard E, Schoonjans K, Annicotte JS, Auwerx J. Liver receptor homolog 1 controls the expression of carboxyl ester lipase. J Biol Chem. 2003 Sep 12;278(37):35725-31. doi: 10.1074/jbc.M302370200. Epub 2003 Jul 9. PMID: 12853459.

Fjeld K, Weiss FU, Lasher D, Rosendahl J, Chen JM, Johansson BB, Kirsten H, Ruffert C, Masson E, Steine SJ, Bugert P, Cnop M, Grützmann R, Mayerle J, Mössner J, Ringdal M, Schulz HU, Sendler M, Simon P, Sztromwasser P, Torsvik J, Scholz M, Tjora E, Férec C, Witt H, Lerch MM, Njølstad PR, Johansson S, Molven A. A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis. Nat Genet. 2015 May;47(5):518-522. doi: 10.1038/ng.3249. Epub 2015 Mar 16. PMID: 25774637; PMCID: PMC5321495.

Hendriks IA, Lyon D, Young C, Jensen LJ, Vertegaal AC, Nielsen ML. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. Nat Struct Mol Biol. 2017 Mar;24(3):325-336. doi: 10.1038/nsmb.3366. Epub 2017 Jan 23. PMID: 28112733.

Hegyi E, Sahin-Tóth M. Genetic Risk in Chronic Pancreatitis: The Trypsin-Dependent Pathway. Dig Dis Sci. 2017 Jul;62(7):1692-1701. doi: 10.1007/s10620-017-4601-3. Epub 2017 May 23. PMID: 28536777; PMCID: PMC5487703.

Hegyi E, Sahin-Tóth M. Human *CPA1* mutation causes digestive enzyme misfolding and chronic pancreatitis in mice. Gut. 2019 Feb;68(2):301-312. doi: 10.1136/gutjnl-2018-315994. Epub 2018 Jul 25. PMID: 30045879; PMCID: PMC6326849.

Hegyi P, Párniczky A, Lerch MM, Sheel ARG, Rebours V, Forsmark CE, Del Chiaro M, Rosendahl J, de-Madaria E, Szücs Á, Takaori K, Yadav D, Gheorghe C, Rakonczay Z Jr, Molero X, Inui K, Masamune A, Fernandez-Del Castillo C, Shimosegawa T, Neoptolemos JP, Whitcomb DC, Sahin-Tóth M; Working Group for the International (IAP – APA – JPS – EPC) Consensus Guidelines for Chronic Pancreatitis. International Consensus Guidelines for Risk Factors in Chronic Pancreatitis. Recommendations from the working group for the international consensus guidelines for chronic pancreatitis in collaboration with the International Association of Pancreatology, the American Pancreatic Association, the Japan

Pancreas Society, and European Pancreatic Club. Pancreatology. 2020 Jun;20(4):579-585. doi: 10.1016/j.pan.2020.03.014. Epub 2020 Apr 8. PMID: 32376198.

Holmstrom SR, Deering T, Swift GH, Poelwijk FJ, Mangelsdorf DJ, Kliewer SA, MacDonald RJ. LRH-1 and PTF1-L coregulate an exocrine pancreas-specific transcriptional network for digestive function. Genes Dev. 2011 Aug 15;25(16):1674-9. doi: 10.1101/gad.16860911. PMID: 21852532; PMCID: PMC3165932.

Horii A, Kobayashi T, Tomita N, Yamamoto T, Fukushige S, Murotsu T, Ogawa M, Mori T, Matsubara K. Primary structure of human pancreatic secretory trypsin inhibitor (PSTI) gene. Biochem Biophys Res Commun. 1987 Dec 16;149(2):635-41. doi: 10.1016/0006-291x(87)90415-3. PMID: 3501289.

Hoskins JW, Jia J, Flandez M, Parikh H, Xiao W, Collins I, Emmanuel MA, Ibrahim A, Powell J, Zhang L, Malats N, Bamlet WR, Petersen GM, Real FX, Amundadottir LT. Transcriptome analysis of pancreatic cancer reveals a tumor suppressor function for HNF1A. Carcinogenesis. 2014 Dec;35(12):2670-8. doi: 10.1093/carcin/bgu193. Epub 2014 Sep 18. PMID: 25233928; PMCID: PMC4247517.

Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. 2019 Jan 8;47(D1):D33-D38. doi: 10.1093/nar/gky822. PMID: 30204897; PMCID: PMC6323978.

Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. 2015 Jan;43 (Database issue):D117-22. doi: 10.1093/nar/gku1045. Epub 2014 Nov 5. PMID: 25378322; PMCID: PMC4383892.

Jakubison BL, Schweickert PG, Moser SE, Yang Y, Gao H, Scully K, Itkin-Ansari P, Liu Y, Konieczny SF. Induced PTF1a expression in pancreatic ductal adenocarcinoma cells activates acinar gene networks, reduces tumorigenic properties, and sensitizes cells to gemcitabine treatment. Mol Oncol. 2018 Jun;12(7):1104-1124. doi: 10.1002/1878-0261.12314. Epub 2018 May 21. PMID: 29719936; PMCID: PMC6026875.

Jancsó Z, Sahin-Tóth M. Tighter Control by Chymotrypsin C (CTRC) Explains Lack of Association between Human Anionic Trypsinogen and Hereditary Pancreatitis. J Biol Chem. 2016 Jun 17;291(25):12897-905. doi: 10.1074/jbc.M116.725374. Epub 2016 Apr 18. PMID: 27129265; PMCID: PMC4933207.

Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J. DNA-binding specificities of human transcription factors. Cell. 2013 Jan 17;152(1-2):327-39. doi: 10.1016/j.cell.2012.12.009. PMID: 23332764.

Kalisz M, Bernardo E, Beucher A, Maestro MA, Del Pozo N, Millán I, Haeberle L, Schlensog M, Safi SA, Knoefel WT, Grau V, de Vas M, Shpargel KB, Vaquero E, Magnuson T, Ortega S, Esposito I, Real FX, Ferrer J. HNF1A recruits KDM6A to activate differentiated acinar cell programs that suppress pancreatic cancer. EMBO J. 2020 May 4;39(9):e102808. doi: 10.15252/embj.2019102808. Epub 2020 Mar 10. PMID: 32154941; PMCID: PMC7196917.

Kereszturi E, Szmola R, Kukor Z, Simon P, Weiss FU, Lerch MM, Sahin-Tóth M. Hereditary pancreatitis caused by mutation-induced misfolding of human cationic trypsinogen: a novel disease mechanism. Hum Mutat. 2009 Apr;30(4):575-82. doi: 10.1002/humu.20853. PMID: 19191323; PMCID: PMC2663013.

Kereszturi É, Sahin-Tóth M. Pancreatic Cancer Cell Lines Heterozygous for the SPINK1 p.N34S Haplotype Exhibit Diminished Expression of the Variant Allele. Pancreas. 2017 Jul;46(6):e54-e55. doi: 10.1097/MPA.0000000000000817. PMID: 28609377; PMCID: PMC5470582.

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 2018 Jan 4;46(D1):D260-D266. doi: 10.1093/nar/gkx1126. Erratum in: Nucleic Acids Res. 2018 Jan 4;46(D1):D1284. PMID: 29140473; PMCID: PMC5753243.

Kim S, Lahmy R, Riha C, Yang C, Jakubison BL, van Niekerk J, Staub C, Wu Y, Gates K, Dong DS, Konieczny SF, Itkin-Ansari P. The basic helix-loop-helix transcription factor E47 reprograms human pancreatic cancer cells to a quiescent acinar state with reduced tumorigenic potential. Pancreas. 2015 Jul;44(5):718-27. doi: 10.1097/MPA.0000000000000328. PMID: 25894862; PMCID: PMC4464938.

Király O, Wartmann T, Sahin-Tóth M. Missense mutations in pancreatic secretory trypsin inhibitor (SPINK1) cause intracellular retention and degradation. Gut. 2007 Oct;56(10):1433-8. doi: 10.1136/gut.2006.115725. Epub 2007 May 24. PMID: 17525091; PMCID: PMC2000263.

Klauss S, Schorn S, Teller S, Steenfadt H, Friess H, Ceyhan GO, Demir IE. Genetically induced vs. classical animal models of chronic pancreatitis: a critical comparison. FASEB J. 2018 Jun 4:fj201800241RR. doi: 10.1096/fj.201800241RR. Epub ahead of print. PMID: 29863911.

Krapp A, Knöfler M, Frutiger S, Hughes GJ, Hagenbüchle O, Wellauer PK. The p48 DNA-binding subunit of transcription factor PTF1 is a new exocrine pancreas-specific basic helix-loop-helix protein. EMBO J. 1996 Aug 15;15(16):4317-29. PMID: 8861960; PMCID: PMC452157.

Krapp A, Knöfler M, Ledermann B, Bürki K, Berney C, Zoerkler N, Hagenbüchle O, Wellauer PK. The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. Genes Dev. 1998 Dec 1;12(23):3752-63. doi: 10.1101/gad.12.23.3752. PMID: 9851981; PMCID: PMC317250.

Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-Alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, Makeev VJ. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. 2016 Jan 4;44(D1):D116-25. doi: 10.1093/nar/gkv1249. Epub 2015 Nov 19. PMID: 26586801; PMCID: PMC4702883.

Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, Medvedeva YA, Magana-Mora A, Bajic VB, Papatsenko DA, Kolpakov FA, Makeev VJ. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. 2018 Jan 4;46(D1):D252-D259. doi: 10.1093/nar/gkx1106. PMID: 29140464; PMCID: PMC5753240.

Roadmap Epigenomics Consortium, **Kundaje** A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30. doi: 10.1038/nature14248. PMID: 25693563; PMCID: PMC4530010.

Kuwata K, Hirota M, Sugita H, Kai M, Hayashi N, Nakamura M, Matsuura T, Adachi N, Nishimori I, Ogawa M. Genetic mutations in exons 3 and 4 of the pancreatic secretory trypsin inhibitor in patients with pancreatitis. J Gastroenterol. 2001 Sep;36(9):612-8. doi: 10.1007/s005350170045. PMID: 11578065.

Kuwata K, Hirota M, Shimizu H, Nakae M, Nishihara S, Takimoto A, Mitsushima K, Kikuchi N, Endo K, Inoue M, Ogawa M. Functional analysis of recombinant pancreatic secretory trypsin inhibitor protein with amino-acid substitution. J Gastroenterol. 2002;37(11):928-34. doi: 10.1007/s005350200156. PMID: 12483248.

Lasher D, Szabó A, Masamune A, Chen JM, Xiao X, Whitcomb DC, Barmada MM, Ewers M, Ruffert C, Paliwal S, Issarapu P, Bhaskar S, Mani KR, Chandak GR, Laumen H, Masson E, Kume K, Hamada S, Nakano E, Seltsam K, Bugert P, Müller T, Groneberg DA, Shimosegawa T, Rosendahl J, Férec C, Lowe ME, Witt H, Sahin-Tóth M. Protease-Sensitive Pancreatic Lipase Variants Are Associated With Early Onset Chronic Pancreatitis. Am J Gastroenterol. 2019 Jun;114(6):974-983. doi: 10.14309/ajg.0000000000000051. PMID: 30789418; PMCID: PMC6624845.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001 Dec;25(4):402-8. doi: 10.1006/meth.2001.1262. PMID: 11846609.

Löhr JM, Dominguez-Munoz E, Rosendahl J, Besselink M, Mayerle J, Lerch MM, Haas S, Akisik F, Kartalis N, Iglesias-Garcia J, Keller J, Boermeester M, Werner J, Dumonceau JM, Fockens P, Drewes A, Ceyhan G, Lindkvist B, Drenth J, Ewald N, Hardt P, de Madaria E, Witt H, Schneider A, Manfredi R, Brøndum FJ, Rudolf S, Bollen T, Bruno M; HaPanEU/UEG Working Group. United European Gastroenterology evidence-based guidelines for the diagnosis and therapy of chronic pancreatitis (HaPanEU). United European Gastroenterol J. 2017 Mar;5(2):153-199. doi: 10.1177/2050640616684695. Epub 2017 Jan 16. PMID: 28344786; PMCID: PMC5349368.

Luo Z, Li Y, Wang H, Fleming J, Li M, Kang Y, Zhang R, Li D. Hepatocyte nuclear factor 1A (HNF1A) as a possible tumor suppressor in pancreatic cancer. PLoS One. 2015 Mar 20;10(3):e0121082. doi: 10.1371/journal.pone.0121082. PMID: 25793983; PMCID: PMC4368635.

Martinelli P, Cañamero M, del Pozo N, Madriles F, Zapata A, Real FX. Gata6 is required for complete acinar differentiation and maintenance of the exocrine pancreas in adult mice. Gut. 2013 Oct;62(10):1481-8. doi: 10.1136/gutjnl-2012-303328. Epub 2012 Sep 21. PMID: 23002247.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. 2014 Jan;42(Database issue):D142-7. doi: 10.1093/nar/gkt997. Epub 2013 Nov 4. PMID: 24194598; PMCID: PMC3965086.

Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016 Jan 4;44(D1):D110-5. doi: 10.1093/nar/gkv1176. Epub 2015 Nov 3. PMID: 26531826; PMCID: PMC4702842.

Matunis MJ, Coutavas E, Blobel G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. J Cell Biol. 1996 Dec;135(6 Pt 1):1457-70. doi: 10.1083/jcb.135.6.1457. PMID: 8978815; PMCID: PMC2133973.

Masamune A, Kotani H, Sörgel FL, Chen JM, Hamada S, Sakaguchi R, Masson E, Nakano E, Kakuta Y, Niihori T, Funayama R, Shirota M, Hirano T, Kawamoto T, Hosokoshi A, Kume K, Unger L, Ewers M, Laumen H, Bugert P, Mori MX, Tsvilovskyy V, Weißgerber P, Kriebs U, Fecher-Trost C, Freichel M, Diakopoulos KN, Berninger A, Lesina M, Ishii K, Itoi T, Ikeura T, Okazaki K, Kaune T, Rosendahl J, Nagasaki M, Uezono Y, Algül H, Nakayama K, Matsubara Y, Aoki Y, Férec C, Mori Y, Witt H, Shimosegawa T. Variants That Affect Function of Calcium Channel TRPV6 Are Associated With Early-Onset Chronic Pancreatitis. Gastroenterology. 2020 May;158(6):1626-1641.e8. doi: 10.1053/j.gastro.2020.01.005. Epub 2020 Jan 10. PMID: 31930989.

Masson E, Rebours V, Buscail L, Frete F, Pagenault M, Lachaux A, Chevaux JB, Génin E, Cooper DN, Férec C, Chen JM. The reversion variant (p.Arg90Leu) at the evolutionarily adaptive p.Arg90 site in CELA3B predisposes to chronic pancreatitis. Hum Mutat. 2021 Apr;42(4):385-391. doi: 10.1002/humu.24178. Epub 2021 Feb 19. PMID: 33565216.

Masui T, Long Q, Beres TM, Magnuson MA, MacDonald RJ. Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. Genes Dev. 2007 Oct 15;21(20):2629-43. doi: 10.1101/gad.1575207. PMID: 17938243; PMCID: PMC2000326.

Masui T, Swift GH, Deering T, Shen C, Coats WS, Long Q, Elsässer HP, Magnuson MA, MacDonald RJ. Replacement of Rbpj with Rbpjl in the PTF1 complex controls the final maturation of pancreatic acinar cells. Gastroenterology. 2010 Jul;139(1):270-80. doi: 10.1053/j.gastro.2010.04.003. Epub 2010 Apr 14. PMID: 20398665; PMCID: PMC2902682.

Molero X, Vaquero EC, Flández M, González AM, Ortiz MÁ, Cibrián-Uhalte E, Servitja JM, Merlos A, Juanpere N, Massumi M, Skoudy A, Macdonald R, Ferrer J, Real FX. Gene expression dynamics after murine pancreatitis unveils novel roles for Hnf1α in acinar cell homeostasis. Gut. 2012 Aug;61(8):1187-96. doi: 10.1136/gutjnl-2011-300360. Epub 2011 Sep 23. PMID: 21948943.

Moore PC, Cortez JT, Chamberlain CE, Alba D, Berger AC, Quandt Z, Chan A, Cheng MH, Bautista JL, Peng J, German MS, Anderson MS, Oakes SA. Elastase 3B mutation links to familial pancreatitis with diabetes and pancreatic adenocarcinoma. J Clin Invest. 2019 Aug 1;129(11):4676-4681. doi: 10.1172/JCI129961. PMID: 31369399; PMCID: PMC6819098.

Nathan JD, Romac J, Peng RY, Peyton M, Macdonald RJ, Liddle RA. Transgenic expression of pancreatic secretory trypsin inhibitor-I ameliorates secretagogue-induced pancreatitis in mice. Gastroenterology. 2005 Mar;128(3):717-27. doi: 10.1053/j.gastro.2004.11.052. PMID: 15765407.

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutyavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature. 2012 Sep 6;489(7414):83-90. doi: 10.1038/nature11212. PMID: 22955618; PMCID: PMC3736582.

Ohmuraya M, Hirota M, Araki K, Baba H, Yamamura K. Enhanced trypsin activity in pancreatic acinar cells deficient for serine protease inhibitor kazal type 3. Pancreas. 2006 Jul;33(1):104-6. doi: 10.1097/01.mpa.0000226889.86322.9b. PMID: 16804421.

Olesen SS, Mortensen LH, Zinck E, Becker U, Drewes AM, Nøjgaard C, Novovic S, Yadav D, Tolstrup JS. Time trends in incidence and prevalence of chronic pancreatitis: A 25-year population-based nationwide study. United European Gastroenterol J. 2021 Feb;9(1):82-90. doi: 10.1177/2050640620966513. Epub 2021 Feb 22. PMID: 33176616; PMCID: PMC8259237.

Pachkov M, Balwierz PJ, Arnold P, Ozonov E, van Nimwegen E. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res. 2013 Jan;41(Database issue):D214-20. doi: 10.1093/nar/gks1145. Epub 2012 Nov 24. PMID: 23180783; PMCID: PMC3531101.

Pagès H (2015). *SNPlocs.Hsapiens.dbSNP142.GRCh37: SNP locations for Homo sapiens (dbSNP Build 142)*. R package version 0.99.5.

Pfützer RH, Barmada MM, Brunskill AP, Finch R, Hart PS, Neoptolemos J, Furey WF, Whitcomb DC. SPINK1/PSTI polymorphisms act as disease modifiers in familial and idiopathic chronic pancreatitis. Gastroenterology. 2000 Sep;119(3):615-23. doi: 10.1053/gast.2000.18017. PMID: 10982753.

Pubols MH, Bartelt DC, Greene LJ. Trypsin inhibitor from human pancreas and pancreatic juice. J Biol Chem. 1974 Apr 10;249(7):2235-42. PMID: 4818832.

Ravnik-Glavac M, Glavac D, di Sant' Agnese P, Chernick M, Dean M. Cystic fibrosis gene mutations detected in hereditary pancreatitis. Pflugers Arch. 1996;431(6 Suppl 2):R191-2. doi: 10.1007/BF02346333. PMID: 8992448.

Rinderknecht H. Activation of pancreatic zymogens. Normal activation, premature intrapancreatic activation, protective mechanisms against inappropriate activation. Dig Dis Sci. 1986 Mar;31(3):314-21. doi: 10.1007/BF01318124. PMID: 2936587.

Romac JM, Ohmuraya M, Bittner C, Majeed MF, Vigna SR, Que J, Fee BE, Wartmann T, Yamamura K, Liddle RA. Transgenic expression of pancreatic secretory trypsin inhibitor-1 rescues SPINK3-deficient mice and restores a normal pancreatic phenotype. Am J Physiol Gastrointest Liver Physiol. 2010 Apr;298(4):G518-24. doi: 10.1152/ajpgi.00431.2009. Epub 2010 Jan 28. PMID: 20110462; PMCID: PMC2853299.

Rosendahl J, Witt H, Szmola R, Bhatia E, Ozsvári B, Landt O, Schulz HU, Gress TM, Pfützer R, Löhr M, Kovacs P, Blüher M, Stumvoll M, Choudhuri G, Hegyi P, te Morsche RH, Drenth JP, Truninger K, Macek M Jr, Puhl G, Witt U, Schmidt H, Büning C, Ockenga J, Kage A, Groneberg DA, Nickel R, Berg T, Wiedenmann B, Bödeker H, Keim V, Mössner J, Teich N, Sahin-Tóth M. Chymotrypsin C (CTRC) variants that diminish activity or secretion are associated with chronic pancreatitis. Nat Genet. 2008 Jan;40(1):78-82. doi: 10.1038/ng.2007.44. Epub 2007 Dec 2. PMID: 18059268; PMCID: PMC2650829.

Rosendahl J, Landt O, Bernadova J, Kovacs P, Teich N, Bödeker H, Keim V, Ruffert C, Mössner J, Kage A, Stumvoll M, Groneberg D, Krüger R, Luck W, Treiber M, Becker M, Witt H. CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated? Gut. 2013 Apr;62(4):582-92. doi: 10.1136/gutjnl-2011-300645. Epub 2012 Mar 17. PMID: 22427236.

Rosendahl J, Kirsten H, Hegyi E, Kovacs P, Weiss FU, Laumen H, Lichtner P, Ruffert C, Chen JM, Masson E, Beer S, Zimmer C, Seltsam K, Algül H, Bühler F, Bruno MJ, Bugert P, Burkhardt R, Cavestro GM, Cichoz-Lach H, Farré A, Frank J, Gambaro G, Gimpfl S, Grallert H, Griesmann H, Grützmann R, Hellerbrand C, Hegyi P, Hollenbach M, Iordache S, Jurkowska G, Keim V, Kiefer F, Krug S, Landt O, Leo MD, Lerch MM, Lévy P, Löffler M, Löhr M, Ludwig M, Macek M, Malats N, Malecka-Panas E, Malerba G, Mann K, Mayerle J, Mohr S, Te Morsche RHM, Motyka M, Mueller S, Müller T, Nöthen MM, Pedrazzoli S, Pereira SP, Peters A, Pfützer R, Real FX, Rebours V, Ridinger M, Rietschel M, Rösmann E, Saftoiu A, Schneider A, Schulz HU, Soranzo N, Soyka M, Simon P, Skipworth J, Stickel F, Strauch K, Stumvoll M, Testoni PA, Tönjes A, Werner L, Werner J, Wodarz N, Ziegler M, Masamune A, Mössner J, Férec C, Michl P, P H Drenth J, Witt H, Scholz M, Sahin-Tóth M; all members of the PanEuropean Working group on ACP. Genome-wide association study identifies inversion in the *CTRB1-CTRB2* locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. Gut. 2018 Oct;67(10):1855-1863. doi: 10.1136/gutjnl-2017-314454. Epub 2017 Jul 28. PMID: 28754779; PMCID: PMC6145291.

Roux E, Strubin M, Hagenbüchle O, Wellauer PK. The cell-specific transcription factor PTF1 contains two different subunits that interact with the DNA. Genes Dev. 1989 Oct;3(10):1613-24. doi: 10.1101/gad.3.10.1613. PMID: 2612907.

Sahin-Tóth M. Genetic risk in chronic pancreatitis: the misfolding-dependent pathway. Curr Opin Gastroenterol. 2017 Sep;33(5):390-395. doi: 10.1097/MOG.0000000000000380. PMID: 28650851; PMCID: PMC5549634.

Saloman JL, Albers KM, Cruz-Monserrate Z, Davis BM, Edderkaoui M, Eibl G, Epouhe AY, Gedeon JY, Gorelick FS, Grippo PJ, Groblewski GE, Husain SZ, Lai KKY, Pandol SJ, Uc A, Wen L, Whitcomb DC. Animal Models: Challenges and Opportunities to Determine Optimal Experimental Models of Pancreatitis and Pancreatic Cancer. Pancreas. 2019 Jul;48(6):759-779. doi: 10.1097/MPA.0000000000001335. PMID: 31206467; PMCID: PMC6581211.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet. 2002 Feb;70(2):425-34. doi: 10.1086/338688. Epub 2001 Dec 27. PMID: 11791212; PMCID: PMC384917.

Scharer CD, McCabe CD, Ali-Seyed M, Berger MF, Bulyk ML, Moreno CS. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. Cancer Res. 2009 Jan 15;69(2):709-17. doi: 10.1158/0008-5472.CAN-08-3415. PMID: 19147588; PMCID: PMC2629396.

Semaniakou A, Croll RP, Chappe V. Animal Models in the Pathophysiology of Cystic Fibrosis. Front Pharmacol. 2019 Jan 4;9:1475. doi: 10.3389/fphar.2018.01475. PMID: 30662403; PMCID: PMC6328443.

Sharer N, Schwarz M, Malone G, Howarth A, Painter J, Super M, Braganza J. Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. N Engl J Med. 1998 Sep 3;339(10):645-52. doi: 10.1056/NEJM199809033391001. PMID: 9725921.

Shannon P, Richards M (2021). *MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs*. R package version 1.34.0.

Sommer L, Hagenbüchle O, Wellauer PK, Strubin M. Nuclear targeting of the transcription factor PTF1 is mediated by a protein subunit that does not bind to the PTF1 cognate sequence. Cell. 1991 Nov 29;67(5):987-94. doi: 10.1016/0092-8674(91)90371-5. PMID: 1720355.

Storz P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol. 2017 May;14(5):296-304. doi: 10.1038/nrgastro.2017.12. Epub 2017 Mar 8. PMID: 28270694; PMCID: PMC6036907.

Sun C, Liu M, An W, Mao X, Jiang H, Zou W, Wu H, Liao Z, Li Z. Heterozygous Spink1 c.194+2T>C mutant mice spontaneously develop chronic pancreatitis. Gut. 2020 May;69(5):967-968. doi: 10.1136/gutjnl-2019-318790. Epub 2019 May 29. PMID: 31142585.

Shelton, CA. Whitcomb, DC. (2016). Hereditary Pancreatitis. Pancreapedia: Exocrine Pancreas Knowledge Base, doi: 10.3998/panc.2016.10.

Szmola R, Sahin-Tóth M. Chymotrypsin C (caldecrin) promotes degradation of human cationic trypsin: identity with Rinderknecht's enzyme Y. Proc Natl Acad Sci U S A. 2007 Jul 3;104(27):11227-32. doi: 10.1073/pnas.0703714104. Epub 2007 Jun 25. PMID: 17592142; PMCID: PMC2040881.

Team TBD (2020). BSgenome.Hsapiens.UCSC.hg19: Full genome sequences for Homo sapiens (UCSC version hg19, based on GRCh37.p13). R package version 1.4.3.

Truninger K, Witt H, Köck J, Kage A, Seifert B, Ammann RW, Blum HE, Becker M. Mutations of the serine protease inhibitor, Kazal type 1 gene, in patients with idiopathic chronic pancreatitis. Am J Gastroenterol. 2002 May;97(5):1133-7. doi: 10.1111/j.1572-0241.2002.05673.x. PMID: 12014716.

Uc A, Husain SZ. Pancreatitis in Children. Gastroenterology. 2019 May;156(7):1969-1978. doi: 10.1053/j.gastro.2018.12.043. Epub 2019 Feb 1. PMID: 30716320; PMCID: PMC6730664.

Villamayor L, Cano DA, Rojas A. GATA factors in pancreas development and disease. IUBMB Life. 2020 Jan;72(1):80-88. doi: 10.1002/iub.2170. Epub 2019 Oct 3. PMID: 31580534.

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano JC, Galli M, Lewsey MG, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, Govindarajan S, Shaulsky G, Walhout AJM, Bouget FY, Ratsch G, Larrondo LF, Ecker JR, Hughes TR. Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014 Sep 11;158(6):1431-1443. doi: 10.1016/j.cell.2014.08.009. PMID: 25215497; PMCID: PMC4163041.

Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, Martin SP, Gates LK Jr, Amann ST, Toskes PP, Liddle R, McGrath K, Uomo G, Post JC, Ehrlich GD. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nat Genet. 1996 Oct;14(2):141-5. doi: 10.1038/ng1096-141. PMID: 8841182.

Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, Neoptolemos JP, Lerch MM, Tector M, Sandhu BS, Guda NM, Orlichenko L; Alzheimer's Disease Genetics Consortium, Alkaade S, Amann ST, Anderson MA, Baillie J, Banks PA, Conwell D, Coté GA, Cotton PB, DiSario J, Farrer LA, Forsmark CE, Johnstone M, Gardner TB, Gelrud A, Greenhalf W, Haines JL, Hartman DJ, Hawes RA, Lawrence C, Lewis M, Mayerle J, Mayeux R, Melhem NM, Money ME, Muniraj T, Papachristou GI, Pericak-Vance MA, Romagnuolo J, Schellenberg GD, Sherman S, Simon P, Singh VP, Slivka A, Stolz D, Sutton R, Weiss FU, Wilcox CM, Zarnescu NO, Wisniewski SR, O'Connell MR, Kienholz ML, Roeder K, Barmada MM, Yadav D, Devlin B. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. Nat Genet. 2012 Dec;44(12):1349-54. doi: 10.1038/ng.2466. Epub 2012 Nov 11. PMID: 23143602; PMCID: PMC3510344.

Witt H, Luck W, Hennies HC, Classen M, Kage A, Lass U, Landt O, Becker M. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet. 2000 Jun;25(2):213-6. doi: 10.1038/76088. PMID: 10835640.

Witt H, Becker M. Genetics of chronic pancreatitis. J Pediatr Gastroenterol Nutr. 2002 Feb;34(2):125-36. doi: 10.1097/00005176-200202000-00006. PMID: 11840029.

Witt H, Sahin-Tóth M, Landt O, Chen JM, Kähne T, Drenth JP, Kukor Z, Szepessy E, Halangk W, Dahm S, Rohde K, Schulz HU, Le Maréchal C, Akar N, Ammann RW, Truninger K, Bargetzi M, Bhatia E, Castellani C, Cavestro GM, Cerny M, Destro-Bisol G, Spedini G, Eiberg H, Jansen JB, Koudova M, Rausova E, Macek M Jr, Malats N, Real FX, Menzel HJ, Moral P, Galavotti R, Pignatti PF, Rickards O, Spicak J, Zarnescu NO, Böck W, Gress TM, Friess H, Ockenga J, Schmidt H, Pfützer R, Löhr M, Simon P, Weiss FU, Lerch MM, Teich N, Keim V, Berg T, Wiedenmann B, Luck W, Groneberg DA, Becker M, Keil T, Kage A, Bernardova J, Braun M, Güldner C, Halangk J, Rosendahl J, Witt U, Treiber M, Nickel R, Férec C. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. Nat Genet. 2006 Jun;38(6):668-73. doi: 10.1038/ng1797. Epub 2006 May 14. PMID: 16699518; PMCID: PMC2746914.

Witt H, Beer S, Rosendahl J, Chen JM, Chandak GR, Masamune A, Bence M, Szmola R, Oracz G, Macek M Jr, Bhatia E, Steigenberger S, Lasher D, Bühler F, Delaporte C, Tebbing J, Ludwig M, Pilsak C, Saum K, Bugert P, Masson E, Paliwal S, Bhaskar S, Sobczynska-Tomaszewska A, Bak D, Balascak I, Choudhuri G, Nageshwar Reddy D, Rao GV, Thomas V, Kume K, Nakano E, Kakuta Y, Shimosegawa T, Durko L, Szabó A, Schnúr A, Hegyi P, Rakonczay Z Jr, Pfützer R, Schneider A, Groneberg DA, Braun M, Schmidt H, Witt U, Friess H, Algül H, Landt O, Schuelke M, Krüger R, Wiedenmann B, Schmidt F, Zimmer KP, Kovacs P, Stumvoll M, Blüher M, Müller T, Janecke A, Teich N, Grützmann R, Schulz HU, Mössner J, Keim V, Löhr M, Férec C, Sahin-Tóth M. Variants in CPA1 are strongly associated with early onset chronic pancreatitis. Nat Genet. 2013 Oct;45(10):1216-20. doi: 10.1038/ng.2730. Epub 2013 Aug 18. PMID: 23955596; PMCID: PMC3909499.

Xie Z, Hu S, Blackshaw S, Zhu H, Qian J. hPDI: a database of experimental human protein-DNA interactions. Bioinformatics. 2010 Jan 15;26(2):287-9. doi: 10.1093/bioinformatics/btp631. Epub 2009 Nov 9. PMID: 19900953; PMCID: PMC2804296.

Xuan S, Borok MJ, Decker KJ, Battle MA, Duncan SA, Hale MA, Macdonald RJ, Sussel L. Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. J Clin Invest. 2012 Oct;122(10):3516-28. doi: 10.1172/JCI63352. Epub 2012 Sep 24. PMID: 23006325; PMCID: PMC3461916.

Zhou J, Sahin-Tóth M. Chymotrypsin C mutations in chronic pancreatitis. J Gastroenterol Hepatol. 2011 Aug;26(8):1238-46. doi: 10.1111/j.1440-1746.2011.06791.x. PMID: 21631589; PMCID: PMC3142265.

Zou WB, Boulling A, Masamune A, Issarapu P, Masson E, Wu H, Sun XT, Hu LH, Zhou DZ, He L, Fichou Y, Nakano E, Hamada S, Kakuta Y, Kume K, Isayama H, Paliwal S, Mani KR, Bhaskar S, Cooper DN, Férec C, Shimosegawa T, Chandak GR, Chen JM, Li ZS, Liao Z. No Association Between CEL-HYB Hybrid Allele and Chronic Pancreatitis in Asian Populations. Gastroenterology. 2016 Jun;150(7):1558-1560.e5. doi: 10.1053/j.gastro.2016.02.071. Epub 2016 Mar 3. PMID: 26946345; PMCID: PMC5380763.

Zou WB, Wang YC, Ren XL, Wang L, Deng SJ, Mao XT, Li ZS, Liao Z. TRPV6 variants confer susceptibility to chronic pancreatitis in the Chinese population. Hum Mutat. 2020 Aug;41(8):1351-1357. doi: 10.1002/humu.24032. Epub 2020 Jun 24. PMID: 32383311.

List of Publications and Manuscripts in Preparation

Published

Variants That Affect Function of Calcium Channel TRPV6 Are Associated With Early-Onset Chronic Pancreatitis.

Masamune A, Kotani H, Sörgel FL, Chen JM, Hamada S, Sakaguchi R, Masson E, Nakano E, Kakuta Y, Niihori T, Funayama R, Shirota M, Hirano T, Kawamoto T, Hosokoshi A, Kume K, Unger L, Ewers M, Laumen H, Bugert P, Mori MX, Tsvilovskyy V, Weißgerber P, Kriebs U, Fecher-Trost C, Freichel M, Diakopoulos KN, Berninger A, Lesina M, Ishii K, Itoi T, Ikeura T, Okazaki K, Kaune T, Rosendahl J, Nagasaki M, Uezono Y, Algül H, Nakayama K, Matsubara Y, Aoki Y, Férec C, Mori Y, Witt H, Shimosegawa T. Gastroenterology. 2020 May;158(6):1626-1641.e8. doi: 10.1053/j.gastro.2020.01.005. Epub 2020 Jan 10. PMID: 31930989.

Under review

Variants in the pancreatic CUB and zona pellucida-like domains 1 (CUZD1) gene are associated with early-onset chronic pancreatitis

Agnieszka Magdalena Rygiel, Lara Sophie Unger, Emmanuelle Masson, Ryotaro Matsumoto, Maren Ewers, Jian-Min Chen, Peter Bugert, Louis Buscail, Tomasz Gambin, Grzegorz Oracz, Maria Winiewska-Szajewska, Agnieszka Mianowska, Jarosław Poznanski, Joanna Kosińska, Piotr Stawinski, Rafał Płoski, Dorota Koziel, Stanisław Gluszek, Helmut Laumen , Fredrik Lindgren , J. Matthias Löhr , Anna Orekhova , Vinciane Rebours , Jonas Rosendahl, Andrea Párniczky, Péter Hegyi, Akira Sasaki, Fumiya Kataoka, Yu Tanaka, Shin Hamada, Miklós Sahin-Tóth, Eszter Hegyi, Claude Férec, Atsushi Masamune, Heiko Witt

In preparation

Investigation of the pathogenic mechanism underlying the common SPINK1 p.N34S pancreatitis risk haplotype.

Lara Sophie Unger, … , Heiko Witt

Appendix

Post-affinity Chromatography control EMSA

Control aliquots of the RX (all nuclear proteins), SN-cplx (all unbound nuclear proteins) and elution fractions (nuclear proteins eluted from ds-oligonucleotides at the given salt concentration) were placed aside and kept at 4°C along the AC procedure. For the EMSA control run, 1.5 µL of the RX and 2 µL of the SN-cplx were combined with 10 µL of washing buffer (WB), whereas 7 µL of the elution fractions were combined with 15 µL of 1x BB. Finally, 1.5 µL of the Cy5-labeled ds-oligonucleotides (1 ng/µL) of the non-risk or risk allele were added to the corresponding reaction. All following steps were performed as outlined for the EMSA baseline conditions in the experimental procedures.

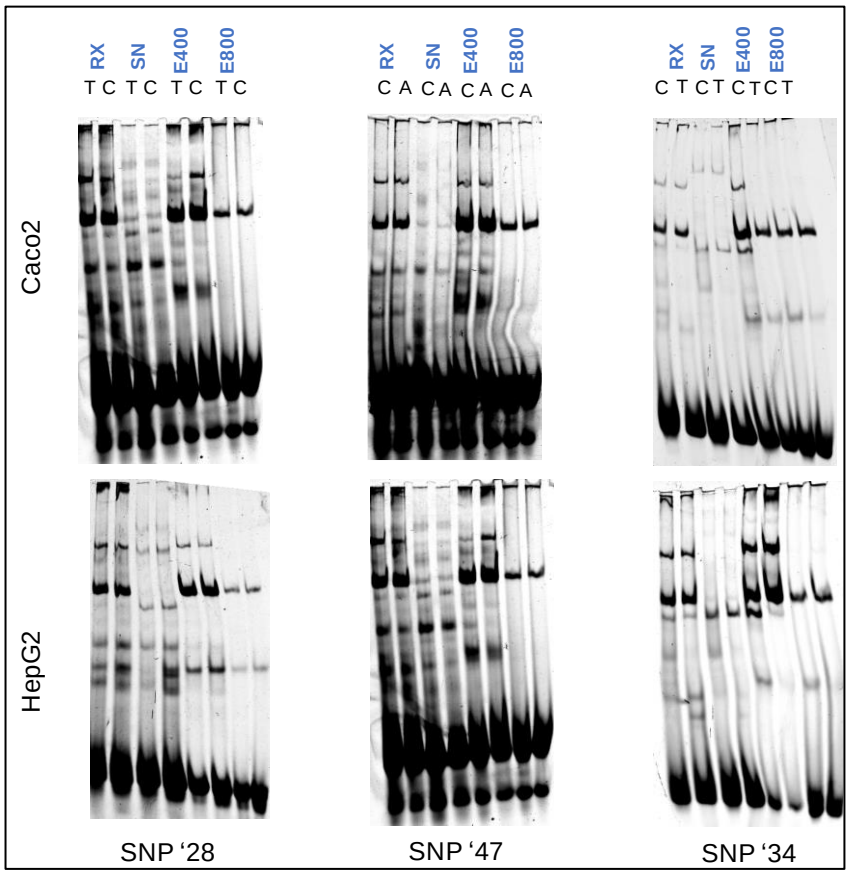


Figure 28 Nuclear protein binding of Caco2 and HepG2 extracts to rs148276928, rs142703147 and rs148911734 after affinity chromatography.

EMSA was performed with nuclear protein of Caco2 and HepG2 after affinity chromatography experiments (n=4, representative shown). The binding reaction with the non-risk allele was applied first, followed by the risk allele RX: Nuclear proteins; SN: unbound and unspecific DNA-binding proteins in supernatant after the addition of polydIdC to the mixture of RX and oligonulceotides; E400: elution at 400 mM NaCl; E800: elution at 800 mM NaCl; SNP’28: rs148276928 c.1-2090T>C; SNP’47 *rs142703147* C>A; SNP’34: rs148911734C>T

Description of the sample processing and data evaluation of the proteomic analysis by Dr. Christine von Törne Helmholtz Zentrum München, German Research Center for Environmental Health, Research Unit Protein Science.

Sample preparation for proteomics

Each 10 µg protein in 0.1x RIPA buffer were subjected to tryptic digest applying a modified filter aided sample preparation (FASP) procedure (Grosche *et al*. 2015; Wiśniewski *et al*. 2009). After protein reduction and alkylation using DTT and iodoacetamide, samples were denatured in UA buffer (8 M urea in 0.1 M Tris /HCl pH 8.5), centrifuged on a 30 kDa cut-off filter device (Sartorius), and washed three times with UA buffer and twice with50 mM ammoniumbicarbonate (ABC). Proteins were proteolysed for 2 h at room temperature using 0.5 µg Lys-C (Wako Chemicals, Neuss, Germany) and subsequently for 16 h at 37°C using 1 µg trypsin (Promega, Mannheim, Germany). Peptides were collected by centrifugation (10 min at 14 000 g) and acidified with 0.5% trifluoroacetic acid (TFA) and stored at -20°C.

Grosche A, Hauser A, Lepper MF, Mayo R, von Toerne C, Merl-Pham J, Hauck SM. The Proteome of Native Adult Müller Glial Cells From Murine Retina. Mol Cell Proteomics. 2016 Feb;15(2):462-80. doi: 10.1074/mcp.M115.052183. Epub 2015 Aug 31. PMID: 26324419; PMCID: PMC4739667.

Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009 May;6(5):359-62. doi: 10.1038/nmeth.1322. Epub 2009 Apr 19. PMID: 19377485.

Mass spectrometric measurements

LC-MSMS analysis was performed in data-dependent acquisition (DDA) mode. MS data were acquired on a Q-Exactive HF-X mass spectrometer (Thermo Scientific) each online coupled to a nano-RSLC (Ultimate 3000 RSLC; Dionex). Tryptic peptides were automatically loaded on a C18 trap column (300 µm inner diameter (ID) × 5 mm, Acclaim PepMap100 C18, 5 µm, 100 Å, LC Packings) at 30 µL/min flow rate. For chromatography, a C18 reversed phase analytical column (nanoEase MZ HSS T3 Column, 100Å, 1.8 µm, 75 µm x 250 mm, Waters) at 250nl/min flow rate in a 95 minutes non-linear acetonitrile gradient from 3 to 40% in 0.1% formic acid was used. The high-resolution (60 000 full width at half-maximum) MS spectrum was acquired with a mass range from 300 to 1500 m/z with automatic gain control target set to 3×10^6 and a maximum of 30 ms injection time. From the MS prescan, the 15 most abundant peptide ions were selected for fragmentation (MSMS) if at least doubly charged, with a dynamic exclusion of 30 seconds. MSMS spectra were recorded at 15 000 resolution with automatic gain control target set to 5×10^2 and a

maximum of 50 ms injection time. The normalized collision energy was 28, and the spectra were recorded in profile mode.

Data Processing – Protein Identification

Proteome Discoverer 2.3 software (Thermo Fisher Scientific; version 2.3.0.523) was used for peptide and protein identification via a database search (Sequest HT search engine) against Swissprot human data base (Release 2020_02, 20432 sequences), considering full tryptic specificity, allowing for up to two missed tryptic cleavage sites, precursor mass tolerance 10 ppm, fragment mass tolerance 0.02 Da. Carbamidomethylation of Cys was set as a static modification. Dynamic modifications included deamidation of Asn and Gln, oxidation of Met; and a combination of Met loss with acetylation on protein N-terminus. Percolator was used for validating peptide spectrum matches and peptides, accepting only the top-scoring hit for each spectrum, and satisfying the cutoff values for FDR <5%. The final list of proteins complied with the strict parsimony principle.

Data processing – Label-free quantification

The quantification of proteins was based on the area of the abundance values for unique plus razor peptides. Abundance values were normalized in a retention time dependent manner to account for sample loading errors. The protein abundances were calculated summing up the abundance values for admissible peptides. The final protein ratio was calculated using median abundance values of four to five replicate analyses each. The statistical significance of the ratio change was ascertained employing the T-test approach described in (Navarro *et al.*, 2014) which is based on the presumption that we look for expression changes for proteins that are just a few in comparison to the number of total proteins being quantified. The quantification variability of the non-changing "background" proteins can be used to infer which proteins change their expression in a statistically significant manner.

General statistical framework for quantitative proteomics by stable isotope labeling.
Navarro P, Trevisan-Herraz M, Bonzon-Kulichenko E, Núñez E, Martínez-Acedo P, Pérez-Hernández D, Jorge I, Mesa R, Calvo E, Carrascal M, Hernáez ML, García F, Bárcena JA, Ashman K, Abian J, Gil C, Redondo JM, Vázquez J. J Proteome Res. 2014 Mar 7;13(3):1234-47. doi: 10.1021/pr4006958. Epub 2014 Feb 10. PMID: 24512137

Endogenous expression levels of *GATA6*, *GATA4* and *YY1*

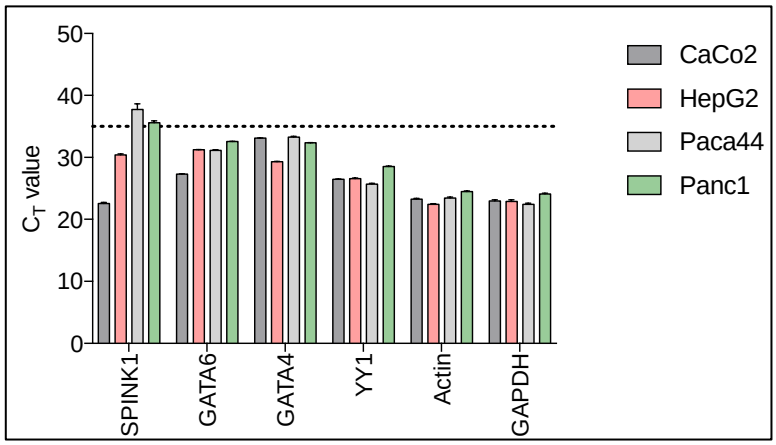


Figure 29 Endogenous mRNA expression of GATA6, GATA4 and YY1 in different cell lines.

Total RNA was collected and reverse transcribed. mRNA expression of the indicated target genes was measured by qPCR with TaqMan assays (technical triplicates). Expression of β-Actin and GAPDH served as reference. SPINK1: serine protease inhibitor Kazal type 1, GATA: GATA binding protein, YY1: Ying Yang, Actin: β-actin and GAPDH: Glyceraldehyde 3-phosphate dehydrogenase; cut off: C_T value > 35 cycles (dotted line).

Relative mRNA expression levels of selected inflammatory markers and *HSPA5* after overexpression of acinar transcription factors in Panc1

Table 33 TaqMan Assays for inflammatory markers

Name	Nucleotide Sequence 5'-3'
CCL2 S	GCTCAGCCAGATGCAATC
CCL2 R	CCACAATGGTCTTGAAGATCACA
CCL2 P	F-CACCTGCTGTTATAACTTCACCAATAGGAA--Q
CFD F	GGACAGCCAGCCCGACA
CFD A	GGCCTTCTCCGACAGCTGTA
CFD P	F-CAGCAGGAGGTCGTGGTCGATG--Q
C1QBP S	TGACATCTTCTCTATCAGGGAAGTTAG
C1QBP R	TTAGGTGGTCATATAAGGCCCAG
C1QBP P	F-CAGTCCACTGGCGAGTCTGAATGG--Q
IL18 F	CAGTCTACACAGCTTCGGGA
IL18 R	CACAAAGTTGATGCAATTGTCTTCT
IL18 P	F-AGATCGCTTCCTCTCGCAACAAACTA--Q
TNFRSF12A S	CTTGGGGGCGCTCTGAG
TNFRSF12A A	CTCTATGGGGGTGGTGAACTTCT
TNFRSF12A P	F-ATCGTCTCCAGACCAAAAAGCCAGAAA--Q

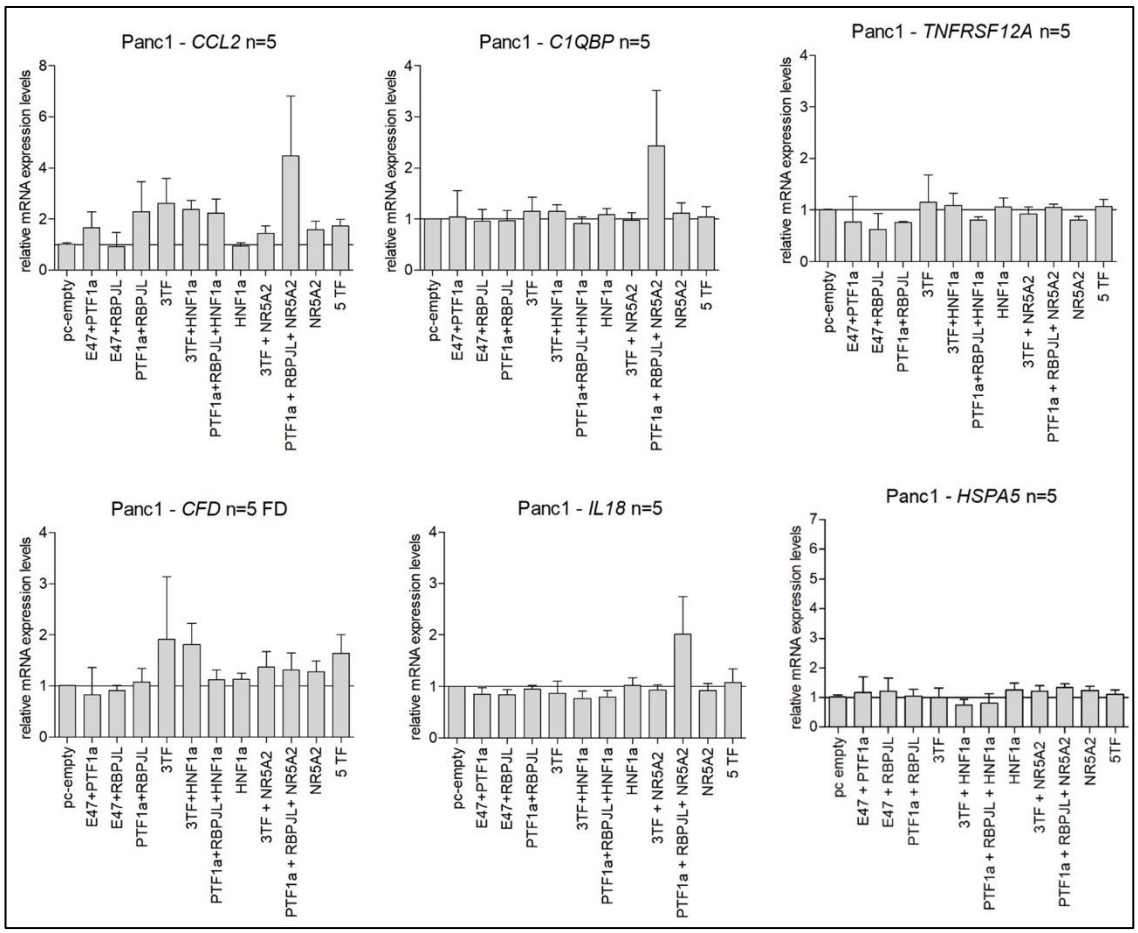


Figure 30 Inflammatory gene expression upon the overexpression of acinar key transcription factors.

Total RNA was collected 72 h post transfection and transcribed to cDNA. Relative mRNA expression of the indicated target genes was measured by qPCR with TaqMan assays (technical triplicates). Expression of β-Actin and GAPDH served as internal controls. Expression levels were normalized to control transfected Panc1 cells (pc empty). The horizontal line illustrates the baseline relative mRNA expression (=1). 3TF represents the trimeric PTF1 complex consisting of E47 (TCF3), PTF1a, and RBPJL. 5TF represents the combined overexpression of E47 (TCF3), PTF1a, RBPJL, HNF1a and NR5A2. Cell line, target gene and number of biological replicates are given above each graph. Standard deviation is shown. TCF3: Transcription factor 3 (E47), PTF1a: Pancreas associated transcription factor 1a, RBPJL: recombination signal binding protein for immunoglobulin kappa J region -like, HNF1a: Hepatocyte nuclear factor 1 homeobox A, NR5A2: nuclear receptor subfamily 5 group A member 2, CCL2:C-C motif chemokine ligand 2, C1QBP: complement C1q binding protein, TNFRSF12A: TNF receptor superfamily member 12A, CFD: complement factor D, IL18: interleukin 18, HSPA5: heat shock protein family A (Hsp70) member 5

Effect of GATA4, GATA6 and HEYL on selected target genes in Panc1

We set out to assess the effect of GATA4, GATA6 and HEYL on pancreatic lipase (PNLIP), pancreatic lipase related protein 1 (PNLIPRP1) and phospholipase A2, group 1B (PLA2G1B). For the first overexpression experiment, GATA4, GATA6 and HEYL were transfected separately in Panc1 cells (Figure 28). For the second experiment, the PTF1 complex (PTF1a, RBPJL and E47) and the PTF1 complex combined with GATA4, GATA6 or HEYL were transfected (Figure 29 (top)) and in the last experiment, the transcription factor HNF1a was added to each condition of the second experiment (Figure 29 (bottom)). The transfection experiments and qPCR measurements using TaqMan assays were carried out in independent biological as well as technical replicate (n=3), respectively. Total RNA was collected 72h after transfection and reverse transcribed. Expression levels were normalized to control transfected Panc1 cells (pc empty). The horizontal line illustrates the baseline relative mRNA expression (=1). 3TF represents the trimeric PTF1 complex consisting of E47 *(TCF3)*, PTF1a, and RBPJL. The experiments were conducted by Melanie Stammler as part of her Bachelor thesis.

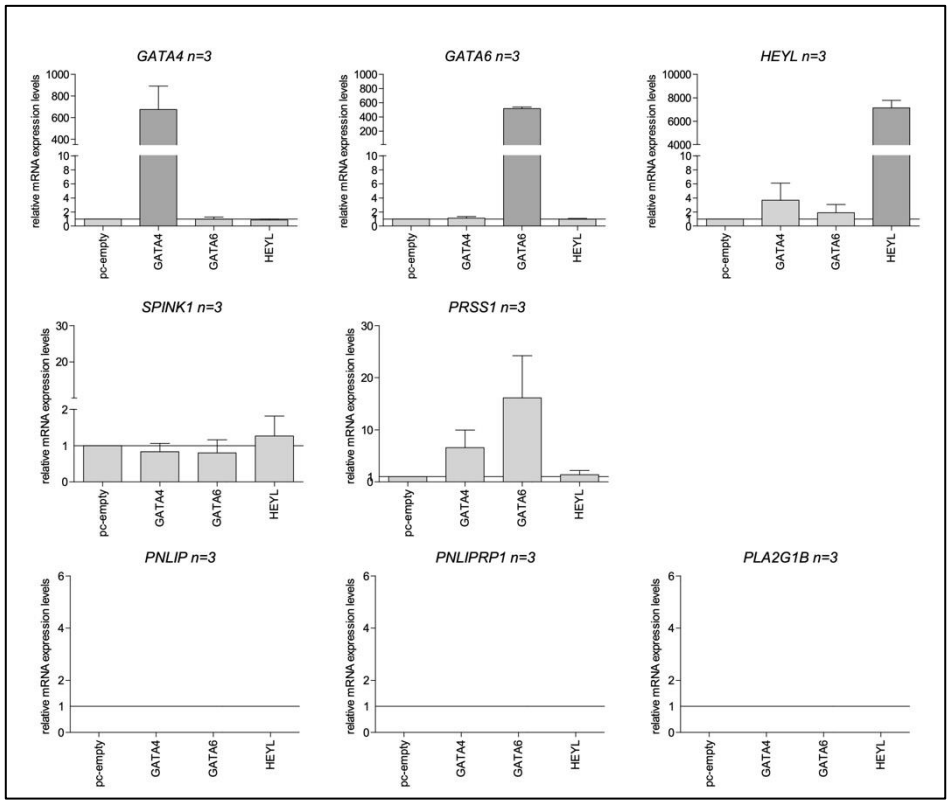


Figure 31 Target gene expression in Panc1 after overexpression of GATA4, GATA6, HEYL as single factors

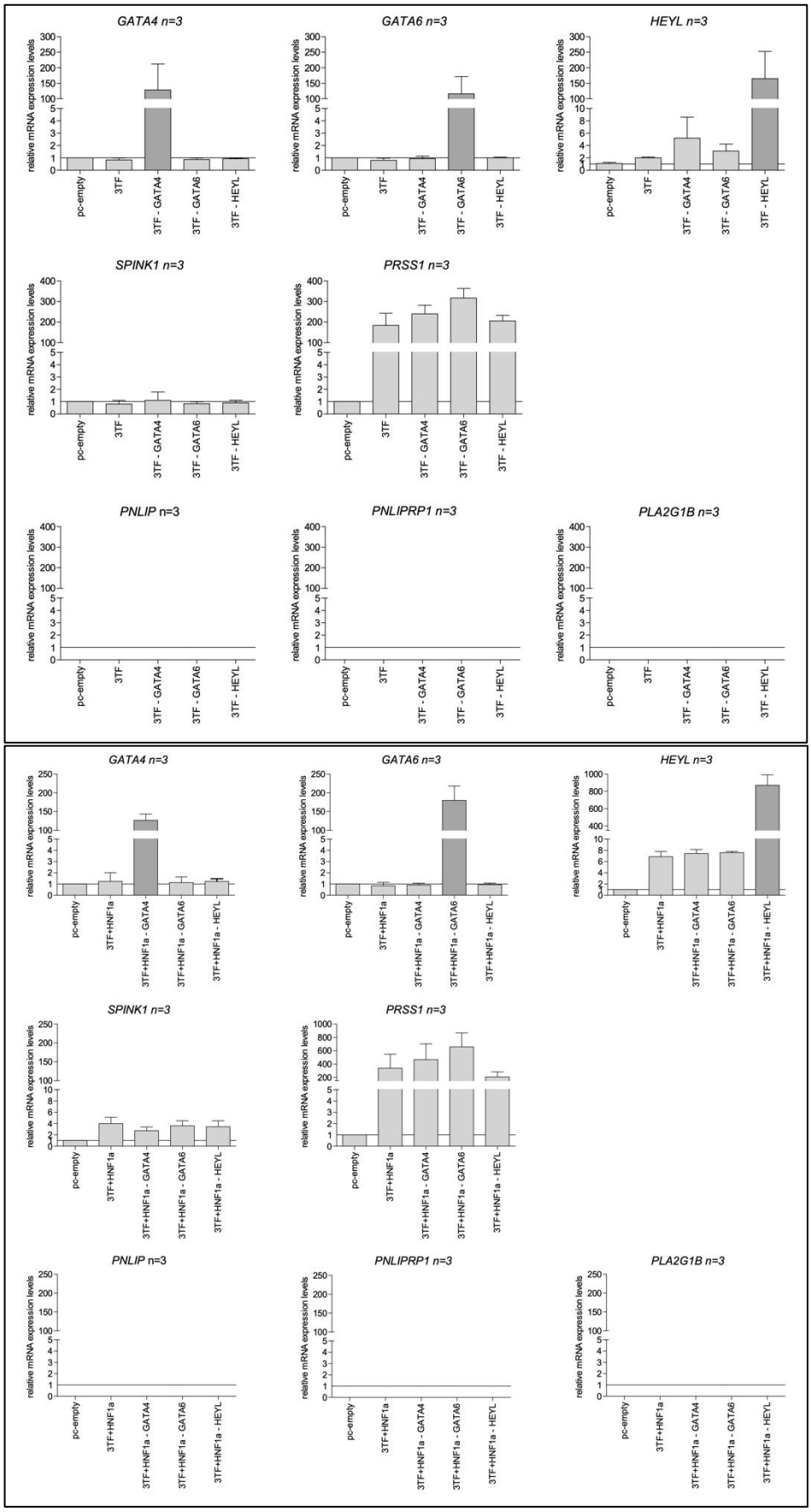


Figure 32 Target gene expression Panc1 after the combined overexpression of the PTF1 complex (top) and HNF1a (bottom) and either GATA4, GATA6, HEYL

Determination of the human pancreatic RBPJL isoform (RBPJL-203)

Concept

For the determination of human pancreatic RBPJL isoform, we followed a two-step approach. In a first step, an isoform-unspecific PCR product was amplified, using primers in the flanking region of interest. In a second step, we sequenced the PCR product in two different approaches. For the first sequencing approach, we used an isoform unspecific sequencing primer which provided the nucleotide sequence of the pre-dominantly expressed isoform in the human tissue. Subsequently, we compared the obtained sequence to the reference sequences of all three isoforms to identify the matching one. In a second sequencing approach, we used isoform-specific sequencing primers. Besides the predominately expressed isoform, this would also indicate the presence of less dominantly expressed isoforms. To control for primer efficiency and specificity, we used the RBPJL-vectors (*Genscript*) as template for the PCR and sequencing.

Experimental Procedure

The human cDNA was sampled from 9 individuals and was a kind gift of Prof. Dr. Dr. Ihsan Ekin Demir (Chirurgische Klinik, Klinikum rechts der Isar, TUM). The over-expression vectors were ordered from Genscript (Piscataway, USA). The PCR reaction was composed in a total volume of 22 µL as follows: 2 µL DNA template (10ng of human cDNA mix of 9 individuals or vector DNA), 0.4 µL of the forward and the reverse primer (10 µM, TIB MOLBIOL, Berlin, Germany), 10 µL of One-Tag® 2x Master Mix (NEB, Ipswich, USA), 9.2 µL of water The PCR conditions were set to an initial denaturation at 95°C, followed by 50 cycles of denaturation at 95°C for 30 sec, primer annealing at 60°C for 40sec and primer extension at 68°C for 2 min, terminating with a final elongation step at 68°C for 5 min. Of the final PCR product, 4 µL were mixed with 0.8 µL of DNA Gel Loading Dye (6x) (Thermo Fisher Scientific, Waltham, USA) and loaded onto a 1% agarose gel. Gel electrophoresis was performed at 100 V for 75 min and the DNA bands were imaged under UV light. PCR products were purified in a PCR clean up reaction adding 0.25 µL of Antarctic Phosphatase (NEB, Ipswich, USA) 0.25 µL Exonuclease I (NEB, Ipswich, USA) and 1.5 µL water to the PCR product (18 µL). The enzymatic digestion was performed at 37°C for 40 min followed by enzyme inactivation at 85°C for 20 min. Each of the RBPJL PCR products was sent to sequencing with the isoform-unspecific sequencing primer (RBPJL-Q1112F). The PCR products form the vector templates were sent with their corresponding isoform specific sequencing primer (positive control) whereas the PCR products of the human template were sent with all three isoform specific sequencing primers (RBPJL-S1-201, RBPJL-S1-202, RBPJL-S1-203). Each sequencing reaction included 5 µL of the PCR product, 10 µL of water and 2 µL of the corresponding primer (10 µM). Sanger sequencing was performed at

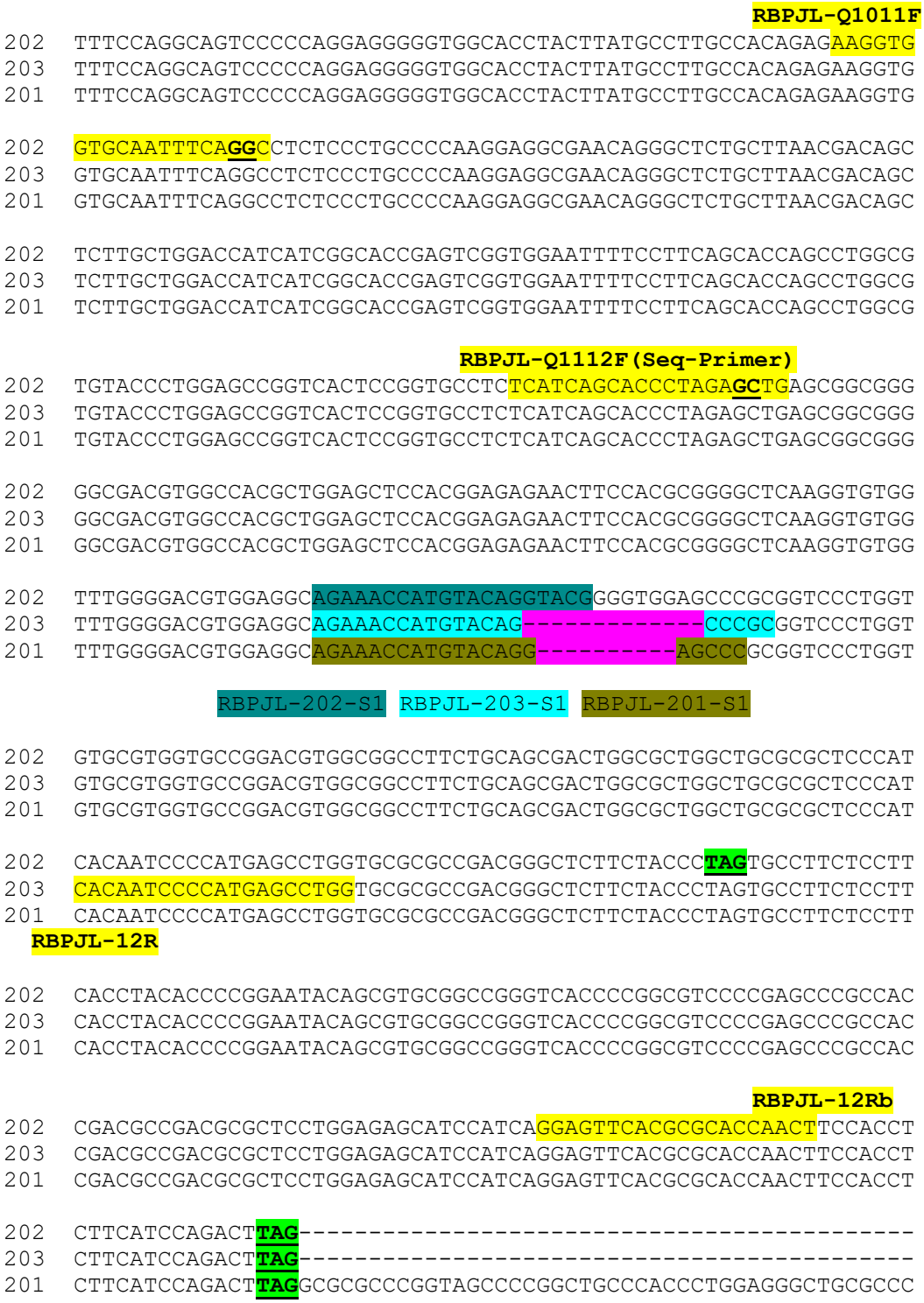
Eurofins Genomics (Ebersberg, Germany). An overview of the relevant primers is provided in the table below. For all reactions, ROTISOLVE® HPLC Gradient Grade water (Roth, Karlsruhe, Germany) was used.

Primer for the identification of the human pancreatic RBPJL isoform

Name	Nucleotide Sequence 5'-3'
RBPJL-Q1011F	AAGGTGGTGCAATTTCAGGC
RBPJL-12R	CCAGGCTCATGGGGATTGTG
RBPJL-12Rb	AGTTGGTGCGCGTGAACTCC
RBPJL-Q1112F	TCATCAGCACCCTAGAGCTG
RBPJL-S1-201	AGAAACCATGTACAGGTACG
RBPJL-S1-202	AGAAACCATGTACAGCCCGC
RBPJL-S1-203	AGAAACCATGTACAGGAGCCC

Note: The two nucleotides forming the exon-exon border are underlined.

Illustration of primer positioning



Company Protocols

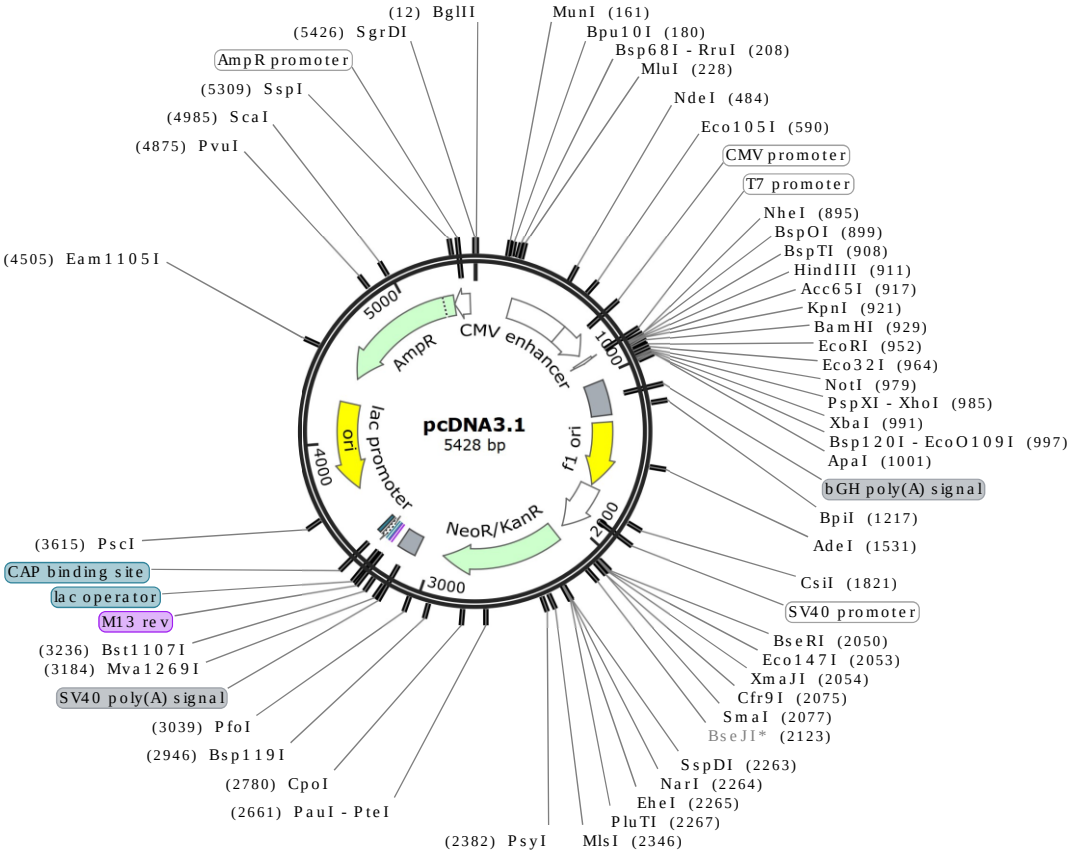
- Protocol-Human SPINK1 DuoSet ELISA Kit-**R&D**
- Protocol-Nuclear Extraction Kit-**Active-Motif**
- Protocol-QIAamp DNA Kit-**Qiagen**
- Protocol-QuantiTect® Reverse Transcription Kit**-Qiagen**
- Protocol-RNeasy Kit-**Qiagen**
- Protocol-Wizard SV Gel and PCR Clean up System-**Promega**

Vector-Maps

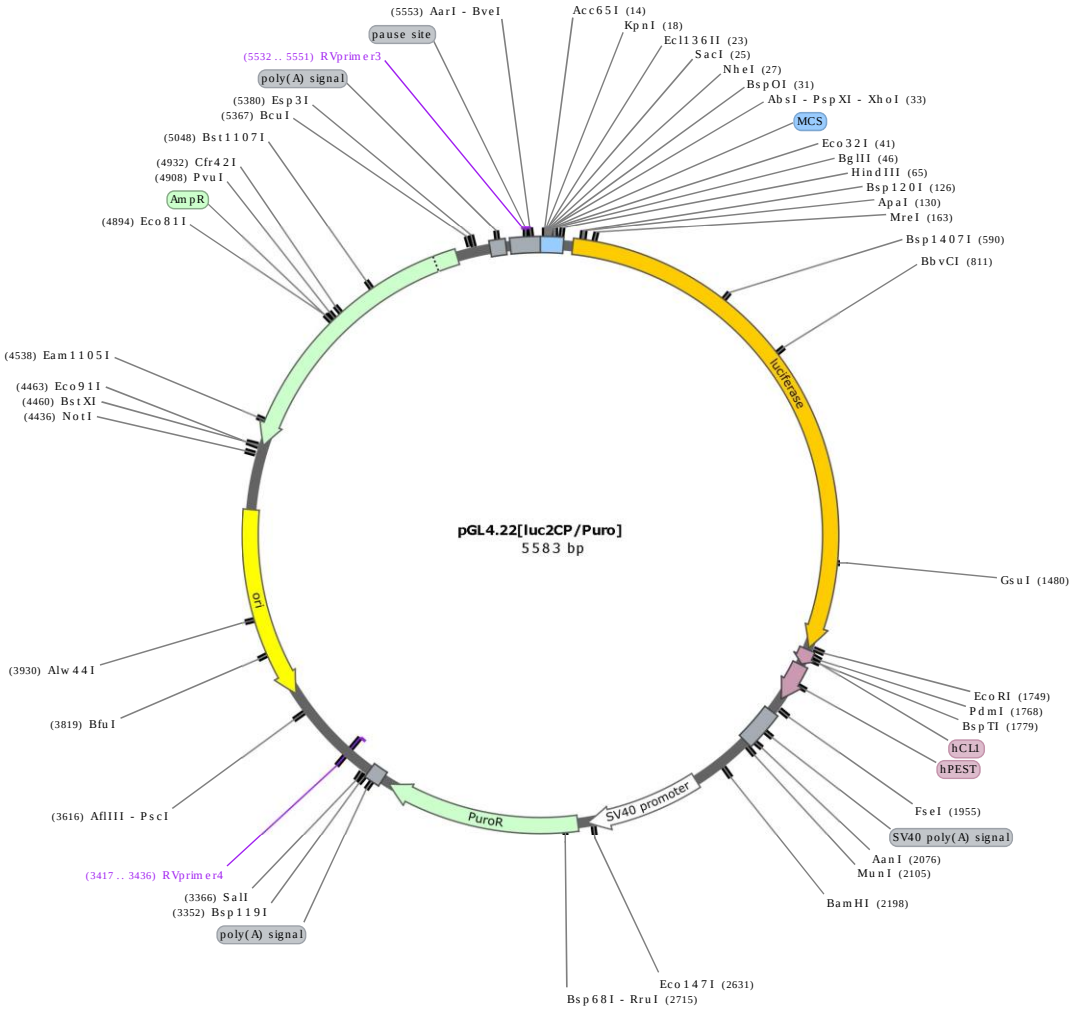
- pcDNA3.1+/C-(K)DYK
- (1) pRL-Ubi
- (2) pGL4.22_SPINK1-Promotor_Luiferase
- (3) pGL4.22_SPINK1-Promotor _flank-rs148911734-wt(C) 500bp_Luciferase
- (4) pGL4.22_SPINK1-Promotor _flank-rs148911734-mt(T)-500bp_Luciferase

Sequence: pcDNA3.1.dna (Circular / 5428 bp)
Enzymes: Unique 6+ Cutters (52 of 653 total)
Features: 16 visible, 16 total

Unique Cutters Bold

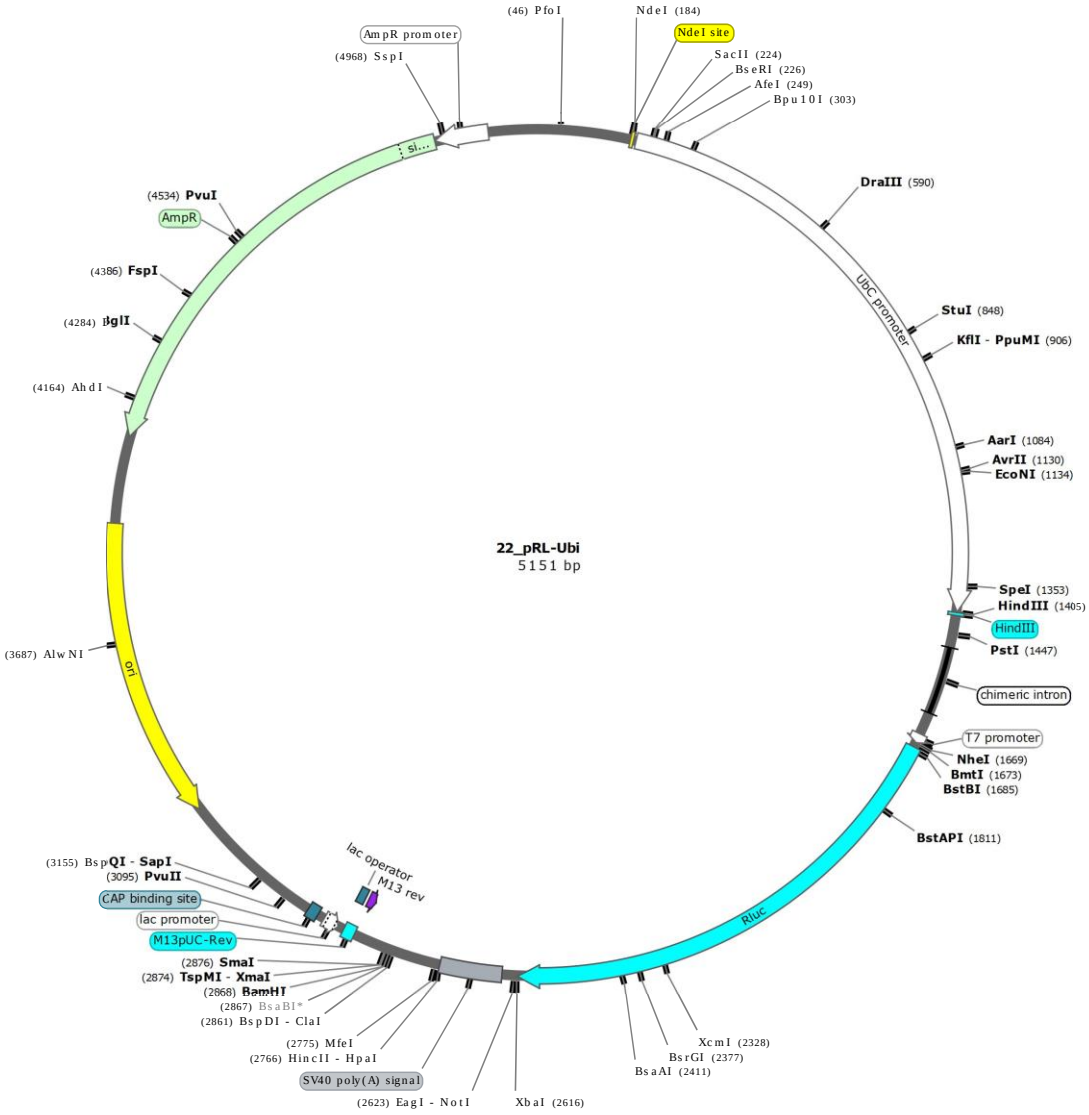


Promoterless vector encoding highly destabilized luciferase for measuring the activity of promoter and enhancer sequences.



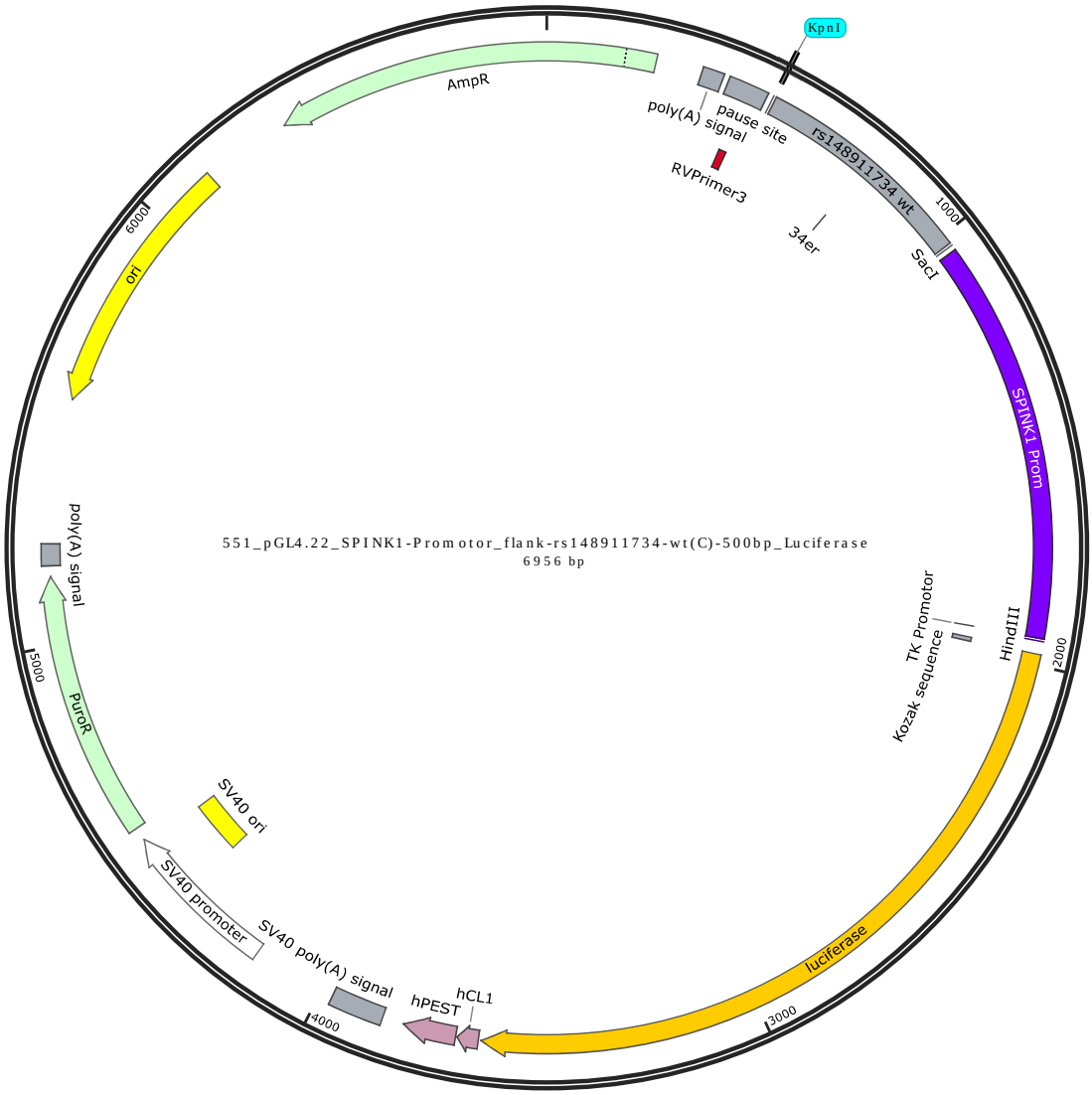
Sequence: 22_pRL-Ubi.dna (Circular / 5151 bp)
Enzymes: Unique 6+ Cutters (45 of 666 total)
Features: 15 total

Unique Cutters Bold




549_pGL4.22_SPINK1-Promotor_Luiferase
6459 bp
poly(A) signal
SacI
pause site
XhoI
KpnI
RVPrimer3
SPINK1 Prom
Kozak sequence
HindIII
R bea Jurc
luciferase
hCL1
hPEST
SV40 poly(A) signal
SV40 promoter
SV40 ori
PuroR
poly(A) signal
ori
AmpR
1000
2000
3000
4000
5000
6000

Sequence: 550_pGL4.22_SPINK1-Promotor _flank-rs148911734-mut(T)-500bp_Luciferase.dna (Circular / 6956 bp)
Features: 21 total


KpnI
AmpR
poly(A) signal
pause site
rs148911734 wt
RVPrimer3
34er
SacI
SPINK1 Prom
HindIII
Tk Promotor
Kozak sequence
luciferase
hCL1
hPEST
SV40 poly(A) signal
SV40 promoter
SV40 ori
PuroR
poly(A) signal
ori
1000
2000
3000
4000
5000
6000
551_pGL4.22_SPINK1-Promotor_flank-rs148911734-wt(C)-500bp_Luciferase
6956 bp

XL