Technische Universität München
TUM School of Life Sciences

Identification and functional analysis of genetic risk factors in idiopathic and hereditary chronic pancreatitis

Katharina Maria Elisabeth Eiseler

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

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Martin Klingenspor

Prüfer der Dissertation:

1. Prof. Dr. Heiko Witt
2. Prof. Dr. Maximilian Reichert
3. Prof. Dr. Robert Jaster

Die Dissertation wurde am 05.06.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 01.12.2023 angenommen.

Acknowledgements

Zuallererst möchte ich mich bei **Prof. Dr. Heiko Witt** bedanken. Danke für deine herausragende wissenschaftliche Betreuung sowie das entgegengebrachte Vertrauen in meine Arbeit. Ich weiß deine stets offene Tür sehr zu schätzen und blicke mit Freude und Dankbarkeit zurück an die vielen Stunden des produktiven Austauschs.

Ein großer Dank gilt auch meiner Mentorin **Maren Ewers**, die mir mit unendlicher Geduld und Wissen immer mit Rat und Tat zu Seite stand. Danke, für deine positive Einstellung und dein stets offenes Ohr. Du bist ein Vorbild für mich.

Der Dank gilt darüber hinaus der **gesamten Arbeitsgruppe**. Die gegenseitige Hilfsbereitschaft und Unterstützung im Labor sind nicht selbstverständlich und haben dadurch ein überaus familiäres und freundschaftliches Umfeld erzeugt. Durch das Teilen von Freude, Erfolg, Sorgen und Problemen sind Arbeitskollegen zu Freunden geworden. Die letzten Jahre werde ich daher in bester Erinnerung halten und möchte die schöne gemeinsame Zeit am Lehrstuhl keinesfalls missen.

Ganz besonders möchte ich mich bei meinen Eltern **Elisabeth und Albert Burreiner** bedanken. Danke, Mama und Papa, dass ihr mir diesen Weg ermöglicht habt. Ohne euch wäre ich nicht, wo ich heute bin. Für eure bedingungslose Liebe und Unterstützung, sowohl emotional als auch finanziell, bin ich euch auf immer dankbar.

Ein weiterer Dank geht an meinen Ehemann und besten Freund **Jens Eiseler**. Danke für deine Unterstützung und Geduld, dass du mir in den letzten Jahren bei vielen Höhen und Tiefen zur Seite standst. In dir habe ich einen lebenslangen Weggefährten gefunden, auf den ich mich immer verlassen kann.

Table of Contents

Acknowledgements II
Table of Contents III
Abstract VI
Zusammenfassung VII
List of figures VIII
List of tables IX
List of Abbreviations XI
1 Introduction 1
1.1 The pancreas and pancreatic organogenesis 1
1.2 Chronic pancreatitis: epidemiology of pancreatitis worldwide 2
1.3 Genetics of chronic pancreatitis: known risk factors 2
1.4 Water channels in the exocrine pancreas 7
1.5 CTRL as a new risk candidate in the panel of digestive enzymes 8
1.6 Transcription regulation in the pancreas: a complex interplay 9
1.6.1 Transcription factors as potential risk genes for pancreatitis 10
2 Aim of the thesis 13
3 Material 14
3.1 Chemicals 14
3.2 Equipment and Consumables 16
3.3 Buffers and solutions 18
3.4 Cell lines 21
3.5 Vector plasmids 21
3.5.1 Human expression vectors 21
3.5.2 Reporter gene vectors (luciferase vectors) 22
3.6 Primers 23
3.7 TaqMan assays for real-time polymerase chain reaction (qPCR) 23
3.8 Enzymes 23
3.9 Antibodies 24
3.10 Oligonucleotides for EMSA 25
3.11 Company-provided kits 25
4 Study cohort 26
4.1 Patient cohort 26
4.2 Whole Exome Sequencing 26
4.3 Control cohorts 26
5 Experimental procedures 27
5.1 Genetic verification of mutations 27
5.1.1 Sanger sequencing 27
5.1.2 Statistical analysis of AQP12 28
5.2 Vector plasmid preparations 28

5.2.1 Human expression vector preparation via MaxiPrep 28
5.2.2 Molecular Cloning of target promoters into a reporter gene vector 29
5.2.3 Primer mutagenesis to introduce a *PacI* site into the pcDNA3.1+/C-(K)-DYK vector 32
5.2.4 Primer mutagenesis to reposition the Flag-tag in the NR5A2 wildtype vector 33
5.2.5 Primer mutagenesis for removal of a Flag-tag in expression vectors 33
5.3 Cell culture experiments 33
5.3.1 General culturing of cells 33
5.3.2 Transient overexpression for protein or RNA isolation 34
5.3.3 Endoplasmic reticulum stress induction 35
5.3.4 Transient overexpression for Luciferase assay 35
5.3.5 Cell culture for nuclear extraction 35
5.4 Sample preparations 36
5.4.1 Total lysate preparation 36
5.4.2 Protein precipitation from supernatant 36
5.4.3 Total RNA isolation 37
5.4.4 Reverse transcription of RNA into cDNA 37
5.4.5 Nuclear protein extraction 37
5.5 SDS-PAGE and Western blot 38
5.6 Activity assay 39
5.6.1 Enzyme activity 39
5.6.2 Enzyme activation 40
5.7 Reporter gene assay / Dual-luciferase assay 41
5.8 Target gene expression real-time polymerase chain reaction 41
5.9 ER stress analysis 42
5.9.1 XBP1 splicing PCR 42
5.9.2 *BiP* expression qPCR 42
5.10 Electrophoretic mobility shift assay (EMSA) 42
5.10.1 Annealing of Cy5-labelled oligonucleotides 42
5.10.2 EMSA 43
6 Results 45
6.1 Genotyping for *AQP12A* and *AQP12B* 45
6.2 Chymotrypsin-like protease 48
6.2.1 Western blot analysis shows reduced secretion of C-terminal mutants 49
6.2.2 CTRL variants show reduced activity or loss of function 51
6.2.3 ER stress as possible consequence of CTRL mutations 52
6.2.4 Summary of functional analyses of CTRL 55
6.3 Transcription Factors 56
6.3.1 Verification of protein expression in Western blot 57
6.3.2 Molecular cloning of potential target gene promoter sequences 58
6.3.3 Luciferase Assay establishment with all promoters 59

6.3.4 Reporter gene assay on transcriptional activity....61
6.3.5 Target gene expression analysis with qPCR....66
6.3.6 Statistics....71
6.3.7 Protein-DNA binding capacity of NR5A2 in EMSA....72
7 Discussion....73
7.1 *AQP12* mutations are not associated with chronic pancreatitis....73
7.2 Mutations in *CTRL* do not show an association with chronic pancreatitis....75
7.2.1 Secretory dysfunction in CTRL mutants does not lead to ER stress....77
7.2.2 Mutant position determines various causes of catalytic defects....77
7.2.3 Mice studies are needed to further unravel the function of CTRL....79
7.3 Transcription factors....80
7.3.1 Mutations in the DNA binding domain of NR5A2 diminish its function....80
7.3.2 Attenuating effects of RBPJL mutations are not associated with CP....81
7.3.3 HNF1a variants do not alter the transactivation of target promoters....83
7.4 Chronic pancreatitis as a multigenic disease....84
7.5 Outlook: the mystery of genetic risk factors in CP remains....84
8 References....I
8.1 List of publications and manuscripts....IX
9 Supplementary Data....X
9.1 Primers....X
9.1.1 Primers for Sanger sequencing....X
9.1.2 Primers for molecular cloning....XII
9.1.3 PCR-mutagenesis primers....XIV
9.1.4 Primers for XBP1 splicing PCR....XIV
9.2 TaqMan assays for qPCR....XV
9.3 Synonymous *AQP12A* and *AQP12B* variants....XVII
9.4 Chymotrypsin-like protease....XIX
9.4.1 Western blot analyses of CTRL mutants....XIX
9.4.2 Activity assay of CTRL mutations with substrate AAPY....XX
9.4.3 Activation assay of CTRL mutations p.G20S and p.G56S....XX
9.5 Transcription factors....XXI
9.5.1 Flag-tag vs. untagged....XXI
9.5.2 Western blot analysis of all NR5A2 mutations....XXIII
9.6 Company provided protocols....XXIII

Abstract

In chronic pancreatitis (CP), a relapsing or continuous inflammation of the pancreas progresses in severe impairment of both endocrine and exocrine functions. Especially in an early onset, genetic factors play a critical role. Numerous risk factors have already been identified since the hypothesis of a genetic origin was first described in 1952. A disturbed balance of the protease pathway seems crucial in the pathogenesis. Besides genetic variants in the *PRSS1* locus, the two major protective mechanisms for pre-maturely activated trypsin, *SPINK1* and *CTRC,* have also been associated with CP. Other digestive enzymes but also ductally expressed genes joined the list of risk factors over the past decades. Nevertheless, a considerable proportion of non-alcoholic chronic pancreatitis (NACP) patients still have no identified underlying cause. This suggests the existence of other to date undiscovered disease-causing loci. Based on an extended candidate approach and whole exome sequencing (WES) data, we identified a gene panel which are expressed in the pancreas. To contribute to the clarification of pathogenic mechanisms in idiopathic CP (ICP), we focused on five genes to investigate different potential pathological pathways.

Aquaporin 12 (AQP12) is a water channel exclusively expressed in zymogen granules of acinar cells. It is suggested to play a role in their maturation and exocytosis. A disturbed water flow can promote accumulation and deformation of vacuoles, thus disturbing zymogen secretion. Our Sanger sequencing of 292 unrelated German CP patients and 143 German control subjects revealed no association of either homologue gene *AQP12A* and *AQP12B* with CP. We found several missense and nonsense mutations in the control cohort as well.
Based on an extended candidate approach, we investigated chymotrypsin-like protease (CTRL) as another digestive enzyme. We hypothesized that genetic alterations interfere with catalytic activity and secretion. As consequence, an intracellular retention of CTRL might lead to ER stress. Western blot and activity analyses shows to some extent severe interference in the protease function and secretion, however, there was no association with CP, as these effects were detectable in mutations found both patients and controls.
We further investigated the transcription factors NR5A2, RBPJL, and HNF1a involved in pancreas development and maintenance. In reporter gene assays and endogenous mRNA expression, we failed to tie mutations in *RBPJL* and *HNF1a* to CP. The investigation of *NR5A2* revealed a significant effect of two variants, p.R142H and p.G150R, on the transactivation on pancreas-specific target gene promoters by strongly reduced NR5A2 function. Future studies in mouse models carrying these variants will be needed to corroborate our results.

Based on burden testing of our WES data and subsequent functional analyses, we identified *NR5A2* as potential novel risk gene for the pathology of CP.

Zusammenfassung

Die chronische Pankreatitis (CP) ist eine rezidivierende oder kontinuierliche Entzündung des Pankreas, die zu einer schweren Beeinträchtigung der endokrinen als auch exokrinen Funktionen führt. Genetische Faktoren spielen insbesondere im frühen Alter eine entscheidende Rolle. Seit der ursprünglichen Hypothese eines genetischen Ursprungs 1952 konnten bereits zahlreiche Risikogene identifiziert werden. Ein Ungleichgewicht der Proteasen scheint entscheidend für die Pathogenese. Neben genetischen Varianten in *PRSS1* wurden auch die beiden wichtigsten Schutzmechanismen für frühzeitig aktiviertes Trypsin, *SPINK1* und *CTRC*, mit CP in Verbindung gebracht. Andere Verdauungsenzyme, aber auch duktal exprimierte Gene wurden in den letzten Jahrzehnten in die Liste der Risikofaktoren aufgenommen. Dennoch ist bei einem beträchtlichen Teil der Patienten noch immer keine zugrundeliegende Ursache bekannt. Dies lässt vermuten, dass es unentdeckte krankheitsverursachende Loci gibt. Auf der Grundlage eines erweiterten Kandidatenansatzes und anhand von Daten aus der *Whole Exome* Sequenzierung (WES) haben wir Gene identifiziert, die im Pankreas exprimiert werden oder dort regulierend wirken. Um zur Klärung der pathogenen Mechanismen bei idiopathischer CP (ICP) beizutragen, konzentrierten wir uns auf fünf Gene, um verschiedene potenzielle pathologische Pfade zu untersuchen.

Aquaporin 12 (AQP12) ist ein Wasserkanal, der ausschließlich in den Zymogen-Granula der Azinuszellen vorkommt. Es wird vermutet, dass er eine Rolle bei deren Reifung und Exozytose spielt. Ein gestörter Wasserfluss kann die Akkumulation und Deformation von Vakuolen fördern und damit die Sekretion stören. Sanger-Sequenzierung von 292 nicht verwandten deutschen CP-Patienten und 143 deutschen Kontrollpersonen ergab keinen Zusammenhang zwischen den beiden homologen Genen *AQP12A* und *AQP12B* und CP. Auch in der Kontrollkohorte fanden wir mehrere Missense- und Nonsense-Mutationen.

Auf der Grundlage eines erweiterten Kandidatenansatzes vermuteten wir, dass genetische Veränderungen im Verdauungsenzym Chymotrypsin-like Protease (CTRL) dessen katalytische Aktivität und Sekretion beeinträchtigen. Infolgedessen könnte eine intrazelluläre Retention zu ER Stress führen. Western Blot und Aktivitätsmessungen ergaben z.T. eine stark beeinträchtigte Funktion und Sekretion, jedoch bei Varianten, die sowohl in Patienten als auch bei Kontrollen gefunden wurden.

Darüber hinaus untersuchten wir pankreas-spezifische Transkriptionsfaktoren NR5A2, RBPJL und HNF1a. Mittels Reportergen-Assays und endogener mRNA-Expression konnten wir *RBPJL* und *HNF1a* als Risikogene ausschließen. Die Untersuchung von NR5A2 ergab jedoch einen signifikanten Funktionsverlust in zwei Varianten, p.R142H und p.G150R, auf die Transaktivierung an pankreas-spezifischen Zielgenpromotoren.

Aufgrund vorangegangener *Burden Tests* sowie unserer funktionellen Daten konnten wir *NR5A2* als potenzielles Risikogen der CP identifizieren.

List of figures

Figure 1.1: Pancreatitis-related genes 6
Figure 1.2 Schematic overview on the complex pancreatic transcription factor interplay 9
Figure 5.1: Schematic display of substrate AAPF digestion by activated CTRL. 39
Figure 6.1: Schematic representation of CTRL mutations within the protein sequence. 49
Figure 6.2: Western blot of HEK293T supernatant (S) and lysate (L) after CTRL overexpression 50
Figure 6.3: CTRL activity in supernatant and cell lysate samples using AAPF 52
Figure 6.4: Schematic representation of the ER stress hypothesis 53
Figure 6.5: ER stress analysis results of CTRL wildtype and mutants 54
Figure 6.6: Summary of functional analyses of CTRL mutations 55
Figure 6.7: Schematic representation of the domain structures of NR5A2, RBPJL, and HNF1a 56
Figure 6.8: Western blot of wildtypes of NR5A2, HNF1A and RBPJL 58
Figure 6.9: Molecular cloning of *CPA2* promoter region into pGL4.22 59
Figure 6.10: Luciferase assay to establish suitable transcription factor-promoter combinations 60
Figure 6.11: Luciferase activity of promoters targeted by NR5A2 62
Figure 6.12: Luciferase activity of CPA1 promoter targeted by RBPJL 63
Figure 6.13: Luciferase activity of promoters targeted by HNF1a 64
Figure 6.14: Luciferase activity of promoters targeted by HNF1a with positive controls included 65
Figure 6.15: Target gene expression in Panc1 upon the overexpression of NR5A2 wt or mutant in complex 67
Figure 6.16: Target gene expression in HEK293 cells upon overexpression of NR5A2 69
Figure 6.17: Target gene expression in Panc1 upon overexpression of RBPJL wt or mutants with PTF1a 70
Figure 6.18: Nuclear protein binding of HEK293 cells upon overexpression of NR5A2 72
Figure 9.1: Original Western blot of HEK293T supernatant (S) and total lysate (L) after CTRL overexpression XIX
Figure 9.2: CTRL activity in supernatant and lysate from HEK293T cells using AAPY XX
Figure 9.3: CTRL activity in supernatant and lysate samples over time XXI
Figure 9.4: qPCR results of transcription factors target gene expression – Flag-tag vs. untagged XXII
Figure 9.5: Western blot analysis of all NR5A2 mutations XXIII

List of tables

Table 3.1: Chemicals 14
Table 3.2: Consumables 16
Table 3.3: Equipment 16
Table 3.4: Buffers for agarose gel electrophoresis 18
Table 3.5: Stock solutions for vector DNA extraction 18
Table 3.6: RIPA lysis buffer for total lysate preparation 18
Table 3.7: Laemmli Buffer for protein precipitation 18
Table 3.8: Buffers for SDS-PAGE and Western blot 19
Table 3.9: Buffer composition for activity assay 19
Table 3.10: Substrate buffers for luciferase assay 20
Table 3.11: Buffers for EMSA 20
Table 3.12: Cell lines used in this project with their specification and origin 21
Table 3.13: Control vectors for transient overexpression experiments 21
Table 3.14: Control vectors for reporter gene assays 22
Table 3.15: Cloned promoter regions in reporter gene vector constructs 22
Table 3.16: Enzymes used for Sequencing PCR and for molecular cloning 23
Table 3.17: Primary antibodies in Western blot 24
Table 3.18: Secondary antibodies for Western blot 24
Table 3.19: Oligonucleotides for EMSA 25
Table 3.20: Company-provided kits for ready-to-use 25
Table 5.1: Promoter regions for molecular cloning with potential transcription factor binding site 30
Table 5.2: compositions of growth media for culturing cell lines 34
Table 5.3: Number of cells seeded depending on cell line and seeding format. 34
Table 5.4: 10 % SDS-polyacrylamide gel preparation for SDS-Page 38
Table 5.5: Conditions for the activity assay for 2 h trypsin activation (in a 96-well plate) 40
Table 5.6: Conditions for the enzyme activation with different time points of trypsin activation 40
Table 5.7: 12 % acrylamide gel preparation for annealing oligonucleotides 43
Table 5.8: 5.3 % acrylamide gel preparation for EMSA 44
Table 6.1: Non-synonymous *AQP12A* variants in German CP patients and controls 46
Table 6.2: Non-synonymous *AQP12B* variants in German CP patients and controls 47
Table 6.3: Deleterious variant prediction in *CTRL* using REVEL 48
Table 6.4: Fisher's Exact test results on CTRL mutations 55
Table 6.5: Deleterious variant prediction in *NR5A2*, *RBPJL* and *HNF1a* using REVEL 57
Table 6.6: Promoter regions with potential TFBS successfully cloned into basic pGL4.22 59
Table 6.7: Promoter region with binding sites for HNF1a cloned into pGL4.22 vector 64
Table 6.8: Fisher's Exact test results on NR5A2 mutations 71
Table 6.9: Fisher's Exact test results on RBPJL mutations 71
Table 7.1: Summarized results on CTRL mutations 76
Table 9.1: Primers for *AQP12A* and *AQP12B* exon amplification and Sanger sequencing X
Table 9.2: Primers for mutation verification in *CTRL* with Sanger sequencing X

Table 9.3: Primers for mutation verification in *NR5A2* with Sanger sequencing XI
Table 9.4: Primers for mutation verification in *RBPJL* with Sanger sequencing XI
Table 9.5: Primers for mutation verification in *HNF1a* with Sanger sequencing XI
Table 9.6: Primers used for molecular cloning of target gene promoters XII
Table 9.7: Primers used for control sequencing and/or colony PCR of cloned vectors..... XII
Table 9.8: Primers used for molecular cloning of a triple promoter fragment XIII
Table 9.9: Primer for vector mutagenesis XIV
Table 9.10: Primers for XBP1 splicing PCR XIV
Table 9.11: TaqMan assay for ER stress analysis XV
Table 9.12: TaqMan assays for housekeepers XV
Table 9.13: TaqMan assays for transcription factors XV
Table 9.14: TaqMan assays for transcription factor target genes XV
Table 9.15: Synonymous and intronic *AQP12A* variants in German CP patients and controls XVII
Table 9.16: Synonymous and intronic *AQP12B* variants in German CP patients and controls XVIII

List of Abbreviations

Abbreviation	Name
6FAM	6-Carboxyfluorescein, fluorescent dye
A	Adenine
AAB	Activity Assay Buffer
AAPF	N-Suc-Ala-Ala-Pro-Phe-pNA
AAPY	N-Suc-Ala-Ala-Pro-Tyr-pNA
ACP	Alcoholic chronic pancreatitis
ACTB	β-actin
ALB	Albumin
AP-1	Activator protein 1
AP	Acute pancreatitis
APS	Ammonium persulfate
AQP12	Aquaporin-12
ATF4	Activating transcription factor 4
ATF6	Activating transcription factor 6
ATP	Adenosin-5'-triphosphate
BBQ	Black Berry® quencher
bHLH	Basic helix-loop-helix
BiP	Immunoglobulin-binding protein
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
$C_2H_3KO_2$	Potassium acetate
C3	Complement component 3
$CaCl_2$	Calcium chloride
CADD	Combined annotation dependent depletion
CALR	Calreticulin
CASR	Calcium sensing receptor
cDNA	Complementary DNA
CEL	Carboxyl ester lipase
CELA3A	Chymotrypsin like elastase 3A
CFTR	Cystic fibrosis transmembrane conductance regulator
ChIP	Chromatin immunoprecipitation
CLDN2	Claudin-2
CO_2	Carbon dioxide
CP	Chronic pancreatitis
CPA1	Carboxypeptidase A1
CPA2	Carboxypeptidase A2
C_t value	Cycle threshold value
CTRB	Chymotrypsin B
CTRC	Chymotrypsin C
CTRL	Chymotrypsin-like protein
CUZD1	CUB and zona pellucida like domains 1
Cy5	Cyanine 5
CYP11A1	Cytochrome P450 family 11 subfamily A member 1
CYP11B1	Cytochrome P450 family 11 subfamily B member 1
CYP17	Cytochrome P450 family 17 subfamily A member 1

Abbreviation	Name
ddH_2O	Double-distilled H2O
dH_2O	Distilled H2O
DMEM-hg	Dulbecco's Modified Eagle Medium - high glucose
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribose nucleoside triphosphates
DTT	1,4-dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF2a	eukaryotic initiation factor 2
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
FBS	Fetal bovine serum
FGB	β-fibrinogen
G	Guanine
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	Genomic DNA
gnomAD	Genome Aggregation Database
GP2	Glycoprotein 2
GSF	Gel filtration
GWAS	Genome wide association study
HEK293	Human embryonic kidney cell transformed by adenovirus type 5
HEK293T	HEK293T cell line with additional plasmid of the SV-40 large T-antigen
HNF1a	Hepatocyte nuclear factor 1 homeobox α
HNF1b	Hepatocyte nuclear factor 1 homeobox β
HPLC	High-performance liquid chromatography
IAPP	Islet amyloid polypeptide
ICP	Idiopathic chronic pancreatitis
IL18	Interleucin 18
IL32	Interleucin 32
INS	Insulin
IRE1α	Inositol requiring 1 α
K_2HPO_4	Dipotassium phosphate
kb	Kilobases
KCl	Potassium chloride
kDa	Kilo Dalton
KH_2PO_4	Potassium dihydrogenphosphate
LB-agar	Luria-Bertani agar
LB-medium	Luria-Bertani medium
M	Molar
mA	Mili ampere
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulfate hydrate
MODY	Maturity-onset diabetes of the young
mRNA	messenger RNA
MTTP	Microsomal triglyceride transfer protein
Na_2HPO_4	Disodium hydrogen phosphate

Abbreviation	Name
NaCl	Sodium chloride
NACP	Non-alcoholic chronic pancreatitis
NaN_3	Sodium azide
NaOH	Sodium hydroxide
NE	Nuclear extract
NR5A2	Nuclear receptor subfamily 5 group A member 2
OR	Odds ratio
PAGE	Polyacrylamide gel electrophoresis
Panc1	Human pancreatic ductal adenocarcinoma
PBS	Phosphate Buffered Saline Medium
PCR	Polymerase chain reaction
PDAC	Pancreatic ductal adenocarcinoma
PDX1	Pancreatic and duodenal homeobox 1
Pen/Strep	Penicillin/streptomycin
PERK	PKR-like ER kinase
PMSF	Phenylmethylsulfonyl fluoride
PNLIP	Pancreatic lipase
poly dIdC	Poly-deoxy-inosinic-deoxycytidylic acid sodium salt
PRSS1	Cationic trypsinogen
PRSS2	Anionic trypsinogen
PTF1	Pancreas transcription factor 1
PTF1a	Pancreas associated transcription factor 1a
qPCR	Semi-quantitative polymerase chain reaction
RBPJ	Recombination signal binding protein for immunoglobulin kappa J
RBPJL	RBPJ-like
REG1A	Regenerating family member 1 alpha
REVEL	Rare exome variant ensemble learner
RIPA	Radio-immunoprecipitation assay
RNA	Ribonucleic acid
rpm	Rounds per minute
RT	Room temperature
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
SOC Medium	Super optimal broth medium
SPINK1	Serine protease inhibitor Kazal type 1
T	Thymine
TAE	Tris-acetic acid-EDTA
TBE	Tris-boric acid-EDTA
TCA	Trichloroacetic acid
TE	Tris-EDTA
TEMED	Tetramethyl ethylenediamine
TFBS	Transcription factor binding site
TG	Thapsigargin
TM	Tunicamycin
Tris	Tris-hydroxymethyl -aminomethane buffer
TRPV6	Transient receptor potential cation channel subfamily V member 6
UPR	Unfolded protein response

Abbreviation	Name
v/v	Volume in volume
w/v	Weight in volume
WES	Whole exome sequencing
XBP1	X-box binding protein 1
β-ME	β-mercaptoethanol

1 Introduction

1.1 The pancreas and pancreatic organogenesis

The mature pancreas is a multi-functional gland consisting of an endocrine and exocrine compartment which both in their own ways regulate homeostasis in nutrient digestion and distribution. The endocrine pancreas harbors five cell types for hormone production which are organized in islets of Langerhans and embedded as clusters throughout the pancreas. Glucagon-producing α-cells and insulin-producing β-cells balance the blood glucose levels. δ-cells secrete somatostatin and mitigate the α- and β-cell function. Ghrelin-producing ε-cells and pancreatic polypeptide-producing cells are the smallest fraction. (Edlund 2002; Méndez-Giménez et al. 2018)

The exocrine compartment forms most of the pancreas with nearly 95 % of its mass (MacDonald et al. 2010). There are two major cell types. Acinar cells produce and secrete hydrolytic enzymes into the highly branched duct when stimulated by cholecystokinin. Ductal cells imbedded in the ductal epithelium secrete HCO_3^- and water upon secretin stimulation to modify the ionic composition of the fluid and the pancreatic juice is flushed into the duodenum (MacDonald et al. 2010; Edlund 2002). Per day, 2 liters are secreted into the duodenum for nutrient digestion (Lee et al. 2012). To prevent premature digestion within the pancreas, most digestive enzymes are synthesized and secreted as inactive precursors, so-called zymogens. Under normal conditions, the key player trypsinogen is catalyzed into its active form trypsin in the duodenum by membrane bound enteropeptidase and autoactivated by trypsin itself. An activation cascade of other digestive enzymes mediated by trypsin follows (Rinderknecht 1986).

The pancreas lays in the upper abdomen behind the stomach where its head is embedded into the duodenal loop. The prior separate dorsal and ventral structures fuse during organogenesis. The ventral structure forms the pancreas head while the body and tail develop from the dorsal structure. Islets of Langerhans differentiate early to produce hormones already in the embryonal state. The exocrine pancreas, however, develops later and becomes fully active only after birth. All pancreatic cell types originate from a shared population of multipotent progenitor cells within the early pancreatic epithelium (MacDonald et al. 2010). Studies of pancreatic gene expression have identified a variety of markers and transcription factors that play a central role in defining pancreatic cells throughout organogenesis (MacDonald et al. 2010).

1.2 Chronic pancreatitis: epidemiology of pancreatitis worldwide

Chronic pancreatitis (CP) is defined as a relapsing or continuous inflammatory disease of the pancreas that can result in a serious impairment of both the endocrine and exocrine function (Witt et al. 2007). Patients often suffer from severe abdominal pain, nausea, and/or vomiting which can severely impact their quality of life. With further pancreatic impairment, maldigestion and steatorrhea can occur as well as diabetes mellitus (Witt and Becker 2002). Early investigations of CP suggested a distinction between acute pancreatitis (AP) and CP to be two different diseases. However, further studies and epidemiological data showed a transition from recurrent AP to CP. Around 21 % of patients with a first episode of AP will have recurrent events and of those, roughly 36 % will develop CP (Sankaran et al. 2015). With longtime CP, the most severe complication is pancreatic ductal adenocarcinoma (PDAC) (Lowenfels et al. 1993; Yadav and Lowenfels 2013; Shelton and Whitcomb 2014). The worldwide CP prevalence ranges from 36.9 - 52.4 per 100,000 individuals due to regional differences in study methodology, diagnostic criteria, and local lifestyle risk factors (Yadav and Lowenfels 2013). The annual incidences are rather consistent with globally 8 - 12 new adult cases per 100,000 general population per year (Xiao et al. 2016). CP diagnosis would be most successful with an adequate surgical biopsy of an early-stage disease. As this is scarce, the diagnosis relies on relevant symptoms, imaging and assessment of the pancreatic function (Witt et al. 2007). To date, there is no gold standard for the classification. Two systems are the TIGAR-O and the M-ANNHEIM. The TIGAR-O system categorizes CP into risk factors: **T**oxic-metabolic, **I**diopathic, **G**enetic, **A**utoimmune, **R**ecurrent and severe AP-associated, or **O**bstructive factors (Etemad and Whitcomb 2001). The M-ANNHEIM system (**M**ultiple factors including **A**lcohol, **N**icotine, **N**utrition, **H**eredity, **E**fferent duct, **I**mmunology, **M**iscellaneous) additionally includes etiology, clinical stage, and severity of the disease (Schneider et al. 2007). The major etiological factor for CP is alcohol abuse, especially in Western countries. After a first event of alcohol-related AP, the risk for a progression to CP is 41 % (Yadav and Lowenfels 2013).

1.3 Genetics of chronic pancreatitis: known risk factors

Despite alcohol abuse, there are also non-alcohol related forms of CP. Especially in an early onset, genetic factors play a critical role (Witt et al. 2007). Already in 1952, Comfort and Steinberg hypothesized a genetic origin as they described repeated episodes of pancreatitis in a pedigree of six family members within three generations (Comfort and Steinberg 1952). With the age of onset between 5 to 23 years, they postulated an autosomal dominant inheritance pattern. Indeed, several genetic risk factors have since been identified, most of which encode pancreatic enzymes. A disturbed balance of the protease pathway seems crucial in the pathogenesis. Whitcomb *et al* gave the first proof of this hypothesis as they

identified a genetic defect in the cationic trypsinogen, *PRSS1.* The p.R112H mutation increases intra-pancreatic autoactivation of trypsinogen (Whitcomb et al. 1996). Ever since, several variants in *PRSS1* have been identified which promote either autoactivation or mitigate its proteolytic cascade. As trypsin harbored the confirmed central role in disease development, the model of a trypsin-dependent pathological pathway developed (Hegyi and Sahin-Tóth 2017). Under normal conditions, two major mechanisms protect pancreatic tissue from prematurely active trypsin. The potent serine protease inhibitor Kazal type 1 (SPINK1) inhibits the activation cascade by binding to prematurely activated trypsin within the pancreas and chymotrypsin C (CTRC) leads to trypsinogen degradation. It selectively cleaves it at the Leu^{81}-Glu^{82} peptide bond within the Ca^{2+} binding loop (Szmola and Sahin-Tóth 2007).

Thus, it was only a matter of time to investigate those protective enzymes as potential CP risk factors. In 2000, Witt *et al* identified *SPINK1* variants in a pediatric CP patient cohort. Especially p.N34S was found in 18 of 96 patients, 6 of whom were homozygous while only one in 227 controls was a heterozygous carrier (Witt et al. 2000). This strong association suggests a defective trypsin inhibition which leads to premature activation. Its actual pathological mechanism, however, remains unsolved until today as functional analyses show that SPINK1 p.N34S does not differ from wildtype. In 2007, Szmola and Sahin-Tóth identified the trypsinogen-degrading enzyme Y originally described by Rinderknecht in 1988 as chymotrypsin C. Its high specificity to degrade all three isoforms of human trypsinogen (Szmola and Sahin-Tóth 2007) arose the hypothesis that loss of function in CTRC increases pancreatitis risk. Soon after, Rosendahl *et al* sequenced 621 subjects with idiopathic (ICP) or hereditary chronic pancreatitis and 614 controls and found several *CTRC* variants. The two most frequent variants, p.R254W and p.K247_R254del, were significantly enriched in patients (3.3 %) compared to controls (0.7 %) (OR = 4.6, $P = 1.3x10^{-7}$) and severely diminish CTRC secretion and thus reduce trypsin inhibition (Rosendahl et al. 2008).

Contrary to disease-causing variants in *PRSS1*, studies on the anionic trypsinogen *PRSS2* never revealed any association with hereditary pancreatitis, despite a 90 % homologue primary structure. In 2016, Jancsó and Sahin-Tóth presented a tighter regulation of PRSS2 autoactivation as explanation to this phenomenon. An additional cleavage site at Leu^{148} and multiple other cleavage sites due to the lack of a conserved disulfide bond Cys^{139}-Cys^{206} increase sensitivity to CTRC-mediated degradation (Jancsó and Sahin-Tóth 2016). While single mutations in *PRSS1* lead to resistance against CTRC, no mutations in *PRSS2* can increase its stability enough to trigger intra-pancreatic activation. Further, a protective variant, p.G191R, leads to a rapid degradation as it introduces a new cleavage site for trypsin (Witt et al. 2006). A European Genome Wide Association study (GWAS) suggested another protective mutation with a 16.6 kb inversion in *CTRB1/CTRB2* (*rs8048956*) (Rosendahl et al. 2018). Chymotrypsin B2 (CTRB2) degrades PRSS2 to mitigate its activity more effectively

than CTRB1. The inversion shifts the expression levels toward a protective trypsinogen degradation and reduces the CP risk. The findings in an alcoholic CP (ACP) cohort were replicated in three non-alcoholic CP (NACP) cohorts of patients and controls (OR 1.62). However, this highly depends on the population genome as in China already 99.6 % carry the inverted allele without contributing to the disease risk (Tang et al. 2018).

Back to pathogenic mutations, research extended from the trypsin-dependent pathway to genetic alterations in other digestive enzymes. The hypothesis of a misfolding-dependent pathway was first described by Sahin-Tóth. He suggested that misfolded secretory enzymes accumulate intracellularly to trigger endoplasmic reticulum (ER) stress (Sahin-Tóth 2017). Mutations in *PRSS1* and carboxypeptidase A1 (*CPA1*) were the first studied examples with a strong CP association. In *in vitro* cell culture, rare *PRSS1* mutations show reduced secretion and intracellular retention (Kereszturi et al. 2009). Elevated markers such as the chaperone *BiP* mRNA and increased splicing of *XBP1* verified ER stress as consequence of PRSS1 p.R116C. Later, further mutations followed (p.C139S and p.L104P) (Balázs et al. 2016).

Especially in early onset CP (≤10 years of age), *CPA1* variants with functional defects are enriched with an OR = 84 (Witt et al. 2013). Loss-of-function variants were found in German and European cohorts as well as in India and Japan. The variant p.N256K showed the strongest CP association in German subjects (0.7 %; 7/944 cases vs 0/3,938 controls). *In vitro* studies in rat pancreatic AR42J cells evidenced an underlying misfolding mechanism which was later supported by *p.N256K*-knock-in mice as they showed spontaneous and progressive CP development (Hegyi and Sahin-Tóth 2019).

The carboxyl ester lipase (*CEL*) gene locus is another representative for the misfolding-pathway model. Fjeld *et al* showed a single nucleotide deletion in the last exon to be associated with maturity-onset diabetes of the young type 8 (MODY8) with a secondary pancreatic exocrine dysfunction (Fjeld et al. 2015). Moreover, a hybrid *CEL* allele (*CEL-HYB1*), which forms by a crossover with the neighboring pseudogene *CELP*, shows an association with ICP and ACP (Fjeld et al. 2015). Hybrid variants impair secretion which leads to intracellular accumulation and consequently to ER stress (Cassidy et al. 2020).

With CEL as the first lipase, the panel of potential CP risk genes extended from proteases and their inhibitors to pancreatic lipase (*PNLIP*). Lasher *et al* found an enrichment of *PNLIP* variants in two European patient cohorts (1.7 %; 18/1,061) compared to controls (0.6 %; 10/1,557). The variant p.F300L (*rs890551695*) was most abundant in German patients and absent in controls (Lasher et al. 2019). *PNLIP* variants were prone to proteolytic degradation by trypsin and chymotrypsin, however the pathogenic mechanism remains enigmatic.

Inconsistent results were generated for the chymotrypsin-like elastase 3B (*CELA3B*). The missense mutations p.R90C and p.R90L were reported to upregulate CELA3B expression for

uncontrolled proteolysis (Moore et al. 2019). However, these findings were not repeated when Tóth *et al.* investigated 550 German CP patients and a European replication cohort (243 patients) (Tóth et al. 2022). Instead, they detected a splice-site variant in intron 2, c.129+1G>A (1.64 % to 0.38 % in controls) predicted to diminish *CELA3B* expression. Nevertheless, the question remains of how reduced CELA3B expression contributes to pancreatitis (Tóth et al. 2022).

Apart from digestive enzymes and inhibitors, alterations in genes expressed other than in the acinar cells have long been studied. In 1998, two groups investigated the chloride/bicarbonate channel *CFTR* (cystic fibrosis transmembrane conductance regulator) and found several variants to be associated with CP, among p.F508del (OR 2.5) was the most frequent variant in patients followed by p.R117H (OR 4.0) (Sharer et al. 1998; Cohn et al. 1998). While both groups agreed on the genetic association, they argued for different biological consequences. Cohn *et al.* suggested that impaired dilution and alkalinization of the pancreatic juice might lead to pancreatic dysfunction (Cohn et al. 1998). On the contrary, Sharer *et al.* presented a reduced pH within the intra-acinar space and ductal lumen to directly cause acinar cell damage (Sharer et al. 1998). Later studies verified the association, however, with less distinctive effects and frequency (Rosendahl et al. 2013).

Another GWAS study identified several single nucleotide polymorphisms (SNPs) in the tight junction protein claudin 2 (*CLDN2-MORC4*) locus on the X-chromosome, though rather distinguished in ACP (Whitcomb et al. 2012). The study suggests a mislocalization of claudin-2 in pancreatic acinar cells as consequence of variants.

Ca^{2+} is known to stimulate premature trypsin activation (Krüger et al. 2000) and secretion. The transient receptor potential cation channel subfamily V member 6 gene (*TRPV6*) locus encodes for an Ca^{2+}-selective ion channel which contributes to the calcium homeostasis. Its dysregulation with elevated intracellular Ca^{2+} levels plays a critical role in the development of pancreatitis. Genetic analysis in Japanese and European subjects revealed enriched p.A210V, p.I223T, and p.D324N in the Japanese and p.L299Q and p.R483W in the European patient cohort. Ca^{2+} imaging assays in HEK293 cells then showed decreased activity in p.A210V and p.R483W, while p.L299Q and p.D324N resulted in a decreased voltage dependent Ca^{2+} flux (Masamune et al. 2020). Subsequent mice studies showed that *Trpv6*$^{mut/mut}$ mice develop severe CP upon caerulein administration due to increased levels of pancreatic enzymes, histological alterations, and fibrosis. The tissue specificity of *TRPV6* however seems to differ between human and mouse. Human RNA-seq data suggests an abundance rather in ductal cells (https://www.proteinatlas.org, (Uhlén et al. 2015) while murine immunoprecipitation studies found TVPV6 to be acinar (Wartenberg et al. 2021).

Felderbauer *et al* described the calcium-sensing receptor (*CASR*) as new potential risk gene. They found familial hypocalciuric hypercalcemia (FHH) patients with heterozygous *CASR* to develop CP if they additionally carried heterozygous *SPINK1* p.N34S (Felderbauer et al. 2003). This shows the complexity of CP once more. Since then, rare *CASR* mutations were described in patients with CP but with conflicting results (Ewers et al. 2021). As CASR plays a crucial role in calcium homeostasis and hypercalcemia is a metabolic risk factor for pancreatitis, the CP association of *CASR* remains a subject of further research.

As the first zymogen granule protein, CUB and zona pellucida like domains 1 (*CUZD1*) has recently been suggested as risk factor for NACP (Rygiel et al. 2022). Thirty non-synonymous variants were found in 1,163 European patients. Subsequent functional studies in Western blot then revealed >50 % reduced secretion in seven of these variants, however no significant enrichment could be predicted compared to the 2,018 European controls (Rygiel et al. 2022). Although this remains a subject to further studies, it introduces zymogen granule proteins as an entire new panel of potential risk factors for CP.

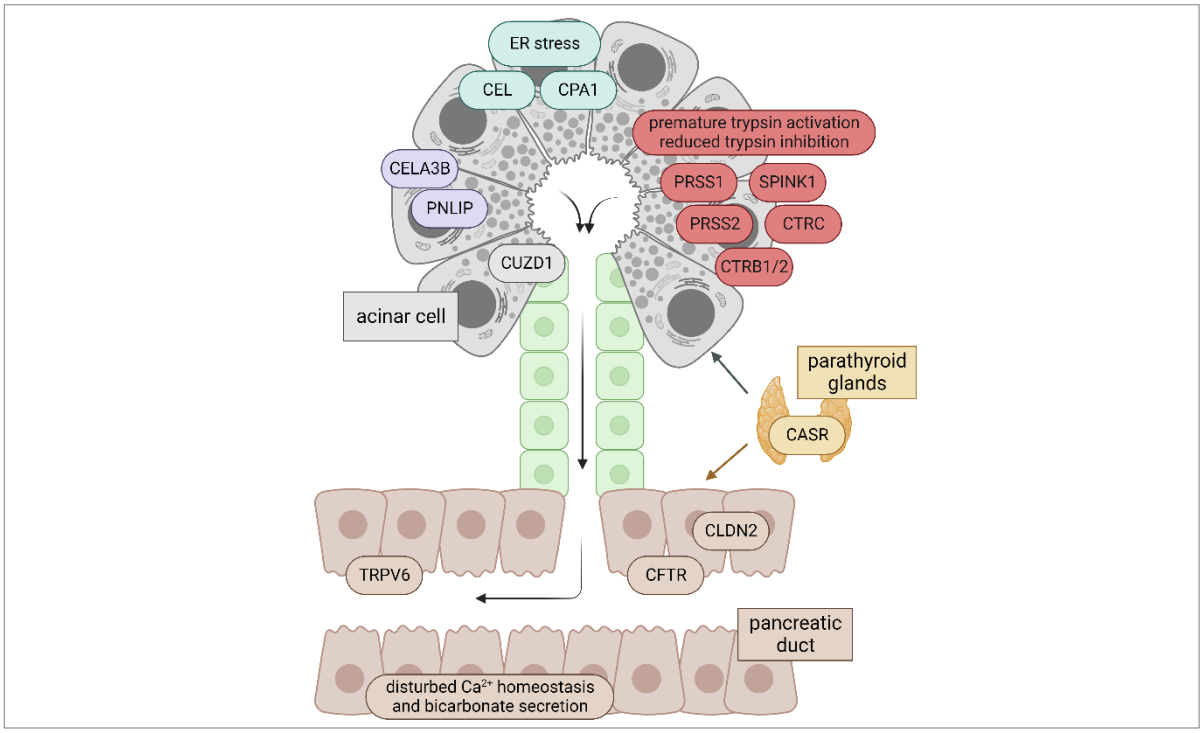


Figure 1.1: Pancreatitis-related genes
Schematic display of known pancreatitis related genes. Genes with the strongest disease association are located in the acinar cells. A disbalance in digestive enzymes and their inhibitors as well as ER stress are causal for disease development. A disturbed Ca^{2+} homeostasis and bicarbonate secretion promote pancreatitis development in the ductal cells. Figure is modified from Rosendahl and Witt (2021) and was created with BioRender.com.

In a considerable proportion of NACP patients the underlying cause is still unknown as they do not carry a known gene mutation. This suggests the existence of other to date undiscovered disease-causing genes. Based on an extended candidate approach as well as whole exome sequencing (WES) data, we identified a gene panel which are expressed in the

pancreas and regulate its development and homeostasis. To contribute to the clarification of pathogenic mechanisms in ICP, we picked five genes to investigate different potential pathological pathways.

1.4 Water channels in the exocrine pancreas

Based on an extended candidate approach, we came across the family of aquaporins (AQPs). These water channels effectuate the transport of water across a cell membrane. As the first water channel, aquaporin-1 was discovered in 1992 in red blood cells and renal proximal tubules when its transport activity was reconstituted in *Xenopus laevis* oocytes (Preston et al. 1992). Its structure consists of six transmembrane α-helices which are connected by five loops. Two of them contain an Asn-Pro-Ala (NPA) motif which is thought to play a key role in water selectivity and channel pore formation (Jung et al. 1994). Ever since, the unique protein fold of AQP1 - the so-called hourglass model - and the homology to its mRNA sequence enabled the discovery of several other water channels. Nowadays, thirteen different members of the aquaporin family are known (AQP0 - AQP12). With focus on the pancreas, up to 2 liters of fluid is secreted into the duodenum per day for nutrient digestion (Lee et al. 2012). Most of the fluids are produced by intercalated ducts cells. As aquaporins mediate water transport, they are also involved in the pancreatic fluid secretion. There are four aquaporins expressed in the exocrine pancreas: AQP1, AQP5, AQP8 and AQP12 (Burghardt et al. 2003; Itoh et al. 2005). On the apical membrane in the intercalated duct, AQP1 and AQP5 are abundantly expressed. Both show a high co-localization with CFTR, which plays an important role in secretin-evoked HCO^{3-} secretion (Burghardt et al. 2003). In contrast, AQP8 and AQP12 are predominantly expressed in acinar cells (Burghardt et al. 2003; Itoh et al. 2005). The expression of AQP12 in acinar cells was demonstrated by RT-PCR from micro-dissected pancreas and *in situ* hybridization. Itoh and colleagues suggested AQP12 in the zymogen granule membrane and its involvement in secretion of digestive enzymes via maturation and exocytosis of secretory granules (Itoh et al. 2005). Furthermore, pancreas-specific *Aqp12*$^{-/-}$ mice showed increased susceptibility to caerulein-induced AP, suggesting a role of AQP12 in the zymogen granule fusion (Ohta et al. 2009; Ishibashi et al. 2014). The two homologous genes, *AQP12A* and *AQP12B*, are located at chromosome 2q37.2 in the human genome. They share a homology of 95.5 % and are directed opposite to each other. Both are almost exclusively expressed in the pancreas (https://www.proteinatlas.org, (Uhlén et al. 2015).

1.5 CTRL as a new risk candidate in the panel of digestive enzymes

The pancreatic fluid contains different digestive enzymes, all of which contribute to nutrient digestion in different ways and affinities. The family of serine proteases consists of enzymes which hydrolyze specific peptide bonds using an activated serine residue in their substrate binding site (Reseland et al. 1997). Their members comprise of trypsins, chymotrypsins, and elastases, all of which are similar in size, structure, and function, however, with different peptide bond specificities. While trypsin as the most abundant serine protease makes up to 19 % in the pancreatic secretion, chymotrypsins still amounts to about 9 % in total pancreatic juice protein (Carrère et al. 1986; Whitcomb and Lowe 2007). They preferably cleave bonds of aromatic and aliphatic amino acids.

A new chymotrypsinogen on chromosome 16q22.1 was first found in 1993 as part of a tight cluster of five unrelated genes. Larsen and colleagues described it to be chymotrypsin-like as it is pancreas specific and shows a 55 % homology with the pro-protein sequence of CTRB. It has seven exons of which two code for active site residues characteristic for the serine protease family (VTAAHC and GDSGG) (Larsen et al. 1993). A few years later, CTRL was indeed suggested as a novel digestive enzyme by Reseland *et al*. Western blots showed increased CTRL levels in the duodenal juice upon stimulation of pancreatic secretion (Reseland et al. 1997). Nowadays, there are four known isoforms of chymotrypsin: CTRB1, CTRB2, CTRC and CTRL. They show similar activity by hydrolyzing polypeptides at certain amino acid residues, especially at tyrosine (Tyr) and phenylalanine (Phe) (Szabó and Sahin-Tóth 2012). However, due to different structures in the substrate binding pockets of these isoforms, substrate recognition differs. CTRC additionally cleaves after leucin (Leu) and methionine (Met) residues and CTRB1 prefers tryptophane (Trp).

As stated earlier, other members of the chymotrypsin family have already been associated with CP. Consequently, it is obvious to place CTRL in the center of attention. Only a few years ago, *Ctrl* knock-out mice were generated to investigate the role of CTRL in caerulein-induced pancreatitis (Mosztbacher et al. 2020). They identified CTRL as rather low-abundance isoform. In an AP model stimulation, these *Ctrl* knock-out mice showed no significant different protease activation to the control C57BL/6N mice. As *CTRL* has not been studied regarding recurrent or chronic pancreatitis, we now investigated *CTRL* variants found in patients on their secretory function and proteolytic capacity.

1.6 Transcription regulation in the pancreas: a complex interplay

As stated earlier, a variety of markers and transcription factors play a central role in defining pancreatic cells at different stages of organogenesis (MacDonald et al. 2010). While transcriptional regulation is quite complex, this holds true for the pancreas and requires a well-functioning and tightly regulated interplay. Figure 1.2 gives a schematic overview.

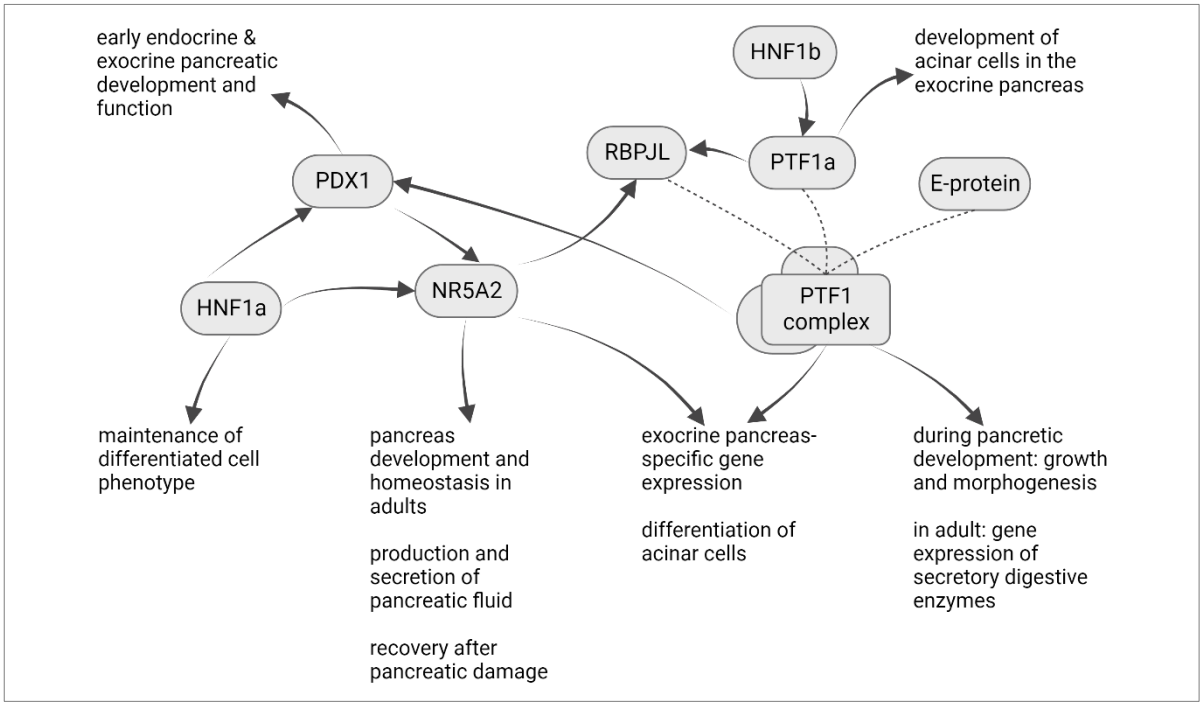


Figure 1.2 Schematic overview on the complex pancreatic transcription factor interplay
Combined literature results from (Holmstrom et al. 2011; Molero et al. 2012; Flandez et al. 2014; Naqvi et al. 2017). In early pancreatic development, the PTF1 complex partner is RBPJ instead of RBPJL. // HNF1a = hepatocyte nuclear factor 1 alpha; HNF1b = hepatocyte nuclear factor 1 beta; PDX1 = pancreatic and duodenal homeobox 1; NR5A2 = nuclear receptor subfamily 5 group A member 2; RBPJL = recombination signal binding protein for immunoglobulin kappa J region like; PTF1a = pancreas associated transcription factor 1a; PTF1 = pancreas transcription factor 1 complex; E-protein = basic helix-loop-helix transcription factor. Figure was created with BioRender.com.

The pancreas specific class B helix-loop-helix (bHLH) factor PTF1a (p48) is a key determinant of acinar cell identity and exocrine pancreas development. Ptf1a null mice die within hours after birth (Krapp et al. 1998). The loss of PTF1a drives acinar to ductal metaplasia (ADM) due to acinar cell reprogramming (MacDonald et al. 2010). PTF1a, together with a commonly expressed basic class B bHLH cofactor and RBPJ, forms the so called PTF1 complex (PTF1-J). Basic helix-loop-helix factors mediate homo- and heterodimerization and help the binding to E-boxes on the DNA with the amino acid sequence CANNTG (Murre et al. 1989; Chaudhary and Skinner 1999). One class of bHLHs are so called E-proteins which includes E47. While PTF1a interacts with the E-box on the DNA after dimerization with a bHLH protein, RBPJ binds to the TC box binding motif (TTTCCCACG) (Cockell et al. 1989; Roux et al. 1989; Beres et al. 2006). Rbpj is a widely expressed DNA-binding transcriptional mediator of the canonical Notch signaling pathway (Ehebauer et al. 2006). Notch signaling enables cell-cell

communication to promote or suppress cell proliferation, cell death, acquisition of specific cell fates, or activation of differentiation programs (Kopan and Ilagan 2009). The PTF1-J complex is responsible for morphogenesis and growth during development. As acinar cell differentiation occurs, Rbpj is gradually replaced by its paralogue Rbpjl as third subunit of the PTF1 complex (PTF1-L) (Masui et al. 2010). Other than Rbpj, Rbpjl is restricted to the pancreas and cerebral cortex and is not involved in Notch signaling (Beres et al. 2006). In the adult pancreas, the complex is essential to the maintenance of an acinar cell homeostasis and regulates the expression of secretory digestive enzymes (Molero et al. 2012). Further, Ptf1a regulates the expression of Pdx1 (pancreatic and duodenal homeobox 1), which is a key regulator of early endocrine and exocrine pancreatic development and function (Wiebe et al. 2007; Jin and Xiang 2019). It plays a role in the early pancreas development and is responsible for the β-cell function as well as insulin secretion (Naqvi et al. 2017). Furthermore, PDX1 induces the expression of the orphan nuclear receptor NR5A2 involved in pancreas organogenesis and homeostasis (Annicotte et al. 2003; Hale et al. 2014; Naqvi et al. 2017). In 2011, the group of Holmstrom demonstrated with luciferase experiments in HEK293T cells, that NR5A2 cooperates with the PTF1 complex - PTF1a and RBPJL - to have the strongest relative activity on target gene promoters *Rbpjl*, *Cel*, *Cpa2*, and *Ela1* (Holmstrom et al. 2011). They used ChIP-seq and RNA-seq analyses to reveal that NR5A2 directly induces the expression of digestive enzymes.

Another important transcription factor group are the hepatocyte nuclear factors HNF1a and HNF1b, which are expressed in both liver and pancreas. HNF1b regulates the expression of PTF1a (Naqvi et al. 2017). The hepatocyte nuclear factor 1a (Hnf1a) plays a role in the maintenance of differentiated cell phenotypes (Molero et al. 2012). In acinar cells, Hnf1a binds to and regulates the promoter of *Nr5a2* (Molero et al. 2012; Zhang et al. 2001). Molero *et al.* analyzed the dynamic regulation of pancreatic genes and the role of transcriptional regulators during the recovery from caerulein-induced AP. Upon caerulein treatment, *Hnf1a* knock-out mice had reduced levels of digestive enzymes, *Ptf1a* and *Nr5a2* (Molero et al. 2012). Later co-expression reporter gene assays in HEK293T cells evidenced cooperation between the PTF1-L complex and HNF1a on the *SPINK1* promoter (Boulling et al. 2017).

1.6.1 Transcription factors as potential risk genes for pancreatitis

In recent years, transcription factors have been investigated and shown to contribute to the development of pancreatic cancer (Petersen et al. 2010; Flandez et al. 2014; Naqvi et al. 2017). We hypothesize now that transcription factors involved in pancreas development and maintenance can predispose CP as well, however, data on an association with pancreatitis is rather scarce. Here, we investigated *NR5A2*, *RBPJL*, and *HNF1a* as representative players in different potential pathogenic pathways.

The orphan nuclear receptor subfamily 5 group A member 2 (NR5A2) was first identified in mouse as liver receptor homolog-1 (Lrh-1) (Tugwood et al. 1991). Nr5a2 is essential in the endoderm during embryonal development to later form liver, intestine, and pancreas (Rausa et al. 1999). In the liver and intestine, NR5A2 plays a role in the regulation of cholesterol, bile acid, and steroid hormone homeostasis (Lee et al. 2008). While there have been several suggestions on phospholipids as ligand (Krylova et al. 2005; Ortlund et al. 2005), the functional relevant ligand remains to be clarified. In the exocrine pancreas, NR5A2 contributes to acinar cell differentiation from pancreas progenitor cells and maintenance of acinar identity in the adult pancreas (Naqvi et al. 2017). Mice with a Nr5a2-deficient pancreas show partial loss of endocrine cells, a disrupted duct and lack >90 % acini (Hale et al. 2014). This leads to reduced digestive enzymes such as lipases and proteases and an impaired pancreatic fluid secretion. The same study strengthens the key role of Nr5a2 regarding the exocrine pancreas-specific transcriptional network, as it cooperates with the PTF1-L complex (Holmstrom et al. 2011). Holmstrom *et al.* generated *Nr5a2* knock-out mice and discovered a reduced concentration of lipases and proteases (*Cel, Cpa1, Ctrl, Ela1, Ela3, Pnlip, Klk1, Cp12, Rbpjl, Shp, Aqp, Sycn*) in the pancreatic fluid as well as impaired secretion. When comparing ChiP-seq and RNA-seq data with *Rbpjl* knock-out mice, they suggested a co-regulation of pancreatic genes by Nr5a2 and the Ptf1 complex. Luciferase assays showed highest promoter activation on *Rbpjl, Cel, Cpa2* and *Ela1* when PTF1a, RBPJL and NR5A2 were co-transfected (Holmstrom et al. 2011). In the past, several GWAS linked SNPs in the vicinity of *NR5A2* to an increased risk of developing pancreatic cancer (Petersen et al. 2010; Amundadottir 2016; Li et al. 2012). As this suggests an important role of NR5A2 in pancreatic homeostasis and disease, Flandez *et al.* hypothesized that genetic variants in *NR5A2* might result in a loss of function. They used heterozygous *Nr5a2*$^{+/-}$ mice to investigate the contribution of Nr5a2 in the recovery from pancreatitis-induced damage. Compared to wildtype mice, the heterozygous mice showed more severe AP and impaired recovery from damage upon caerulein implementation (Flandez et al. 2014).

Subsequent research aimed to identify the underlying mechanisms of impaired pancreas regeneration due to heterozygous *Nr5a2* loss. A common understanding was that transcriptional networks responsible for inflammatory responses differ from those involved in cell differentiation. Cobo *et al.* then introduced a new concept when RNA-seq showed a basal pre-inflammatory state in *Nr5a2*$^{+/-}$ mice even without caerulein-treatment. Of the 926 genes up-regulated in *Nr5a2*$^{+/-}$ mice compared to wildtype, 68 % belong to inflammatory pathways and code for chemokines (*i.e., Ccl5*, *Ccl7*), complement components (*i.e., C1qb*, *C3*) and metalloproteases. Subsequent RT-qPCR confirmed these results. Despite this, the adult pancreas in *Nr5a2*$^{+/-}$ mice did not differ histologically from the wildtype with equal expression levels of transcription factors and digestive enzymes. To confirm a subclinical pancreatitis-like

state in *Nr5a2*$^{+/-}$ mice, RNA-seq data was compared to wildtype mice with caerulein-induced AP. Indeed, while the basal transcriptomics of heterozygous and wildtype mice differ completely, dynamics changed upon caerulein treatment. Even one caerulein dose was sufficient to promote a transcriptional switch in wildtype mice similar to that of *Nr5a2*$^{+/-}$. ChIP-qPCR results show differential binding of Nr5a2 from the promotors of pancreatic to inflammatory genes. (Cobo et al. 2018) To assess similar transcriptomics in humans, Cobo *et al.* further support the hypothesis that NR5A2 levels are subject to cancer susceptibility as they showed an enrichment of tumors with reduced NR5A2 expression in PDAC patients with a history of CP (Cobo et al. 2018). Therefore, functional analyses of *NR5A2* mutations give us valuable insights into the cellular mechanisms that lead to the development of (subclinical) pancreatitis and even PDAC.

RBPJL as cofactor of the PTF1 complex replaces RBPJ during acinar differentiation. In the adult pancreas, the PTF1 complex with RBPJL plays a prominent role in the regulation of exocrine pancreas-specific gene expression and acinar cell identity. In 2010, Masui *et al.* showed in RNA-seq experiments not only an incomplete differentiation of acinar cells in *Rbpjl*$^{-/-}$ mice, but also diminished expression of Ptf1 target genes such as digestive enzymes and proteins which regulate exocytosis and mitochondrial metabolism (Masui et al. 2010). This underlines the important role of RBPJL for the PTF1 complex. As RBPJL is mainly responsible for a high transcriptional activation (Beres et al. 2006), genetic variants could disrupt its DNA binding affinity, the PTF1 complex formation and might impair or abolish transcriptional activity. Disturbed acinar cell homeostasis would be the consequence.

HNF1a exhibits an important role in the complex transcriptional interplay in the pancreas. Regarding the endocrine compartment of the pancreas, already in 1998, *Hnf1a* null mice have shown to be viable but develop early diabetes, fatty liver, and renal dysfunction (Pontoglio et al. 1998). Even prior, *HNF1a* was identified as maturity-onset diabetes of the young 3 (MODY3) risk gene (Yamagata et al. 1996). Since then, several genetic variants in *HNF1a* have been identified to predispose MODY3 (Chi et al. 2002; Barrio et al. 2002; Bjørkhaug et al. 2003; Najmi et al. 2017). Further, GWAS suggest *HNF1a* to be another risk locus that contribute to pancreatic cancer development (Petersen et al. 2010; Pierce and Ahsan 2011)]. As Hnf1a is known to cooperate with exocrine specific genes, the hypothesis arose of *HNF1a* variants to contribute to a pancreatitis phenotype. Experiments on tissue recovery after AP confirmed its crucial role in the maintenance of acinar cell homeostasis (Harries 2006; Molero et al. 2012). *Hnf1a* null mice showed increased acinar cell proliferation and lower levels of *Ptf1a*, *Rbpjl* and digestive enzyme mRNA (Molero et al. 2012). Functional analyses on *HNF1a* mutations found in CP patients sought to contribute to the understanding of CP underlying mechanisms.

2 Aim of the thesis

The aim of this thesis was to identify novel genetic risk factors associated with chronic pancreatitis (CP). As there are still many CP patients not clarified of their pathogenic mechanism, we identified a panel of candidate genes which are expressed in the pancreas.

One strategy was based on an extended candidate approach. For this, we did literature research combined with expression data on genes that are expressed in the pancreas and have not yet been investigated regarding chronic pancreatitis. One of those genes identified was *AQP12* as a member of the aquaporin family. Its exclusive expression in the zymogen granules of acinar cells prompted us to sequence 292 unrelated German CP patients and 143 German control subjects using Sanger Sequencing.

Another gene of interest based on the extended candidate approach is chymotrypsin-like protease (CTRL). As a chymotrypsin, it is suggested to have similar activity to CTRC. Therefore, we investigated genetic alterations of *CTRL* found in CP patients. For this, a pediatric patient cohort of 1,059 individuals was sequenced using whole exome sequencing (WES). The results could be analyzed regarding mutation frequencies compared to healthy controls and the public dataset from 56,885 non-finish European individuals in gnomAD v.2.1.1. This allows us to identify several missense as well as nonsense mutations in coding regions. After mutation verifications using Sanger sequencing, our next step was to functionally characterize the variants. Western blot was used for secretion analysis and activity assays were conducted to test the proteolytic capacity of CTRL. To investigate the possible effects of *CTRL* mutations on ER stress, we conducted Taqman assays on *BiP* mRNA expression using real-time polymerase chain reaction (qPCR) as well as *XBP1* splicing PCR.

A rather novel approach is the hypothesis of transcription factors involved in pancreas development and maintenance. We investigated *NR5A2*, *RBPJL*, and *HNF1a* as representative players in different potential pathogenic pathways. Especially *NR5A2* showed promising data from the bioinformatical analysis of our WES data. Functional analysis included reporter gene assays to investigate the transcriptional activity on target gene promoters. Furthermore, qPCR allowed us to determine mRNA expression of target genes depending on the wildtype or mutant transcription factor. Eventually, DNA binding capacity of NR5A2 was tested using EMSA. These functional analyses sought to contribute to the understanding of CP underlying mechanisms.

3 Material

3.1 Chemicals

Table 3.1: Chemicals

Chemical	Company
1,4-DTT	Sigma-Aldrich, Schnelldorf, Germany
2-Propanol, ROTISOLV® ≥99.9 %	Carl Roth, Karlsruhe, Germany
30 % acrylamide and bis-acrylamide solution, 37.5:1	Carl Roth, Karlsruhe, Germany
40 % acrylamide and bis-acrylamide solution, 37.5:1	Carl Roth, Karlsruhe, Germany
5X Passive Lysis Buffer	Promega, Walldorf, Germany
Acetic acid	Merck, Darmstadt, Germany
Ampicillin sodium salt ≥97 %	Carl Roth, Karlsruhe, Germany
APS	Bio-Rad, Feldkirchen, Germany
ATP	VWR, Darmstadt, Germany
BioRad Dye Protein Assay	Bio-Rad, Feldkirchen, Germany
Boric acid	Merck, Darmstadt, Germany
Bromphenol blue	Merck, Darmstadt, Germany
BSA, Fraction V, Blotting grade	AppliChem, Darmstadt, Germany
Calcium chloride ($CaCl_2$)	Carl Roth, Karlsruhe, Germany
Coelenterazine	Carl Roth, Karlsruhe, Germany
Coenzyme A	PJK Biotech, Kleinblittersdorf, Germany
Dipotassium phosphate (K_2HPO_4)	Sigma-Aldrich, Schnelldorf, Germany
Disodium phosphate (Na_2HPO_4)	Carl Roth, Karlsruhe, Germany
D-Luciferin	PJK Biotech, Kleinblittersdorf, Germany
DMEM-hg (cat# D6429)	Sigma-Aldrich, Schnelldorf, Germany
DMF	Sigma-Aldrich, Schnelldorf, Germany
DMSO (ROTIPURAN® ≥99.8 %, p.a.)	Carl Roth, Karlsruhe, Germany
dNTPs - Set1	Carl Roth, Karlsruhe, Germany
EDTA	Merck, Darmstadt, Germany
Ethanol, ultra-pure	Merck, Darmstadt, Germany
FastGene Agarose	Nippon Genetics Europe, Düren, Germany
FBS Superior (cat# S0615)	Sigma-Aldrich, Schnelldorf, Germany
Fugene-HD Transfection Reagent	Promega, Walldorf, Germany
GeneRuler 100 bp Plus DNA Ladder	Thermo Fisher Scientific, Darmstadt, Germany
GeneRuler™ DNA Ladder Mix	Thermo Fisher Scientific, Darmstadt, Germany
Glycerol	Carl Roth, Karlsruhe, Germany
Glycine	AppliChem, Darmstadt, Germany
H2O ROTISOLV® HPLC Gradient Grade	Carl Roth, Karlsruhe, Germany
HD Green™ DNA Stain	Intas Science Imaging Instruments, Göttingen, Germany
LB-Agar	Carl Roth, Karlsruhe, Germany
LB-Medium	Carl Roth, Karlsruhe, Germany
Lipofectamine™2000 Reagent	Invitrogen™ (Thermo Fisher Scientific), Darmstadt, Germany
Magnesium chloride ($MgCl_2$)	Carl Roth, Karlsruhe, Germany
Magnesium sulfate ($MgSO_4$)	Sigma-Aldrich, Schnelldorf, Germany
Methanol, ROTIPURAN® ≥99.9 %	Carl Roth, Karlsruhe, Germany
Monopotassium phosphate (KH_2PO_4)	Carl Roth, Karlsruhe, Germany

Chemical	Company
N-Suc-Ala-Ala-Pro-Phe-pNA	Sigma-Aldrich, Schnelldorf, Germany
N-Suc-Ala-Ala-Pro-Tyr-pNA	Chempep, Wellington, FL, USA
Opti-MEM® I (1X) (cat# 31985047)	Gibco™ (Thermo Fisher Scientific), Darmstadt, Germany
Orange G	Sigma-Aldrich, Schnelldorf, Germany
PageRuler™ Prestained Protein Ladder	Thermo Fisher Scientific, Darmstadt, Germany
PBS Solution (cat# D8537)	Sigma-Aldrich, Schnelldorf, Germany
Penicillin/Streptomycin (cat# P4333)	Sigma-Aldrich, Schnelldorf, Germany
PMSF	Carl Roth, Karlsruhe, Germany
Poly dIdC	Sigma-Aldrich, Schnelldorf, Germany
Porcine trypsin	Sigma-Aldrich, Schnelldorf, Germany
Potassium acetate ($C_2H_3KO_2$)	Sigma-Aldrich, Schnelldorf, Germany
Potassium chloride (KCl)	Carl Roth, Karlsruhe, Germany
SDS	Carl Roth, Karlsruhe, Germany
SOC medium	Takara, Saint-Germain-en-Laye, France
Sodium azide (NaN_3)	Merck, Darmstadt, Germany
Sodium chloride (NaCl)	Carl Roth, Karlsruhe, Germany
Sodium hydroxide (NaOH)	Carl Roth, Karlsruhe, Germany
Sucrose	Sigma-Aldrich, Schnelldorf, Germany
TCA, ≥99.0 %	Sigma-Aldrich, Schnelldorf, Germany
TEMED (≥99 %, p.a., for electrophoresis)	Carl Roth, Karlsruhe, Germany
Thapsigargin (cat# AG-CN2-0003-M001)	BIOMOL, Hamburg, Germany
Tricine	Sigma-Aldrich, Schnelldorf, Germany
Tris-hydroxymethyl-aminomethane buffer	Merck, Darmstadt, Germany
Triton X-100	Sigma-Aldrich, Schnelldorf, Germany
Trypane Blue Solution, 0.4 %	Sigma-Aldrich, Schnelldorf, Germany
Trypsin-EDTA-Solution (cat# T3924)	Sigma-Aldrich, Schnelldorf, Germany
Tunicamycin (cat# LKT-T8153.1)	BIOMOL, Hamburg, Germany
Tween® 20	VWR, Darmstadt, Germany
Xylene cyanol	Sigma-Aldrich, Schnelldorf, Germany
β-ME	Thermo Fisher Scientific, Darmstadt, Germany

3.2 Equipment and Consumables

Table 3.2: Consumables

Consumable	Catalog No.	Company
48-well plates; cell culture	734-2326	VWR, Darmstadt, Germany
6-well plates; cell culture	734-2323	VWR, Darmstadt, Germany
96-well microplates	391-3615	VWR, Darmstadt, Germany
Amersham™ Protran® Western blotting nitrocellulose membrane	GE10600003	Sigma-Aldrich, Schnelldorf, Germany
Cell scraper, size M	83.3951	Sarstedt, Nürnbrecht, Germany
FrameStar® 384 Well Skirted qPCR plate	4ti-0382	4titude
Luciferase assay PS tubes	115101	Greiner Bio-One, Frickenhausen, Germany
PCR cover foil	4ti-0500	4titude, Surrey, UK
PCR Multiply 4er stripes	72.990.002	Sarstedt, Nürnbrecht, Germany
PCR Multiply 8er stripes	72.985.002	Sarstedt, Nürnbrecht, Germany
PCR plate	4ti-0750	4titude, Surrey, UK
Petri dishes for agar plates	82.1473	Sarstedt, Nürnbrecht, Germany
Petri dishes, (ØxH): 100 x 20 mm; cell culture	83.3902	Sarstedt, Nürnbrecht, Germany
qPCR cover foil	4ti-0560	4titude, Surrey, UK
T175 cell culture flasks	83.3912.002	Sarstedt, Nürnbrecht, Germany
T75 cell culture flasks	83.3911.002	Sarstedt, Nürnbrecht, Germany
Whatmann blotting paper	GB 46	A. Hartenstein GmbH, Würzburg, Germany

Table 3.3: Equipment

Equipment	Usage	Company
5810	Centrifuge	Eppendorf, Hamburg, Germany
5415D	Centrifuge	Eppendorf, Hamburg, Germany
5417R	Centrifuge	Eppendorf, Hamburg, Germany
ARE Heating Magnetic Stirrer	Heating plate	VELP, Usmate Velate, Italy
Axiovert 40C	Microscope	Carl Zeiss, Oberkochen, Germany
Azure Sapphire™ Biomolecular Imager	WB scanner	Biozym, Hessisch Oldendorf, Germany
B28	Incubator	Binder, Tuttlingen, Germany
Chromas 2.6.6	Software	Technelysium, South Brisbane, Australia
FastGene FAS V Gel Imaging System	Gel visualization	NIPPON Genetics, Düren, Germany
Heracell 150	Incubator	Thermo Fisher Scientific, Darmstadt, Germany
Heraeus Multifuge X3R	Centrifuge	Thermo Fisher Scientific, Darmstadt, Germany
IKA® HS501 digital	Vertical shaker	IKA, Staufen im Breisgau, Germany
Intelli-Mixer RM-2S	Shaker	LTF Labortechnik, Wasserburg am Bodensee, Germany
KM-2	Vertical shaker	Edmund Bühler, Bodelshausen, Germany
LightCycler® 480 II	qPCR	Roche, Basel, Switzerland
Maxigel system	EMSA chamber	Biometra, Göttingen, Germany
Mini-PROTEAN Tetra Cell	SDS-PAGE chamber	Bio-Rad, Feldkirchen, Germany
NanoPhotometer® N60	Nanometer	Implen, Munich, Germany

Equipment	Usage	Company
Neubauer improved cell-counting chamber (0.0025 mm², depth 0.100 mm)	Cell counting	Paul Marienfeld, Lauda-Königshofen, Germany
Owl™ Gel Electrophoresis System	Gelelectrophoresis	Thermo Fisher Scientific, Darmstadt, Germany
P25 Standard Power Pack	EMSA power supply	Biometra, Göttingen, Germany
PowerPac™ 200 power supply	Gelelectrophoresis	Bio-Rad, Feldkirchen, Germany
PowerPac™ Basic	SDS-PAGE power supply	Bio-Rad, Feldkirchen, Germany
Professional Thermocycler	PCR	Biometra, Göttingen, Germany
QBD2	Heat shock	Grant Instruments Europe B.V., Amsterdam, Netherlands
SIRIUS Luminometer	Luciferase assay	Berthold Detection Systems, Pforzheim, Germany
SnapGene Viewer	Software	Dotmatics, Boston, MA, USA
SVL-6.LC, Serial No. 04 14423	UV light	Vilber, Marne-la-Vallée, France
SW22	Water bath	JULABO GmbH, Seelbach, Germany
Thermo Haake® DC10-K20	EMSA cooling	Thermo Fisher Scientific, Darmstadt, Germany
Thermomixer comfort	Shaker/incubator	Eppendorf, Hamburg, Germany
Thermomixer compact	Shaker/incubator	Eppendorf, Hamburg, Germany
Typhoon TRIO+® Variable Mode Imager	EMSA scanner	Amersham Bioscience Europe, Freiburg, Germany
UP200S	Sonification	Hielscher Ultrasonics, Teltow, Germany
VarioScan™ Flash Multimode Reader	Photometer	Thermo Fisher Scientific, Darmstadt, Germany
ZWY-100H	Bacterial shaker	Labwit Scientific, Ashwood, Australia

3.3 Buffers and solutions

All buffers and solutions are dissolved and prepared in double-distilled H_2O unless stated otherwise. Chemicals used for the preparation of buffers are listed in Table 3.1.

Table 3.4: Buffers for agarose gel electrophoresis

50X TAE buffer (autoclaved)	
2 M	Tris
64 mM	EDTA
5.7 %	Acetic acid (v/v)
6X Loading dye	
50 mg	Bromphenol blue
50 mg	Xylene cyanol
8 g	Sucrose
ad to 20 ml with dH_2O	

Table 3.5: Stock solutions for vector DNA extraction

FlexiPrep solution 1	
100 mM	Tris (pH 7.5)
10 mM	EDTA (pH 8.0)
0.1 %	RNase (v/v) – freshly added
FlexiPrep solution 2	
200 mM	NaOH
1 %	SDS (v/v)
FlexiPrep solution 3	
3 M	Potassium acetate
11.5 %	Acetic acid (v/v)

Table 3.6: RIPA lysis buffer for total lysate preparation

RIPA buffer stock (pH 7.4)	
10 mM	Tris (pH 7.4)
150 mM	NaCl
5 mM	EDTA
1 %	Triton X-100 (v/v)
10 mM	PMSF

Table 3.7: Laemmli Buffer for protein precipitation

1X Laemmli-DTT buffer (pH 6.8)	
50 mM	Tris (pH 6.8)
10 %	Glycerol (v/v)
2 %	SDS (v/v)
12.5 mM	EDTA
0.02 %	Bromphenol blue (w/v)
1 %	β-ME (v/v)
100 mM	DTT

Table 3.8: Buffers for SDS-PAGE and Western blot

4X Laemmli Buffer (pH 6.8)	
250 mM	Tris (pH 6.8)
40 %	Glycerol (v/v)
8 %	SDS (v/v)
0.02 %	Bromphenol blue (w/v)
10 %	β-ME (v/v)
SDS stacking gel buffer	
139 mM	Tris (pH 6.8)
0.11 %	SDS (v/v)
SDS separation gel buffer	
1.12 mM	Tris (pH 8.8)
0.3 %	SDS (v/v)
5X SDS running buffer	
125 mM	Tris
960 mM	Glycine
17.3 mM	SDS
2X transfer buffer	
40 mM	Tris
300 mM	Glycine
40 %	Methanol (v/v)
0.0004 %	SDS (v/v)
10X PBS	
1.37 M	NaCl
27 mM	KCl
100 mM	Na_2HPO_4
18 mM	KH_2PO_4

Table 3.9: Buffer composition for activity assay

Activity assay buffer	
0.1 M	Tris (pH 8.0)
10 mM	$CaCl_2$

Table 3.10: Substrate buffers for luciferase assay

Firefly luciferase buffer (pH 7.8)	
33.3 mM	DTT
20 mM	Tricine
2.67 mM	$MgSO_4$
0.1 mM	EDTA
530 µM	ATP
470 µM	D-Luciferin
270 µM	Coenzyme A
Renilla luciferase buffer (pH 5.0)	
1.1 M	NaCl
220 mM	K_2HPO_4 (pH 5.1)
2.2 mM	EDTA
1.3 mM	NaN_3
1.43 µM	Coelenterazine
0.44 mg/ml	BSA

Table 3.11: Buffers for EMSA

TE buffer (pH 7.6)	
10 mM	Tris
1 mM	EDTA
5X TBE buffer (pH 8.3)	
445 mM	Tris
445 mM	Boric acid
10 mM	EDTA
5X binding buffer (pH 7.5)	
50 mM	Tris
20 %	Glycerol (v/v)
5 mM	$MgCl_2$
2.5 mM	EDTA
2.5 mM	DTT
10X loading buffer (Orange G; pH 7.5)	
250 mM	Tris
40 %	Glycerol (v/v)
0.2 %	Orange G (v/v)

3.4 Cell lines

Table 3.12: Cell lines used in this project with their specification and origin

Cell line	Specification	Origin
Stellar™ chemo competent cells	*E. coli* HST08 strain	Takara (Cat. #636763), Saint-Germain-en-Laye, France
HEK293	Human embryonic kidney cell transformed by adenovirus type 5	DSMZ (ACC05), Braunschweig, Germany
HEK293T	HEK293T cell line with additional plasmid of the SV-40 large T-antigen (tsA1609)	DSMZ (ACC-635), Braunschweig, Germany
Panc1	Human Pancreatic Ductal Adenocarcinoma	Kind gift of Prof. Dr. Jonas Rosendahl, Halle (Saale)

3.5 Vector plasmids

3.5.1 Human expression vectors

We ordered human expression vectors from GenScript Biotech B.V. (Leiden, Netherlands). All constructs were ordered in the pcDNA3.1+/C-(K)-DYK backbone vector (Table 3.13).

Vectors for the wildtype CTRL (OHu00893D; NM_001907.2) and 17 mutations (p.L2S, p.S15Y, p.G20S, p.G37E, p.G56S, p.G61S, p.R106W, p.S112N, p.T150I, p.H173R, p.V177L, p.C201Y, p.S208F, p.G215R, p.L218Rfs*33, p.C220G, and p.G230S) were used for experiments as ordered with DYK-tag (Flag-tag).

Ordered vector constructs for the wildtype NR5A2 (OHu20723D; NM_205860.2) and a total of 15 mutants (p.L10I, p.G24R, p.R28Q, p.R54Q, p.P63S, p.R142H, p.G150R, p.A195T, p.P224H, p.D236H, p.T248R, p.R267Q, p.K321E, p.G343E, and p.D412E) were re-cloned using polymerase chain reaction (PCR) mutagenesis to remove the Flag-tag. The same was done for the wildtype RBPJL (OHu01387; NM_001281449.1) and wildtype HNF1a (OHu25248; NM_000545.6). All primers used for PCR mutagenesis are listed in Table 9.9. For RBPJL, we additionally ordered ten mutants (p.E32K, p.E84K, p.Q200X, p.V215A, p.R222P, p.R265X, p.V459L, p.P476A, p.F505L, and p.N509H) and five mutants for HNF1a (p.A161T, p.R168C, p.G183R, p.R287Q, and p.P580L) without Flag-tag.

Table 3.13: Control vectors for transient overexpression experiments

Name	Vector properties
pcDNA3.1+/C-(K)-DYK	DYK-tag; driven by a CMV promoter
pcDNA3.1_Flag_*PacI*	pcDNA3.1+/C-(K)-DYK with additional *PacI* restriction enzyme cutting site at the 3'-end of DYK-tag

DYK tag amino acid sequence: DYKDDDDK

3.5.2 Reporter gene vectors (luciferase vectors)

Unless stated otherwise, the laboratory of Prof. Heiko Witt provided the control reporter gene vectors. All constructs with selected promoter regions listed in Table 3.15 were cloned from genomic DNA of healthy donors into the basic reporter gene vector pGL4.22.

Table 3.14: Control vectors for reporter gene assays

Name	Reporter gene	Coding sequence
pRL-Ubi	Renilla luciferase	luciferase from the anthozoan coelenterate Renilla reniformis (*sea pansy*)
pGL4.22	Firefly luciferase	luciferase reporter gene *luc2* (*Photinus pyralis*); designed for high expression and reduced anomalous transcription
pGL3-basic (*)	Firefly luciferase	modified coding region for firefly luciferase *luc*+ (*Photinus pyralis*)

(*) kind gift of Prof. Anders Molven, University of Bergen

Table 3.15: Cloned promoter regions in reporter gene vector constructs

Name	Vector	Promoter sequence
pGL4.22_CPA2prom	pGL4.22	human *CPA2* promoter region -501/+20bp (chr7:130,266,301-130,266,877, hg38)
pGL4.22_CELprom	pGL4.22	human *CEL* promoter region -551/+22bp (chr9:133,061,428-133,061,994, hg38)
pGL4.22_CYP11B1prom	pGL4.22	human *CYP11B1* promoter region -1061/+29bp (chr8:142,879,824-142,880,901, hg38)
pGL4.22_CYP17prom	pGL4.22	human *CYP17* promoter region -2856/+31bp (chr10:102,837,389-102,840,265, hg38)
pGL4.22_CALRprom	pGL4.22	human *CALR* promoter region -1802/+93bp (chr19:12,936,780-12,938,664, hg38)
pGL4.22_SPINK1prom	pGL4.22	human *SPINK1* promoter region -1154/+215bp (chr5:147,831,578-147,832,935, hg38)
pGL4.22_CELA3Aprom	pGL4.22	human *CELA3A* promoter region -1965/+23bp (chr1:21,999,704-22,001,673, hg38)
pGL4.22_REG1Aprom	pGL4.22	human *REG1A* promoter region -1136/+384bp (chr2:79,119,355-79,120,866, hg38)
pGL4.22_GP2prom	pGL4.22	human *GP2* promoter region -1402/+1084bp (chr16:20,326,435-20,328,909, hg38)
pGL4.22_PRSS2prom	pGL4.22	human *PRSS2* promoter region -1575/+18bp (chr7:142,769,398-142,770,979, hg38)
pGL4.22_CPA1prom	pGL4.22	human *CPA1* promoter region -2051/+27bp (chr7:130,378,443-130,380,520, hg38)
pGL4.22_INSprom	pGL4.22	human *INS* promoter region -2314/+366bp (ch11:2,160,976-2,163,655, hg38)
pGL4.22_IAPPprom	pGL4.22	human *IAPP* promoter region -2058/+494bp (chr12: 21,370,810-21,373,361, hg38)
pGL4.22_MTTPprom	pGL4.22	human *MTTP* promoter region -1613/+201bp (chr4:99,573,094-99,574,907, hg38)
pGL3-RA (*)	pGL3-basic	firefly luciferase gene under the control of a rat albumin promoter (-170/+5bp) fragment
pGL3-HNF4AP2 (*)	pGL3-basic	firefly luciferase gene under the control of a human *HNF4aP2* promoter (-418/+13bp) (MIM: 600281) fragment
pGL4.22_2xFGBprom	pGL4.22	two copies of human *FGB* promoter region -120/-29bp (NM_005141.5)
pGL4.22_2xALBprom	pGL4.22	two copies of human *ALB* promoter region -129/+42bp (NM_000477.7)

(*) kind gift of Prof. Anders Molven, University of Bergen; exemplified explanation on promoter sequence -501/+20 bp = 501 base pairs in 5'-direction of 5'-UTR and 20 base pairs into 5'-UTR.

3.6 Primers

We designed primers manually and verified their properties and binding capacity using the SnapGene Viewer Software. All primers were then ordered from TIB MOLBIOL (Berlin, Germany). Sequences can be found in the supplementary data section 9.1.

3.7 TaqMan assays for real-time polymerase chain reaction (qPCR)

All primers and probes for TaqMan assays were designed and obtained from TIB MOLBIOL. For each assay, we tested two forward and two reverse primers for their optimal combination. Based on the cycle threshold (C_t) value and the slope of the results, we selected the best combination. The hydrolysis probe was labeled with a donor (6FAM-) and acceptor (—BBQ). All primer and probe sequences are listed in the supplementary data section 9.2.

3.8 Enzymes

Enzymes were used for PCR, for restriction digestion and for ligation.

Table 3.16: Enzymes used for Sequencing PCR and for molecular cloning

Name	Catalog No.	Company
AmpliTaq Gold™ DNA Polymerase (5 U/µl)	4311818	Applied Biosystems (Thermo Fisher Scientific), Darmstadt, Germany
Antarctic Phosphatase	M0289L	New England Biolabs, Frankfurt am Main, Germany
Exonuclease I	M0293L	New England Biolabs, Frankfurt am Main, Germany
FastDigest *ApaI*	FD1414	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *BglII*	FD0083	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *BpiI* (IIs class)	FD1014	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *EcoRI*	FD0274	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *HindIII*	FD0504	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *MunI*	FD0753	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *NheI*	FD0973	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *PacI*	FD2204	Thermo Fisher Scientific, Darmstadt, Germany
FastDigest *XhoI*	FD0694	Thermo Fisher Scientific, Darmstadt, Germany
MyTaq™ DNA Polymerase (5 U/µl)	BIO-21107-BL	BioCat, Heidelberg, Germany
OneTaq® Hot Start DNA Polymerase (5 U/µl)	M0481X	New England Biolabs, Frankfurt am Main, Germany
PCRBIO HiFi Polymerase (2 U/µl)	PB10.41-02	PCR Biosystems, London, UK
Phusion™ High-Fidelity DNA Polymerase (2 U/µl)	F530L / F534L	Thermo Fisher Scientific, Darmstadt, Germany
Rnase A (17,500 U)	19101	Qiagen, Hilden, Germany
Platinum™ SuperFi II DNA Polymerase	12361050	
T4 DNA Ligase (5 U/µl)	EL0011	Thermo Fisher Scientific, Darmstadt, Germany

3.9 Antibodies

Primary antibodies are directed against the Flag-tag of the protein of interest or specifically against the protein itself. Anti β-actin was used for monitoring an equal loading of samples. Secondary antibodies were ordered from LI-COR Bioscience (Bad Homburg, Germany) labelled with a fluorophore, which is excited by a near-infrared radiation for detection.

Table 3.17: Primary antibodies in Western blot

Name	Species	Dilution	Catalog No.	Company
Anti β-actin antibody (polyclonal)	Goat	1:2,500	ab8229	abcam, Berlin, Germany
Monoclonal ANTI-FLAG® M2 antibody	Mouse	1:1,000	F1804-200UG	Sigma-Aldrich, Schnelldorf, Germany
Anti-NR5A2 antibody (polyclonal)	Rabbit	1:1,000	ABE2867	Sigma-Aldrich, Schnelldorf, Germany
Anti-NR5A2 antibody (polyclonal)	Rabbit	1:800	HPA005455	Sigma-Aldrich, Schnelldorf, Germany
Anti-NR5A2 antibody (polyclonal)	Rabbit	1:1,000	PA5-28347	Invitrogen™ (Thermo Fisher Scientific), Darmstadt, Germany

Table 3.18: Secondary antibodies for Western blot

Target	Species	Catalog No.	Company
Goat	IRDye® 800CW donkey anti-goat IgG	926-32214	LI-COR Biosciences, Bad Homburg, Germany
Mouse	IRDye® 680RD donkey anti-mouse IgG	926-68072	LI-COR Biosciences, Bad Homburg, Germany
Rabbit	IRDye® 680RD donkey anti-rabbit IgG	926-68023	LI-COR Biosciences, Bad Homburg, Germany

3.10 Oligonucleotides for EMSA

We ordered a 45 bp promoter region containing the corresponding transcription factor binding motif as forward and reverse strand from TIB MOLBIOL. All forward strands are labelled at the 5'-end with a Cy5-fluorophor.

Table 3.19: Oligonucleotides for EMSA

Name	Sequence 5'→3'
NR5A2_CPA1-fwd_Cy5	Cy5-CTATGGCTCTCACCAGTGCCCAAGGTCGGGGTCCTGACACAGAGC
NR5A2_CPA1-rev	GCTCTGTGTCAGGACCCCGACCTTGGGCACTGGTGAGAGCCATAG
NR5A2_CYP17-fwd_Cy5	Cy5-GGCAAGAGATAACACAAAGTCAAGGTGAAGATCAGGGTAGCCCTT
NR5A2_CYP17-rev	AAGGGCTACCCTGATCTTCACCTTGACTTTGTGTTATCTCTTGCC
NR5A2_CPA2-fwd-Cy5	Cy5-CAATCATAACTATATTGCCCTTGGTATGCTCTAGCAATT
NR5A2_CPA2-rev	AATTGCTAGAGCATACCAAGGGCAATATAGTTATGATTG
NR5A2_CYP11A1-fwd_Cy5	Cy5-CTTCTGGTATGGCCTTGAGCTGGTAGTT
NR5A2_CYP11A1-rev	AACTACCAGCTCAAGGCCATACCAGAAG

3.11 Company-provided kits

Table 3.20: Company-provided kits for ready-to-use

Kit	Catalog No.	Company
Mix2Seq Kits	-	Eurofins Genomics, Ebersberg, Germany
Nuclear Extract Kit	40010	Active Motif Europe, Waterloo, Belgium
QIAamp DNA Mini Blood Kit	51104	Qiagen, Hilden, Germany
QIAshredder	79656	Qiagen, Hilden, Germany
QuantiTect® Reverse Transcription Kit	205313	Qiagen, Hilden, Germany
RNeasy® Mini Kit	74106	Qiagen, Hilden, Germany
SensiFast™ Probe No-ROX Kit	BIO-86020	BioCat, Heidelberg, Germany
Wizard® SV Gel and PCR Clean-Up System	A9282	Promega, Walldorf, Germany

4 Study cohort

4.1 Patient cohort

We enrolled 1,059 CP non-related European CP patients for exome sequencing. Genomic DNA from peripheral blood leukocytes was extracted using the QIAamp DNA Mini Blood Kit (Table 3.20). All data was kindly provided by Prof. Heiko Witt as well as other research groups from Berlin (Germany), Halle (Germany), Brest (France), UK and Sweden.

4.2 Whole Exome Sequencing

Whole exome sequencing (WES) was performed by BGI Genomics in Hongkong. In the Genome Center in Tübingen, Germany, Dr. Marc Sturm performed the alignment based on the hg19 reference genome using the megSAP pipeline. Our lab eventually obtained VCF files for each sample. For the prediction of deleterious variants, we annotated CADD 1.6 (Combined Annotation Dependent Depletion; (Rentzsch et al. 2021) and REVEL (Rare Exome Variant Ensemble Learner; (Ioannidis et al. 2016) predictions. Burden testing of rare variants (Guo et al. 2018) in candidate causal genes against the gnomAD database was performed by Andreas Schmidt.

4.3 Control cohorts

For *AQP12* mutation screening, we sequenced genomic DNA from 143 German control subjects (72 females; mean age ± SD: 40.7 ± 9.0 years) using direct sequencing. These healthy blood donors were medical students or staff.

For the bioinformatical analysis of WES data, we used a public dataset from 56,885 non-Finish European individuals in gnomAD v.2.1.1 (https://gnomad.broadinstitute.org) as controls.

Provided by the Genome Center of Tübingen, 2,099 patients with rather neurological disorders were included in the NGS sequencing as an additional control cohort.

5 Experimental procedures

5.1 Genetic verification of mutations

5.1.1 Sanger sequencing

We verified all mutations found in whole exome sequencing (WES) by Sanger sequencing. We used genomic DNA from peripheral blood leukocytes. For PCR, we used 0.75 U MyTaq polymerase (Table 3.16) and 0.4 µM primers in a total volume of 20 µl. All primer sequences are listed in the supplementary data section 9.1.1 in Table 9.1 (*AQP12A* and *AQP12B*), Table 9.2 (*CTRL*), Table 9.3 (*NR5A2*), Table 9.4 (*RBPJL*), and Table 9.5 (*HNF1a*) and are designed to bind to the intronic region flanking the exons which contains the alleged mutation. Cycle conditions were as follows: initial denaturation for 1 min at 95 °C; 50 cycles of 20 sec denaturation at 95 °C, 40 sec primer annealing at 60 °C and 2 min extension at 72 °C; and a final extension step for 2 min at 72 °C in a thermocycler. For the amplification of exon 4 of *AQP12A* or exon 3 of *AQP12B*, we performed a PCR with 0.75 U OneTaq polymerase (Table 3.16), 0.4 mM deoxynucleoside triphosphates (dNTPs) and 0.1 µM primers in a total volume of 22 µl. Cycle conditions were as follows: initial denaturation for 5 min at 95 °C; 50 cycles of 20 sec at 95 °C, 40 sec at 60 °C and 2 min at 68 °C; and a final extension step for 5 min at 68 °C. In case of PCR products longer than 3 kb, we used Platinum SuperFi II DNA Polymerase (Table 3.16). For this, 0.4 µl SuperFi II polymerase, 160 µM dNTPs and 0.16 µM primers in a total volume of 20 µl was used. Cycle conditions were as follows: initial denaturation for 30 sec at 98 °C; 45 cycles of 10 sec at 98 °C, 10 sec at 60 °C and 2 min at 72 °C; and a final extension step for 5 min at 72 °C.

To confirm successful PCR amplification, we performed an agarose gel electrophoresis. A 5 µl aliquot of the PCR products was mixed with 1 µl 6X Loading dye (see Table 3.4), loaded onto a 1.5 % agarose gel (with HD Green: 5 µl/100 ml agarose) in 1X TAE buffer (Table 3.4) and ran for 45 min at 100 V. We digested the PCR products with 0.25 µl Antarctic phosphatase and 0.25 µl exonuclease I (Table 3.16) in a thermocycler (40 min at 37 °C; 20 min at 85 °C) and sent them for Sanger sequencing with the Mix2Seq Kit (Table 3.20) to Eurofins Genomics (ABI 3730XL DNA Analyzer; Ebersberg, Germany). For this, we mixed 2 µl PCR product with 2 µl sequencing primer [10 µM] and filled it to 17 µl total volume with HPLC grade H_2O. Sequencing primers are listed in the supplementary data section 9.1.1. Those solely used for sequencing are marked with $^{(*)}$. The results were then visualized in the DNA sequencing software Chromas 2.6.6 (Technelysium, South Brisbane, Australia).

5.1.2 Statistical analysis of AQP12

To analyze the genotype results on *AQP12A* and *AQP12B*, we used the Fisher's exact test with a 2x3 contingency table for genotypes and a 2x2 contingency table for heterozygous carriers, allele frequencies, dominant and recessive models using the Interactive Statistical Calculation Pages (https://statpages.info; retrieved in July 2022). A two-tailed *p*-value less than 0.05 was considered significant. We tested the distributions of the genotypes for the Hardy-Weinberg equilibrium.

5.2 Vector plasmid preparations

5.2.1 Human expression vector preparation via MaxiPrep

We ordered human expression vectors from GenScript as listed in chapter 3.5.1. Upon arrival, the lyophilized vectors were dissolved in 20 µl HPLC grade H_2O and stored at -20 °C. Prior to their usage in any experimental setup, we needed the vector to be retransformed in *E. coli* cells for a higher yield and its correct sequence to be verified. We mixed 1 µl vector DNA carefully with 25 µl freshly thawed Stellar™ chemo competent *E. coli* cells on ice and incubated them for 15 min. After a heat shock of 1 min at 42 °C, we put the cell-DNA mix back on ice for another minute to cool down. The mixture was removed from ice and 250 µl room-temperated SOC medium was added. The cells incubated at 37 °C and 300 rpm on a Thermomixer compact for 30 min. We performed subsequent work in a sterile environment using a Bunsen burner when the retransformed *E. coli* were added to 30 ml LB-medium fortified with 1 mg/ml ampicillin. The samples then incubated over night at 37 °C and 190 rpm in a bacterial shaker for a maximum of 16 h.

The next day, we transferred an aliquot onto a LB-agar plate with 1 mg/ml ampicillin and incubated at 37 °C for 6 h in a bacteria incubator. This master plate was used for a later re-growth of the cells for vector preparation if needed. For the MaxiPrep, the 30 ml overnight culture was transferred into a 50 ml centrifugation tube and centrifuged at 25,000 g for 5 min (Heraeus Multifuge X3R). The medium was discarded, and the pellet was completely dissolved in 2 ml FlexiPrep solution 1 (Table 3.5) with 0.1 % RNase (Table 3.16) by gentle shaking. Transferred into a new 15 ml centrifugation tube, we added 2 ml FlexiPrep solution 2 (Table 3.5), inverted the tube carefully and incubated it 5 min for alkaline lysis. Afterwards, 2 ml FlexiPrep solution 3 (Table 3.5) were added. The sample was inverted again until the mixture showed a white flaky precipitation. For the separation of the dissolved plasmid DNA from the white cell debris, we centrifuged the samples again for 10 min at 25,000 g. The supernatant containing the DNA was then transferred into a new 15 ml centrifugation tube and mixed with 4.2 ml isopropanol. Following a 30 min incubation at -20 °C, we centrifuged the samples at 25,000 g for 30 min. We then discarded the supernatant completely with an

additional short centrifugation and dissolved the DNA pellet in 200 µl nuclease-free H_2O. For purification of the total DNA, we used the Wizard SV Gel and PCR Kit (Table 3.20) according to the manufacturer's protocol (supplementary section 9.6). The final elution of the DNA in 150 µl was done twice and the concentration was measured at the NanoPhotometer® N60.

For the verification of the correct sequence, Sanger sequencing was carried out as described in chapter 5.1 with the primers pcDNA3.1-2F and pcDNA3.1-4R (Table 9.9). We visualized the results in the DNA sequencing software Chromas 2.6.6 and analyzed them using the SnapGene Viewer Software and the Emboss Needle Tool provided by EMBL-EBI (Madeira et al. 2022). This holds true for all future Sanger sequencing results.

5.2.2 Molecular Cloning of target promoters into a reporter gene vector

The promoter region of a potential transcription factor target gene was selected from literature or by using the transcription factor binding site (TFBS) prediction tool JASPAR2020 (Castro-Mondragon et al. 2022) as listed in Table 5.1. Human gDNA samples from healthy individuals served as template for the PCR with primers listed in Table 9.6. We designed the primers to introduce a restriction enzyme cutting site at the 5'-end (*NheI* or *XhoI*) and the 3'-end (*XhoI* or *HindIII*) of the PCR product to clone them into the multiple cloning site of the pGL4.22 basic luciferase vector (Table 3.14).

Promoter region amplification

For most promoter regions, we conducted a PCR in a total volume of 55 µl with 1 U HiFi polymerase (Table 3.16) and 0.4 µM primers. The reaction was prepared in four equal approaches for a high yield. For this, we pipetted 5 µl gDNA template into a PCR strip and prepared a master mix for four reactions. The master mix was gently shaken and centrifuged shortly before adding the polymerase. Afterwards, 50 µl mix was added to the gDNA template, centrifuged (5430, Eppendorf, Hamburg, Germany) and the PCR run in a thermocycler. Cycle conditions were as follows: initial denaturation for 1 min at 95 °C; 35 cycles of 15 sec at 95 °C, 15 sec at 60 °C and 2 min at 72 °C; and a final extension step for 2 min at 72 °C. For the *INS* promoter, the PCR worked best with 2.4 µl SuperFi II polymerase (see Table 3.16), 160 µM dNTPs and 0.16 µM primers in a total volume of 55 µl. Here, the cycle conditions were as follows: initial denaturation for 30 sec at 98 °C; 45 cycles of 10 sec at 98 °C, 10 sec at 60 °C and 2 min at 72 °C; and a final extension step for 5 min at 72 °C. For the amplification of the promoters of *ALB* and *FGB*, we used 1 U Phusion polymerase (Table 3.16), 0.2 mM dNTPs and 0.4 µM primers. Cycle conditions were set here as follows: initial denaturation for 2 min at 98 °C; 35 cycles of 1 min at 98 °C, 1 min at 60 °C and 1.5 min at 72 °C; and a final extension step for 10 min at 72 °C.

Table 5.1: Promoter regions for molecular cloning with potential transcription factor binding site

Gene	Name	Promoter region	Potential target gene of
CPA2	Carboxypeptidase A2	-501/+20bp	NR5A2 (Holmstrom et al. 2011)
CPA1	Carboxypeptidase A1	-2051/+27bp	NR5A2 (Hale et al. 2014); preliminary work)
CEL	Carboxyl ester lipase	-551/+22bp	NR5A2 (Holmstrom et al. 2011); RBPJL (preliminary work)
CELA3A	Chymotrypsin like elastase family member 3A	-1965/+23bp	NR5A2 (JASPAR2020; (Castro-Mondragon et al. 2022)
PRSS2	Serine protease 2	-1479/+18bp	RBPJL (JASPAR2020; (Castro-Mondragon et al. 2022)
SPINK1	Pancreatic secretory trypsin inhibitor	-1154/+215bp	HNF1a/PTF1 (Boulling et al. 2011); preliminary work)
REG1A	Regenerating family member 1 alpha	-1136/+384bp	HNF1a (JASPAR2020; (Castro-Mondragon et al. 2022); RBPJL (preliminary work)
GP2	Glycoprotein 2	-1402/+1084bp	RBPJL/PTF1 (preliminary work)
CYP11B1	Steroid-11b-hydroxylase	-1061/+29bp	NR5A2 (Wang et al. 2001; Sirianni et al. 2002)
CYP17	Steroid-17a-hydroxylase	-2856/+31bp	NR5A2 (Hanley et al. 2001; Sirianni et al. 2002)
INS	Insulin	-2314/+366bp	HNF1a (Okita et al. 1999)
IAPP	Islet amyloid polypeptidase	-2058/+494bp	HNF1a (Green et al. 2003)
MTTP	Microsomal triglyceride transfer protein	-1613/+201bp	HNF1a (Au et al. 2008)
CALR	Calreticulin	-1802/+93bp	NR5A2 (Arvaniti et al. 2016)
ALB	Albumin	-129/+42bp	HNF1a (Najmi et al. 2017; Althari et al. 2020)
FGB	β-fibrinogen	-120/-29bp	HNF1a (Chi et al. 2002)

Exemplified explanation on promoter sequence -501/+20 bp = 501 base pairs in 5'-direction of 5'-UTR and 20 base pairs into 5'-UTR.

We used a 1.5 % agarose gel to purify the PCR products in gel electrophoresis (100 V, 40 min) and visualized them with the FastGene FAS V Gel Imaging System. The expected fragments were cut out and purified in one tube using the Wizard SV Gel and PCR Kit (Table 3.20) according to the manufacturer's protocol (supplementary section 9.6). The final elution of the products was done twice in a total volume of 38 µl.

Restriction digest and ligation of the PCR product and vector pGL4.22

Depending on the introduced restriction enzyme cutting site, we digested the PCR product with 1.5 µl FastDigest enzyme 1 (*NheI* or *XhoI*), 1.5 µl FastDigest enzyme 2 (*XhoI* or *HindIII*) and 4 µl 10X FastDigest Green buffer in a total volume of 40 µl. Simultaneously, 6 µg pGL4.22 vector was digested with the same enzymes. Additionally, a control "digestion" of the vector was prepared without enzymes. We incubated the reactions at 37 °C for 1 h in a Thermomixer compact. Another gel electrophoresis for separation with subsequent purification (Wizard SV Gel and PCR Kit; Table 3.20) was performed and the concentration was measured at the NanoPhotometer® N60.

To calculate the ratio of how much promoter fragment and pGL4.22 vector needs each to be added to the reaction, we used the Promega formula:

$$\text{ng insert} = \frac{\text{ng vector} \times \text{kb insert}}{\text{kb vector}} \times 3$$

The ligation reaction was mixed with 100 ng vector, the calculated ng insert, 2 µl T4 ligation buffer, and 1 µl T4 ligase (Table 3.16) in a total volume of 20 µl. For a control ligation, we replaced the volume for the insert with HPLC grade H_2O. The ligation then incubated over night at RT.

Vector transformation into *E. coli* and colony PCR

For transformation, 3 µl of the ligation mix were carefully added to 25 µl freshly thawed Stellar™ chemo competent *E. coli* cells on ice and incubated for 15 min. After a heat shock of 1 min at 42 °C, we put back the *E. coli* on ice for 1 min to cool down and added 250 µl room-temperated SOC medium. The cells were incubated at 37 °C and 300 rpm on a Thermomixer compact for 30 min. Subsequent work was performed in a sterile environment using a Bunsen burner. We distributed the transformed *E. coli* equally onto a LB-agar plate with 25 µg/ml ampicillin and incubated them over night at 37 °C.

The next day, we picked single colonies to test for successful transformation. For this, the colony was dissolved in 21 µl HPLC grade H_2O in a PCR plate and 2 µl of the *E. coli*-water mix were transferred onto a new LB-agar plate with ampicillin (master plate) for incubation. We conducted a PCR with 0.85 U AmpliTaq polymerase, 1.5 mM $MgCl_2$, 0.1 mM dNTPs, and 0.12 µM primer added to the reaction. Cycle conditions were as follows: initial denaturation for 12 min at 95 °C; 25 cycles of 20 sec at 95 °C, 40 sec at 60 °C and 1.5 min at 72 °C; and a final extension step for 2 min at 72 °C. A following gel electrophoresis allowed the discrimination of colonies with the correct vector construct from false positives by size.

To verify the correct sequence of the vector construct, we inoculated positive colonies regrown on the master plate in 3 ml LB-medium with 1 mg/ml ampicillin and incubated them over night at 37 °C and 190 rpm in a bacterial shaker for a maximum of 16 h.

Vector preparation (MiniPrep) and sequence verification

The next day, we performed a MiniPrep. For this, the 3 ml overnight culture was transferred into a 2 ml reaction tube and centrifuged at 21,330 g for 1 min (5415D, Eppendorf). The medium was discarded, and the pellet was completely dissolved in 200 µl FlexiPrep solution 1 (Table 3.5) with 0.1 % RNase (Table 3.16) by vigorous shaking. With 200 µl FlexiPrep solution 2 (Table 3.5) added, we carefully inverted the tube and incubated it 5 min for alkaline

lysis. Afterwards, we added 200 µl FlexiPrep solution 3 (Table 3.5) and inverted the sample again until the mixture showed a white flaky precipitation. For the separation of the vector DNA from the white cell debris, the samples were centrifuged again for 5 min at 21,330 g. We then transferred the supernatant containing the vector DNA into a new 1.5 ml reaction tube with 420 µl isopropanol and mixed well. Following a 10 min incubation at RT, we centrifuged the samples at 21,330 g for 10 min. The supernatant was then discarded completely with an additional short centrifugation and the vector DNA pellet was dissolved in 20 µl nuclease-free H_2O. We further purified the vector DNA using the Wizard SV Gel and PCR Kit (Table 3.20). We eluted the DNA twice in total volume of 50 µl and measured concentration at the NanoPhotometer® N60. The purified vector construct was then sent for Sanger sequencing as described above in chapter 5.2.1 with the sequencing primers RVprimer3, Luci-R-Seq, and a promoter specific primer; all listed in Table 9.7.

5.2.3 Primer mutagenesis to introduce a *PacI* site into the pcDNA3.1+/C-(K)-DYK vector

A restriction enzyme cutting site for *PacI* at the 3'-end of the Flag-tag region was introduced into the pcDNA3.1+/C-(K)-DYK vector (Table 3.13) using primer mutagenesis. In two separate PCR, we amplified 200 ng vector with 1 U Phusion polymerase (Table 3.16), 0.2 mM dNTPs, and 0.4 µM primer each (PCR 1: pcDNA3.1 *EcoRI* F + pcDNA3.1 *PacI* R; PCR 2: pcDNA3.1 *PacI* F + pcDNA3.1 *BpiI* R; see Table 9.9) in a total volume of 100 µl. Cycle conditions were as follows: initial denaturation for 2 min at 98 °C; 35 cycles of 1 min at 98 °C, 1 min at 56 °C and 1.5 min at 72 °C; and a final extension step for 10 min at 72 °C. The reverse primer pcDNA3.1 *PacI* R in PCR 1 and the forward primer pcDNA3.1 *PacI* F in PCR 2 were designed to introduce the *PacI* cutting site -TTAAT'TAA-. Subsequent to a 1.5 % agarose gel electrophoresis (100 V, 45 min), we visualized the PCR products with the FastGene FAS V Gel Imaging System, purified them using the Wizard SV Gel and PCR Kit (Table 3.20) and eluted the products in 38 µl nuclease-free H_2O. A recombinant PCR amplified the two resulting PCR products with 0.5 U Phusion polymerase and 0.2 mM dNTPs. Cycle conditions were as follows: 5 cycles of 30 sec at 95 °C, 1 min at 55 °C and 2 min at 72 °C; and a final extension step for 10 min at 72 °C. To this reaction, we added another 0.5 U Phusion polymerase, 30 µM dNTPs, and 0.5 µM primer (pcDNA3.1 *EcoRI* F + pcDNA3.1 *BpiI* R; see Table 9.9) for a third recombinant PCR. Cycle conditions were as follows: 10 cycles of 1 min at 95 °C, 1 min at 58 °C and 2 min at 72 °C; and a final extension step for 10 min at 72 °C. After another agarose gel electrophoresis (100 V, 45 min), we cut the expected fragments and purified them in a final volume of 38 µl.

All further steps from restriction digest with *EcoRI* and *BpiI* as well as the ligation into the pcDNA3.1+/C-(K)-DYK vector to MiniPrep were performed as described in chapter 5.2.2. For

the final verification of the *PacI* cutting site in the vector sequence, we sent the vector for Sanger sequencing with the primers pcDNA3.1-2F and pcDNA3.1-4R (Table 9.9).

5.2.4 Primer mutagenesis to reposition the Flag-tag in the NR5A2 wildtype vector

As the C-terminal position of the Flag-tag in the pcDNA3.1_NR5A2-WT-iso1-Flag vector did not lead to an expression of NR5A2, we repositioned the Flag-tag onto the N-terminal end of the protein. For this, primer mutagenesis was performed similar to the procedure described in chapter 5.2.3. For the first recombinant PCR, we used the primers pcDNA3.1 *MunI* F with pcDNA3.1 Flag NR5A2 R and pcDNA3.1 Flag NR5A2 F with NR5A2 *PacI* R (Table 9.9). The final PCR product was digested with *MunI* and *PacI* and eventually ligated into the prior produced pcDNA3.1_Flag_*PacI* vector (see chapter 5.2.3).

5.2.5 Primer mutagenesis for removal of a Flag-tag in expression vectors

For some experimental approaches, we had to use expression vectors without a Flag-tag. Therefore, we removed the C-terminal tag by primer mutagenesis to replace it with a sole stop codon. We performed primer mutagenesis comparable to the procedure described in section 5.2.3. However, only one recombinant PCR had to be performed with the expression vector containing the Flag-tag as DNA template. Primers were pcDNA3.1 *HindIII* F with either pcDNA3.1 NR5A2 *PacI* R, pcDNA3.1 RBPJL *PacI* R or pcDNA3.1 HNF1a *PacI* R depending on the insert coding sequence. We digested the insert product as well as the empty vector pcDNA3.1_Flag_*PacI* with *HindIII* and *PacI* and eventually ligated them into the prior produced pcDNA3.1_Flag_*PacI* vector (see 5.2.3). Following transformation of *E. coli*, we purified the produced vector construct via MiniPrep and verified the correct sequence by Sanger sequencing using pcDNA3.1-2F and pcDNA3.1-4R as primers (Table 9.9).

5.3 Cell culture experiments

5.3.1 General culturing of cells

Cell lines (Table 3.12) were stored in 1 ml vials in liquid nitrogen in their corresponding growth medium (see Table 5.2) containing additional 10 % DMSO. To take cells into culture, we mixed them carefully with 11 ml pre-warmed growth media in a 15 ml centrifugation tube and spin down for 5 min at 200 g (5810, Eppendorf). We then replaced the media with 12 ml fresh growth media to remove all remaining DMSO. The cell suspension was transferred into a T75 tissue culture flask and kept in 37 °C and 5 % CO_2.

We passaged cells with an 80-90 % confluency. For this, the media was aspirated carefully from the cell layer and cells were washed with warm PBS. To detach the cells from the flask, we added a trypsin-EDTA solution and incubated the cells for roughly 5 min at 37 °C and

5 % CO_2. Added growth media then stopped the trypsin activity and cells were re-suspended. Depending on the cell line and its growth rate, we transferred a dilution of 1:10 1:20 into a new cell culture flask and kept it in 37 °C and 5 % CO_2. When we needed a higher yield, cells were regrown in a T175 tissue culture flask.

Table 5.2: compositions of growth media for culturing cell lines

Cell line	Growth medium composition
HEK293 / HEK293T	DMEM-hg 10 % FBS 1 % Pen/Strep
Panc1	DMEM-hg 10 % FBS 1 % Pen/Strep

Medium was fortified with 10 % DMSO for freezing cells in liquid nitrogen for long-term storage.

At the beginning of experiments, an aliquot of 20 µl of the cell suspension after trypsin treatment was mixed with 20 µl of a 0.2 % trypane-blue solution (1:1 dilution) and pipetted onto a Neubauer improved cell-counting chamber (0.0025 mm², depth 0.100 mm). We determined the total cell count to seed a certain number of cells. By using the following equations, we obtained the number of cells per ml as well as the volume needed for seeding.

$$\frac{\text{cells counted}}{\text{number of squares counted}} \text{x2x10}^4\text{=cells/ml}$$

$$\frac{\text{cells/ml}_{\text{final}}\text{x volume}_{\text{final}}}{\text{cells/ml}_{\text{counted}}} = \text{volume}_{\text{needed}}$$

Cells were seeded to have an 80 % confluency the next day for transfection. For most experiments, we used a 6-well or a 48-well format. For a high yield in nuclear extract (NE), the cells were seeded in 10 cm petri dishes.

Table 5.3: Number of cells seeded depending on cell line and seeding format.

Cell line	6-well format	48-well format	10 cm petri dish
Surface area	9.6 cm^2	0.83 cm^2	58.8 cm^2
HEK293 / HEK293T	$0.5x10^6$	$8.0x10^4$	$3.0x10^6$
Panc1	$0.5x10^6$	--	$3.0x10^6$

Number of seeded cells is given in cells/well.

5.3.2 Transient overexpression for protein or RNA isolation

Cells were seeded in a 6-well format as listed in Table 5.3 and incubated for 24 h to reach a confluency of 80 % at the time of transfection. On the day of transfection, we performed a transient overexpression of vectors with Lipofectamine 2000. As preparation, mix A (250 µl Opti-MEM® I with 4 µg total expression vector DNA) and mix B (250 µl Opti-MEM® I with 10 µl Lipofectamine 2000) incubated separately for 5 min at RT. Depending on single

transfection or co-transfection of expression vectors, the amount of each vector DNA in mix A was always adjusted to a combined total of 4 µg. After the incubation, we transferred mix B into mix A and incubated it for another 20 min at RT. In each well of the cell culture plate, 500 µl of the total of 2 ml medium was then replaced by the transfection mix and incubated at 37 °C and 5 % CO_2. After either 6 h incubation or 18 h incubation, we washed the cells once with room-temperated PBS. The cells were then incubated with 2 ml new Opti-MEM® at 37 °C and 5 % CO_2. We harvested the cell 72 h post-transfection.

5.3.3 Endoplasmic reticulum stress induction

To establish positive controls for ER stress analyses, we treated cells with either 5 µg/ml tunicamycin (TM, stock: 6 mM), 0.5 µM thapsigargin (TG, stock: 0.1 mM), or 2 mM DTT (stock: 1 M) and incubated them for 6 h until cell harvest and RNA isolation. These reagents are known to induce ER stress and therefore served as positive controls.

5.3.4 Transient overexpression for Luciferase assay

We seeded the cells in a 48-well format as listed in Table 5.3 with a 24 h incubation to reach a confluency of 80 % at the time of transfection. On the day of transfection, we co-transfected human expression vector DNA with reporter gene DNA and an internal standard DNA (pRL-Ubi, see Table 3.14) in a ratio of 1:1 and 1:0.5, respectively. We transfected with Fugene HD. For this, we replaced 2 ml growth medium by 2 ml Opti-MEM® I approximately 1 h before transfection. The transfection mix was prepared in a 1.5 ml reaction tube, each condition for a technical duplicate: 20 µl Opti-MEM® I, 500 ng total vector DNA and 1.25 µl Fugene HD. After an incubation for 20 min at RT, we carefully added 12 µl transfection mix to the cells and further kept them at 37 °C and 5 % CO_2. After 24 h, the cells were washed with PBS and growth medium was added for another 24 h incubation until cell harvest.

5.3.5 Cell culture for nuclear extraction

We conducted transient overexpression in HEK293T, HEK293 and Panc1 cells for nuclear protein extraction. Cells were seeded in a 10 cm petri dish 24 h prior to transfection as listed in Table 5.3. We conducted transfection as described in section 5.3.2. Due to the larger surface area in the 10 cm petri dish, we adjusted the volume of the transfection mixes to mix A of 1.5 ml Opti-MEM® I with 24 µg total expression vector DNA and mix B of 1.5 ml Opti-MEM® I with 60 µl Lipofectamine 2000. We changed the medium after 6 h and harvested the cells 72 h post-transfection.

5.4 Sample preparations

5.4.1 Total lysate preparation

We harvested cells 72 h post-transfection. For functional analyses of secreted proteins such as CTRL, conditioned medium was collected in a 2.0 ml reaction tube on ice. Otherwise, medium was aspirated, and cells were washed once with ice-cold PBS. With additional 500 µl PBS, we scraped off the cells and collected them in a 1.5 ml reaction tube on ice. Both supernatant and cells were centrifuged for 3 min at 500 g and 4 °C (5417R, Eppendorf). We transferred the supernatant into a new 1.5 ml reaction tube on ice. PBS was removed from the cells and the pellet was stored at -80 °C. For total lysate preparation, we thawed the cell pellets on ice and resuspended them in 50 µl RIPA buffer (Table 3.6) with 1 µl PMSF or in activity assay buffer (AAB; Table 3.9), depending on subsequent experiments. The samples were then exposed twice to sonification (5 beats, cycle 0.5, amplitude 40) and put back on ice. After centrifugation for 5 min at 1,000 g, we transferred the supernatant containing total lysate into a new 1.5 ml reaction tube.

We measured total protein concentration in technical triplicates in a 96-well microplate using the Bradford Protein Assay supplied by Bio-Rad (Feldkirchen, Germany). The assay originally developed by Marion M. Bradford in 1976 (Bradford 1976) allows the determination of total protein concentration using colorimetry. The reagent contains an acidic Coomassie® Brilliant Blue G-250 dye which binds to amino acid residues. In response, the dyes absorbance maximum shifts from 465 nm to 595 nm (Bradford 1976) which we measured photometrically (VarioScan™ Flash Multimode Reader, Thermo Fisher Scientific). A standard curve with defined BSA concentrations enables the calculation of the protein concentration. For this, a 4 mg/ml concentration of BSA was dissolved in the corresponding buffer and a range of 1:1 dilutions was prepared from 4 mg/ml to 0.0625 mg/ml and sole lysis buffer as blank.

5.4.2 Protein precipitation from supernatant

We removed the conditioned medium from the cells and collected it in a 2.0 ml reaction tube on ice. We centrifuged the samples for 5 min at 200 g and 4 °C (5417R, Eppendorf) to get rid of cell debris and transferred the supernatant to a new 1.5 ml reaction tube. For the precipitation of proteins, we mixed 360 µl supernatant with 40 µl TCA and centrifuged for 10 min at 16,000 g at RT (5415D, Eppendorf). The supernatant was then discarded, the protein pellets dissolved in 15 µl Laemmli buffer-DTT solution (Table 3.7) and denatured for 5 min at 90 °C in a thermomixer comfort. After a short centrifugation, we put the samples back on ice to avoid evaporation and further processed them for Western Blot analysis.

5.4.3 Total RNA isolation

Cells in 6-well plate were washed with 1 ml cold PBS. As the expected number of pelleted cells is < $5x10^6$ cells, we added 350 µl ice-cold RLT buffer (RNeasy Mini Kit, Qiagen) to lyse the cells. The lysate was collected from the dish with a cell scraper and transferred into a QIAshredder spin column (Table 3.20) on ice. For sample homogenization, we centrifuged the columns for 2 min at 25,000 g (5417R, Eppendorf) and mixed the flow-through with 350 µl 70 % ethanol. For subsequent RNA isolation, we used the RNeasy Mini Kit (Table 3.20; Qiagen) according to the manufacturer's protocol (supplementary section 9.6). All pipetting steps were done on ice, while centrifugation steps were set at 15 °C. In a final step, we eluted the RNA twice with 35 µl RNAse-free water and stored it at -80 °C until further use.

5.4.4 Reverse transcription of RNA into cDNA

The NanoPhotometer® N60 gave us the RNA concentration to calculate and prepare a working solution of 200 ng/µl. For the reverse transcription of RNA into cDNA, we used the QuantiTect® Reverse Transcription Kit (Table 3.20; Qiagen) and instructions were followed as provided (supplementary section 9.6). In a total reaction volume of 20 µl, 1 µg RNA was used. We made alterations to the manual for the incubation durations: the gDNA elimination was carried out for 5 min at 42 °C and the reverse transcription was set to 25 min at 42 °C. The final cDNA was diluted 1:10 with HPLC grade H_2O and stored at -20 °C.

5.4.5 Nuclear protein extraction

For a yield of 2-3 µg/µl, we used the Nuclear Extract Kit (Table 3.20; Active Motif Europe) according to the supplier's manual (supplementary section 9.6). We prepared all buffers fresh on ice on the day of the harvest as provided by the kit. Due to precipitation, the phosphatase inhibitor was prior heated on a magnetic stirrer to 50 °C for 10 min. We washed the cells once with 6 ml ice-cold 1X PBS containing 0.05 % phosphatase inhibitor and added another 1 ml PBS/phosphatase inhibitor. After scraping the cells off the dish, we transferred them to a pre-cooled 1.5 ml reaction tube and centrifuged them for 5 min at 200 g and 4 °C (5417R, Eppendorf). We discarded the supernatant and performed all subsequent steps on ice. The cell pellets were carefully resuspended with 500 µl 1X hypotonic buffer as provided by the kit and incubated for 15 min on ice in order for the cells to swell. After adding 25 µl detergent, we gently shook the samples on a vortexer for 10 sec at highest level and centrifuged them for 30 sec at 14,000 g and 4 °C. The supernatant being the cytoplasmic fraction was aliquoted to new 1.5 ml reaction tubes and stored at -80 °C until further use. For the nuclear extraction, we resuspended the remaining pellet in 50 µl complete lysis buffer (provided components in the kit: lysis buffer AM1, 10 mM DTT, 0.01 % protease inhibitor cocktail) and shook it for 10 sec at highest level. The samples were incubated at 4 °C for 30 min on a shaker and again

vigorously shaken for 30 sec at highest level. After a subsequent centrifugation for 10 min at 14,000 g and 4 °C, we transferred and aliquoted the supernatant which contains the NE into new 1.5 ml reaction tubes. For concentration measurement, we performed a Bradford Protein Assay as described in chapter 5.4.1. The NE was stored in -80 °C until further use for Western blot or electrophoretic mobility shift assay (EMSA).

5.5 SDS-PAGE and Western blot

For protein detection from overexpression experiments, we conducted SDS-PAGE with subsequent Western blot. Samples were prepared on ice. For this, we mixed 10-20 µg total lysate with ddH_2O to a volume of 15 µl. To each sample, 5 µl 4X Laemmli buffer (Table 3.8) was added, mixed well and shortly centrifuged (5417R, Eppendorf). We loaded the final volume of 20 µl onto a 10 % SDS-polyacrylamide gel (Table 5.4). In a separate lane, 1 µl of a 1:10 dilution of PageRuler™ Prestained Protein Ladder served as indicator for protein size (kDa). In the case of secreted proteins, we loaded the total volume of 15 µl precipitated supernatant as described in section 5.4.2 onto the gel. SDS-PAGE ran in 1X SDS running buffer (Table 3.8) at 120 V until the dye front left the stacking gel followed by 45 min at 160 V for separation of the proteins.

Table 5.4: 10 % SDS-polyacrylamide gel preparation for SDS-Page

10 % separation gel	
Amount	Components
373 mM	Tris (pH 8.8)
0.1 %	SDS (v/v)
10 %	Acrylamide (v/v, ROTIPHORESE®Gel 30 (37.5:1)
0.06 %	TEMED (v/v)
0.11 %	APS (v/v)
4.5 % stacking gel	
Amount	Components
118 mM	Tris (pH 6.8)
0.1 %	SDS (v/v)
4.5 %	Acrylamide (v/v, ROTIPHORESE®Gel 30 (37.5:1)
0.08 %	TEMED (v/v)
0.06 %	APS (v/v)

Gel format is 8.6x6.8 cm (width x height)

We then transferred the separated proteins onto a nitrocellulose membrane by wet blot in 1X transfer buffer (Table 3.8) for 30 min at 360 mA. The membrane then incubated for 1 h with 5 % BSA blocking solution (5 % BSA fraction V blotting grade in 1X PBS) at RT on a vertical shaker (IKA® HS501 digital, IKA). The primary antibody solution consisted of 5 ml 5 % BSA blocking solution with 0.1 % Tween20 and incubated with the membrane over night at 4 °C on a vertical shaker (KM-2, Edmund Bühler). The dilutions of the primary antibodies are listed in Table 3.17. The following day, we washed the membrane three times with 1X

PBS-T (Table 3.8; 0.05 % Tween20) each for 10 min at RT on a vertical shaker. The secondary antibodies were diluted 1:10,000 in a 5 ml 1X PBS-T solution as listed in Table 3.18. The membrane then incubated for 2 h at RT on a vertical shaker protected from light followed by another three washing steps: twice with 1X PBS-T and once with 1X PBS (Table 3.8). We visualized the proteins using the Azure Sapphire™ Biomolecular Imager (Biozym).

5.6 Activity assay

5.6.1 Enzyme activity

To investigate the proteolytic capacity of secreted CTRL, we performed activity assays. Since CTRL cleaves polypeptides at tyrosine and phenylalanine residues (Szabó and Sahin-Tóth 2012), we used N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (AAPF) and N-succinyl-Ala-Ala-Pro-Tyr-p-nitroanilide (AAPY) as substrates. When cleaving the phenylalanine or tyrosine residue as shown in Figure 5.1, p-nitroaniline (F-pNA or Y-pNA) is released as yellow color, whose absorption can be measured at 405 nm.

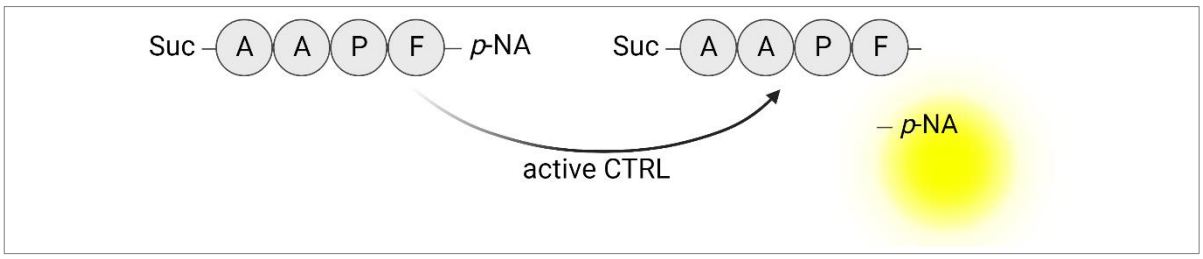


Figure 5.1: Schematic display of substrate AAPF digestion by activated CTRL.

We dissolved the substrate powder in DMF (1 µl for each 1 mg) and then added activity assay buffer (AAB; Table 3.9) to a final concentration of 0.2 mM. On the day of the assay, supernatant as well as cells were harvested. To remove cell debris, we shortly spun down the supernatant and processed the cells for total lysate preparation in AAB as described in section 5.4.1. For the initial activation of CTRL, the samples incubated in a 96-well plate with porcine trypsin for exactly 2 h at 37 °C with different conditions as listed in Table 5.5. Each condition was performed as technical triplicate.

Table 5.5: Conditions for the activity assay for 2 h trypsin activation (in a 96-well plate)

	AAB	Trypsin [0.1 mg/ml]	Tris/$CaCl_2$	Sample
Trypsin activity control	-	50 µl [1 mg/ml]	-	-
Trypsin control	-	50 µl	-	-
AAB control	50 µl	-	-	-
AAB control (activated)	48.75 µl	1.25 µl	-	-
Supernatant	-	1.25 µl	10 µl	38.75 µl
Supernatant control	-	-	10 µl	40 µl
Lysate	calculated*	1.25 µl	-	20 µg
Lysate control	calculated*	-	-	20 µg

*AAB volume was calculated depending on 20 µg total lysate used; AAB = activity assay buffer.

Subsequent to trypsin activation, we added 150 µl AAPF or AAPY to each well to measure the enzyme activity. Color intensity was measured immediately, after 4 min and after 8 min incubation at 405 nm (VarioScan™ Flash Multimode Reader, Thermo Fisher Scientific). To ensure the exact same enzymatic digestion time, we pipetted and measured each technical replicate separately.

5.6.2 Enzyme activation

As data of some variants of CTRL suggested an altered trypsin activation, we performed another assay approach. Instead of an exact incubation of 2 h, we activated the CTRL samples in a larger scale (1.5 ml reaction tube) with trypsin at 37 °C in a Thermomixer compact as stated in Table 5.6. At ten different time points, we extracted the samples from the activation reaction for substrate measurement, which allowed us to determine enzyme activation over time.

Table 5.6: Conditions for the enzyme activation with different time points of trypsin activation

	AAB	Trypsin [0.1 mg/ml]	Tris/$CaCl_2$	Sample
Trypsin control	-	1650 µl [1 mg/ml]	-	-
AAB control	1650 µl	-	-	-
AAB control (activated)	1608.75 µl	41.25 µl	-	-
Supernatant	-	41.25 µl	330 µl	1278.75 µl
Lysate	calculated*	41.25 µl	-	660 µg

*AAB volume was calculated depending on 20 µg total lysate used; AAB = activity assay buffer; preparation in 1.5 ml reaction tubes

We measured after 5 min, 10 min, 20 min, 30 min, 40 min, 50 min, 60 min, 80 min, 100 min, and 120 min activation. For each time point, the samples were mixed shortly in a Thermomixer compact and aliquots of 50 µl per well were withdrawn onto a 96-well plate. While the trypsin activation reaction continued until the next time point, we pipetted 150 µl

AAPF or AAPY onto each well and substrate cleavage was measured as described in chapter 5.6.1. Measurement times, however, were only after 4 min and after 8 min.

5.7 Reporter gene assay / Dual-luciferase assay

The dual-luciferase assay allowed the determination of a transcription factors transactivation activity. For this, we co-transfected a reporter gene (firefly luciferase) vector containing a promoter region of a target gene with an expression vector of a transcription factor. In an ideal situation, the overexpressed transcription factor binds to the promoter region on the reporter gene vector and initiates the expression of a firefly luciferase. With a subsequent assay, the activity of expressed firefly luciferase can be measured. When adding its substrate, the enzyme oxidizes D-luciferin to oxyluciferin which leads to an emission of light. This can be measured in a bioluminometer and quantified as relative light unit (RLU). An internal control, namely a vector with the gene for renilla luciferase, was co-transfected into the cells to normalize for the transfection efficiency, cell viability, and number of lysed cells.

On the day of luciferase assay, the transfected cells as described in chapter 5.3.4 were washed once with cold PBS. For each well, we added 40 µl 1X passive lysis buffer (PLB, diluted from 5X with PBS) and placed the 48-well plate on a Thermomixer comfort at RT for 10 min at 1050 rpm to allow cell lysis. We measured the luciferase activity in technical duplicates of biological triplicates at RT using the SIRIUS Luminometer (Berthold Detection Systems). First, we measured the empty luciferase PS tube to determine the background. Afterwards, we transferred 5 µl cells from the 48-well plate into the PS tube as well as 25 µl firefly substrate (Table 3.10) for a second measurement. Immediately after, 25 µl renilla substrate (Table 3.10) was added to the reaction for a third measurement. Each measurement was set to a delay of 3 sec and a measurement time of 10 sec.

To determine the RLU, we subtracted the blank value from the second and third measurement value. The firefly activity value was then divided by the renilla activity value for an internal standardization. Finally, the average of each technical duplicate was calculated in reference to the wildtype value, which was set to 1.

5.8 Target gene expression real-time polymerase chain reaction

We performed quantitative real-time polymerase chain reaction (qPCR) to determine RNA expression of target genes depending on the wildtype or mutant transcription factor. For each reaction, we mixed 2.1 µl H_2O with 400 nM forward and reverse primer, 100 nM TaqMan probe, and 5 µl SensiFast Probe No-Rox Kit (Table 3.20) to a final volume of 8 µl. All primer and probe sequences are listed in the supplementary data section 9.2. The housekeeper genes β-actin and GAPDH served as internal control and 2 µl cDNA (of 1:10 dilution, see chapter 5.4.4) were used as template. We carried out each reaction in a 384-well plate as

technical triplicates in the LightCycler® 480 II (Roche). The qPCR conditions were as follows: initial denaturation for 2 min at 95 °C; 55 cycles of 10 sec denaturation at 95 °C, and 30 sec annealing/extension at 60 °C. The LightCycler® 480 Software then allowed the analysis and calculation of the C_t value of each run.

To calculate the relative fold change in gene expression, we applied the delta-delta-C_t ($\Delta\Delta C_t$) method (Livak and Schmittgen 2001) by subtracting the mean C_t values of β-actin and GAPDH from each C_t value of the target genes (= ΔC_t). Additionally, the resulting mean ΔC_t of the wildtype transfection conditions was further subtracted from the ΔC_t of each different mutant transfection conditions (= $\Delta\Delta C_t$). We then calculated the fold change of the mutant gene expression with the formula $2^{-\Delta\Delta Ct}$ as the gene expression of the wildtype transfection condition was set to 1.

5.9 ER stress analysis

5.9.1 XBP1 splicing PCR

For amplification of spliced and unspliced X-box binding protein 1 (XBP1), we used 2 µl cDNA samples as templates (of 1:10 dilution, see chapter 5.4.4). The reaction further contained 0.75 U MyTaq polymerase (Table 3.16) and 0.4 µM primers (Table 9.10) in a total volume of 20 µl. Cycle conditions were as follows: initial denaturation for 5 min at 95 °C; 35 cycles of 30 sec at 95 °C, 30 sec at 60 °C and 30 sec at 72 °C; and a final extension step for 5 min at 72 °C.

We separated PCR products on a 3 % agarose gel (100 V, 180 min) and visualized them with the FastGene FAS V Gel Imaging System. Depending on the product size, the samples could be interpreted regarding ER stress. Spliced XBP1 has a product size of 416 bp indicating ER stress compared to unspliced XBP1 with a product size of 442 bp.

5.9.2 *BiP* expression qPCR

We analyzed the expression of the ER stress marker *BiP* via real-time PCR with *BiP* specific oligonucleotides (Table 9.11) as described in section 5.8.

5.10 Electrophoretic mobility shift assay (EMSA)

5.10.1 Annealing of Cy5-labelled oligonucleotides

We ordered a 45 bp promoter region containing the corresponding transcription factor binding motif as forward and reverse strand (Table 3.19) and dissolved in nuclease-free H_2O to a concentration of 100 µM. All forward strands are labelled at the 5'-end with a Cy5-fluorophor. For annealing the single-stranded (ss-) oligonucleotides, we mixed 7 µl each of both strands

and 6 µl TE buffer (Table 3.11) and heated the mix to 86 °C in a thermocycler, which lays at least 10 °C above the melting temperature of the single strands. After a gradual cooling of 5 °C/10 min down to 10 °C, the ss-oligonucleotides were separated from the annealed double-stranded (ds-) oligonucleotides in a 12 % acrylamide gel (see Table 5.7), which pre-ran for 30 min at 4 °C in 250 V. For this, we combined the 20 µl reaction mixture with 4 µl 10X loading buffer (Orange G; Table 3.11) and loaded it onto the gel. As an annealing control, we included the unlabeled reverse strands in the run (7 µl ss-oligonucleotide, 13 µl HPLC grade H_2O, and 4 µl Orange G). To avoid cross-contamination, the samples were loaded with at least two pockets empty in between. The electrophoresis was set to 200 V at 4 °C and ran for 5 h in 0.5X TBE buffer (Table 3.11). Afterwards, we visualized the bands under UV light, cut the ds-oligonucleotides from the gel and transferred them into a light-protected 1.5 ml reaction tube. Subsequently, we added 150 µl TE buffer to the samples and incubated them over night at 37 °C on a Thermomixer comfort with 1000 rpm to extract the ds-oligonucleotides from the gel. The next day, we cut a 1 ml filter tip with a scalpel to use the part with the filter on a new light-protected 1.5 ml reaction tube. We then added the solution with the ds-oligonucleotides onto the filter of the tip and centrifuged for 1 min at 3,300 g (5417R, Eppendorf). To wash the remaining samples through the filter, we added additional 30 µl TE buffer and spun down again for 1 min at 3,300 g. Eventually, we adjusted the concentration to 40 ng/µl with TE buffer using the NanoPhotometer® N60 and prepared diluted aliquots (1:10 dilution in H_2O) for subsequent EMSA experiments.

Table 5.7: 12 % acrylamide gel preparation for annealing oligonucleotides

Amount	Components
0.5X	TBE (pH 8.3)
12 %	Acrylamide (v/v, ROTIPHORESE®Gel 40 (37.5:1)
2.5 %	Glycerol (v/v)
0.075 %	APS (v/v)
0.05 %	TEMED (v/v)

Gel format is 17.0x18.0 cm (width x height)

5.10.2 EMSA

To test the DNA-binding capacity of a transcription factor, we performed EMSA. On the day of the experiment, a 5.3 % acrylamide gel (Table 5.8) pre-ran in 0.5X TBE buffer (Table 3.11; ≥30 min, 4 °C, 250 V) to remove excessive ammonia and persulfate ions, which can disturb the integrity of labile protein-DNA complexes. It also removes non-polymerized acrylic acid.

For the sample preparation, we diluted 20 µg NE of transfected HEK293 cells with HPLC H_2O to 6 µl. After adding 2 µl 5X binding buffer (Table 3.11) and 700 ng of poly dIdC (700 ng/µl stock) the samples spun shortly at 4 °C (5417R, Eppendorf). After an incubation on ice for 10 min, we added 1 µl Cy5-labeled ds-oligonucleotides (1 ng/µl working solution) and spun

the sample shortly again at 4 °C. A subsequent incubation for 20 min on ice (light protected) allowed the formation of potential DNA-protein complexes. Prior to loading the gel, we added 1.5 µl 10X loading buffer (Orange G, Table 3.11), shortly centrifuged the samples at 4 °C, vigorously mixed and shortly centrifuged them again. The total volume of 11.5 µl was then loaded onto the pre-run gel and native gel electrophoresis was performed (≥3 h, 4 °C, 250 V) until the running front reached the bottom of the gel. Due to the Cy5-labeled ds-oligonucleotides, we performed electrophoresis under a light protective hood. For visualization of DNA-protein complexes, we scanned the gel on the Typhoon TRIO+® Variable Mode Imager (Amersham) at high sensitivity at 633 nm and 100 µm resolution.

Table 5.8: 5.3 % acrylamide gel preparation for EMSA

Amount	Components
0.5X	TBE (pH 8.3)
5.3 %	Acrylamide (v/v, ROTIPHORESE®Gel 40 (37.5:1)
2.5 %	Glycerol (v/v)
0.075 %	APS (v/v)
0.05 %	TEMED (v/v)

Gel format is 17.0x18.0 cm (width x height)

6 Results

6.1 Genotyping for *AQP12A* and *AQP12B*

We conducted genotyping for the homologous genes *AQP12A* and *AQP12B* in 292 unrelated German CP patients and 143 German control subjects. Table 6.1 shows the results for *AQP12A*. In total, we discovered 12 nonsynonymous exonic alterations including one nonsense variant. The variants were in Hardy-Weinberg equilibrium except for p.T190M, as here a single homozygous individual in the control cohort was detected. Overall, the distribution of genotypes in patients and controls was similar (p>0.05). The p-values showed no significance (p>0.05) (table 1.1), even when we calculated the variants with a recessive (wildtype and heterozygous *vs*. homozygous for the minor allele) or a dominant model (wildtype *vs*. heterozygous and homozygous for the minor alleles) or the allele frequencies (data not shown). For bioinformatics analysis, we annotated CADD (Combined Annotation Dependent Depletion) and REVEL (Rare Exome Variant Ensemble Learner) predictions. The analyses showed, that with a rigorous CADD score ≥25, only the frameshift variant p.S140Tfs*24, which was found once in controls, was predicted deleterious. Using REVEL (score ≥0.75), none of the variants were predicted as deleterious, also because this tool cannot predict frameshift variants such as p.S140Tfs*24. With less stringent values applied (CADD ≥20 and REVEL ≥0.5), three patients (1.0 %; 2x p.L182Q, 1x p.Y219H) and one control (0.3 %; p.S140Tfs*24) carried a predicted deleterious variant (p=1.0).

AQP12B genotyping results are shown in Table 6.2. We detected a total of 19 missense and two nonsense variants. Here, all variants were in Hardy-Weinberg equilibrium. As with *AQP12A*, the distribution of *AQP12B* genotypes were similar between patients and controls (p>0.05) (table 1.2). When comparing allele frequencies or the distribution of variants using a recessive or dominant model or neither, all p-values were statistically insignificant (p>0.05) (data not shown). With a CADD score ≥25, only one *AQP12B* variant, p.S152Tfs*24, was predicted as deleterious while REVEL (score ≥0.5) resulted in no deleterious prediction of any of the variants. Notably, the allele frequency of p.S152Tfs*24 was higher in controls than patients (15.7 % vs. 13.4 %), with four controls (2.8 %) being homozygous.

Further data in the supplementary section 9.3 (supplementary Table 9.15 and Table 9.16) shows synonymous and adjacent intronic variants in both genes that we detected additionally. Again, for these variants there was no statistically significant difference between patients and controls (p>0.05) in any of the statistical calculations. For the intronic and synonymous variants, a less stringent CADD score (score ≥10) was used however, no variant was predicted to be deleterious.

Table 6.1: Non-synonymous *AQP12A* variants in German CP patients and controls

Exon	Position (hg38)	Nucleotide change	AA change	rs number	CADD	REVEL	CP (n=292) n (%)	Controls (n=143) n (%)	*p*-value
1	2:240691972	c.59C>G	p.A20G	rs141084461	0.19	0.021	1 (0.3)	0	1.0
1	2:240691981	c.68G>A	p.R23Q	rs149487583	22.2	0.099	1 (0.3)	0	1.0
1	2:240691996	c.83T>G	p.L28R (het) p.L28R (hom)	rs71428454	24.3	0.204	125 (42.8) 38 (13.0)	58 (40.6) 18 (12.6)	0.87
2	2:240692300	c.350T>C	p.M117T	rs58604656	2.61	0.045	0	1 (0.7)	0.33
2	2:240692368	c.419delG	p.S140Tfs*24	rs754059155	31.0	-	0	1 (0.7)	0.33
2	2:240692381	c.431G>C	p.R144P (het) p.R144P (hom)	rs142225246	7.22	0.440	38 (13.0) 0	13 (9.1) 1 (0.7)	0.13
2	2:240692422	c.472G>A	p.A158T (het) p.A158T (hom)	rs12477907	9.47	0.225	129 (44.2) 42 (14.4)	61 (42.7) 21 (14.7)	0.95
2	2:240692495	c.545T>A	p.L182Q	rs752347589	23.4	0.570	2 (0.7)	0	1.0
2	2:240692519	c.569C>T	p.T190M (het) p.T190M (hom)	rs201497460	24.2	0.409	5 (1.7) 0	2 (1.4) 1 (0.7)	0.43
3	2:240694516	c.655T>C	p.Y219H	-	24.4	0.529	1 (0.3)	0	1.0
4	2:240698190	c.722G>A	p.R241H	rs1162389579	1.22	0.140	8	5 (3.5)	0.77
4	2:240698199	c.731A>G	p.H244R (het) p.H244R (hom)	rs201744512	12.4	0.150	128 (43.8) 132 (45.2)	62 (43.4) 67 (46.9)	0.90

P-values were calculated by two-tailed Fisher's Exact test with 2x3 contingency table comparing genotype frequencies and with a 2x2 contingency table comparing carrier frequencies. CP = chronic pancreatitis; AA = amino acid; CADD = combined annotation dependent depletion; REVEL = rare exome variant ensemble learner; het = heterozygous; hom = homozygous.

Table 6.2: Non-synonymous *AQP12B* variants in German CP patients and controls

Exon	Position (hg38)	Nucleotide change	AA change	rs number	CADD	REVEL	CP (n=292) n (%)	Controls (n=143) n (%)	*p*-value
1	2:240682792	c.46A>G	p.T16A (het) p.T16A (hom)	rs4081908	0.83	0.005	111 (38) 174 (59.6)	43 (30.1) 91 (63.6)	0.06
1	2:240682791_2	c.46_47AC>GA	p.T16D	-	-	-	1 (0.3)	1 (0.7)	0.55
1	2:240682770	c.68G>A	p.R23Q (het) p.R23Q (hom)	rs4081909	19.0	0.041	40 (13.7) 1 (0.3)	25 (17.5) 1 (0.7)	0.33
1	2:240682726	c.112G>C	p.A38P	rs769213334	22.0	0.243	1 (0.3)	0	1.0
1	2:240682704	c.134insG	p.V45Gfs*	rs560641543	22.6	-	2 (0.7)	0	1.0
1	2:240682693	c.145G>C	p.A49P	rs771884214	23.9	0.151	0	1 (0.7)	0.33
1	2:240682686	c.152T>G	p.F51C (het) p.F51C (hom)	rs116505219	18.6	0.032	21 (7.2) 2 (0.7)	14 (9.8) 1 (0.7)	0.60
1	2:240682686	c.152T>A	p.F51Y	-	20.6	0.036	1 (0.3)	0	1.0
1	2:240682617	c.221C>G	p.T74S (het) p.T74S (hom)	rs184125501	24.1	0.242	38 (13.0) 4 (1.4)	14 (9.8) 0	0.28
1	2:240682585	c.253T>G	p.V85F (het) p.V85F (hom)	rs182755222	13.7	0.038	22 (7.5) 2 (0.7)	14 (9.8) 1 (0.7)	0.74
1	2:240682491	c.347C>T	p.A116V	rs190367401	2.15	0.040	8 (2.7)	5 (3.5)	0.77
1	2:240682452	c.386C>T	p.T129M (het) p.T129M (hom)	rs74882485	22.4	0.185	130 (44.5) 25 (8.6)	55 (38.5) 12 (8.4)	0.29
1	2:240682449	c.389G>A	p.R130H (het) p.R130H (hom)	rs62187796	0.05	0.049	103 (35.3) 17 (5.8)	48 (33.6) 5 (3.5)	0.52
1	2:240682382	c.455delG	p.S152Tfs*24 (het) p.S152Tfs*24 (hom)	rs201917032	26.8	-	64 (21.9) 6 (2.1)	37 (25.9) 4 (2.8)	0.53
1	2:240682381	c.457T>G	p.S153A	rs754522916	22.2	0.127	1 (0.7)	1 (0.7)	0.55
1	2:240682371	c.467G>C	p.R156P	rs771419765	6.65	0.365	0	1 (0.7)	0.33
1	2:240682330	c.508G>A	p.A170T (het) p.A170T (hom)	rs201329918	9.52	0.140	22 (7.5) 2 (0.7)	13 (9.1) 1 (0.7)	0.81
1	2:240682311	c.527C>T	p.T176I	rs200681385	2.59	0.030	1 (0.3)	0	1.0
1	2:240682242	c.596C>T	p.T199M (het) p.T199M (hom)	rs200045990	8.03	0.035	22 (7.5) 2 (0.7)	13 (9.1) 1 (0.7)	0.81
2	2:240680386	c.690G>T	p.E230D	rs748167930	12.6	0.238	0	1 (0.7)	0.33
3	2:240676702	c.767A>G	p.H256R (het) p.H256R (hom)	rs1807128	11.8	0.041	26 (8.9) 1 (0.3)	14 (9.8) 1 (0.7)	0.70

P-values were calculated by two-tailed Fisher's Exact test with 2x3 contingency table comparing genotype frequencies and with a 2x2 contingency table comparing carrier frequencies. CP = chronic pancreatitis; AA = amino acid; CADD = combined annotation dependent depletion; REVEL = rare exome variant ensemble learner; het = heterozygous; hom = homozygous.

6.2 Chymotrypsin-like protease

Mutations in a large panel of digestive enzymes have already been identified as risk factors for chronic pancreatitis. Based on an extended candidate gene approach, we investigated the chymotrypsin-like protease (*CTRL*) as a potential novel risk gene. Our WES data revealed 15 variants located in the coding sequence of *CTRL*, seven of whom were found in controls only (p.G37E, p.G61S, p.S112N, p.V177L, p.G215R, p.L218Rfs*33, p.G230S).

For the prediction of deleterious variants, we annotated REVEL predictions and burden testing of rare variants in candidate causal genes against the gnomAD database. The burden test approach compares the number of individuals that carry rare, protein-altering variants in each gene between case subjects and the control gnomAD database (Guo et al. 2018). These analyses were performed by Andreas Schmidt and results are listed in Table 6.3 presenting three variants in *CTRL* predicted deleterious (p.C201Y, p.S208F and p.C220G) with a REVEL score ≥0.5 found in our patient cohort (p = 0.143; OR = 2.03).

Table 6.3: Deleterious variant prediction in *CTRL* using REVEL

Gene	**REVEL score**	**case count**			**control count (gnomAD)**		
		het	hom	total	het	hom	total
CTRL	≥0.5	4	0	4	106	0	106
	p-value	OR	Position (hg19)		Nucleotide change	AA change	
	0.143	2.03	16:67963974 16:67964098 16:67964119		A>C G>A C>T	p.C220G p.S208F p.C201Y	

REVEL = rare exome variant ensemble learner; het = heterozygous; hom = homozygous; OR = odds ratio; AA = amino acid, case count = 1,059; control count = 56,885; calculations were performed by Andreas Schmidt.

Despite the lack of statistical significance, we functionally examined the *CTRL* mutations because all prediction tools are of limited value. After confirmation of the mutations by Sanger sequencing, we analyzed secretion and activity in HEK293T cells. Additionally, we included two common missense mutations found in the patient cohort (p.T150I and p.H173R) which in gnomAD showed an allele count >1000 (allele frequency >0.01). Figure 6.1 gives a schematic illustration of the CTRL mutations and their position within the amino acid sequence.

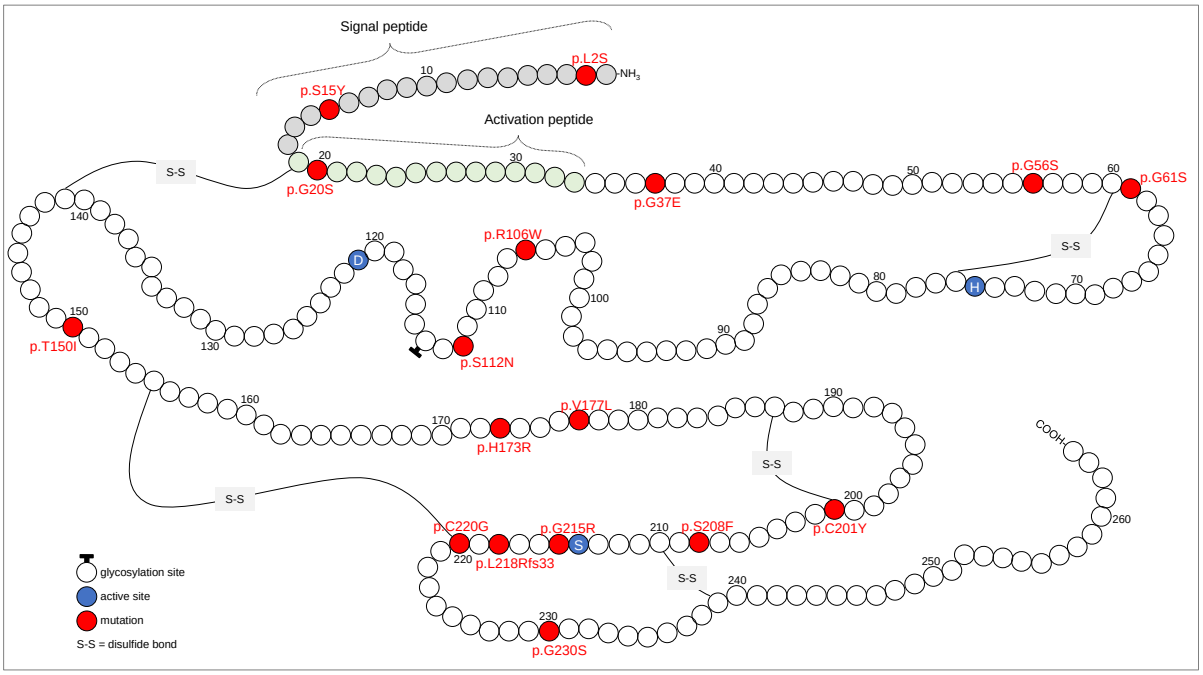


Figure 6.1: Schematic representation of CTRL mutations within the protein sequence.
Illustration of the CTRL polypeptide chain harboring single amino acid changes investigated in this project. Grey = signal peptide; green = activation peptide; blue = active site amino acids; red = investigated mutations; S-S = disulfide bond.

6.2.1 Western blot analysis shows reduced secretion of C-terminal mutants

We conducted Western blot analyses to investigate whether *CTRL* mutations lead to defective secretion. We overexpressed CTRL wildtype and mutants in HEK293T cells and analyzed both supernatant and total lysate (Figure 6.2). Please note that Western blots are shown in black and white for better visibility of faint bands. The original figures can be found in the supplementary data (Section 9.4.1; Supplementary Figure 9.1).

Since CTRL is a secretory enzyme, we expected that the protein would be found not only in the lysate but primarily in the supernatant. We split the mutations into three sets for analysis. Set 1 (Figure 6.2 **A**) includes mutations found in CP patients, set 2 (Figure 6.2 **B**) composes of both patient mutations and control mutations (p.G37E and p.G230S), and set 3 (Figure 6.2 **C**) contains mutations found in the control cohort only. Each set of mutations was investigated compared to the CTRL wildtype and included the pcDNA3.1+/C-(K)-DYK empty vector as negative control.

We detected almost all variants and the CTRL wildtype in the lysate, indicating successful protein synthesis. Solely the frameshift mutation p.L218Rfs*33 (set 3; Figure 6.2 **C**) did not show a band. As the frameshift not only results in a severely altered amino acid sequence, but also in a premature stop codon, this leads either to direct intracellular degradation of the protein or to no synthesis at all through nonsense-mediated decay.

The situation is different when looking at the supernatant, which represent the secretory properties. Several mutations led to reduced secretion since no bands or weak bands were

detectable: p.C201Y and p.C220G (Figure 6.2 **A**), p.G230S (Figure 6.2 **B**), as well as p.G215R and p.L218Rfs*33 (Figure 6.2 **C**). In p.C220G, we detected a weak band in one biological replicate indicating a reduced secretion. Strikingly, all mutations with diminished secretion are located at the C-terminal end.

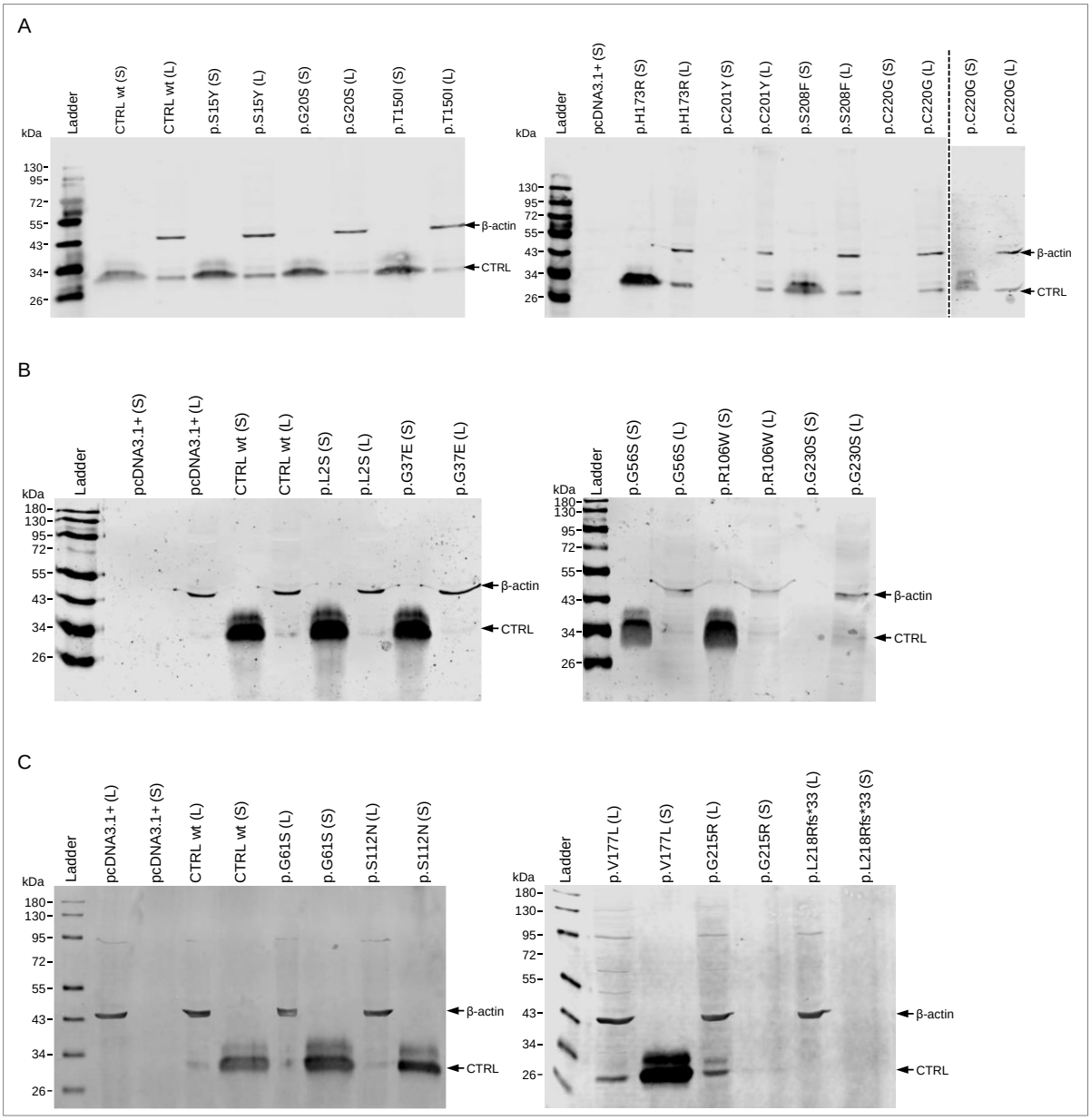


Figure 6.2: Western blot of HEK293T supernatant (S) and lysate (L) after CTRL overexpression
Antibody against Flag-Tag was used to detect overexpressed CTRL (28 kDa). The housekeeper gene β-actin (42 kDa) was detected in the total lysate samples as internal loading control. The supernatant was precipitated from 400 µl before loading onto the gel; 40 µg total lysate was used. The figures are representative for a total of four biological replicates. (**A**) Mutation set 1 found in CP patients by WES. Some experiments were conducted by Lea Neppl for her Bachelor thesis. In some replicates, p.C220G shows a reduced secretion. (**B**) Mutation set 2 found in CP patients (p.L2S, p.G56S, p.R106W) as well as controls (p.G37E, p.G230S). Anja Dollinger conducted some Western blots during her research internship. (**C**) Mutation set 3 found in the control cohort. Some experiments were conducted by technical assistant Beate Rauscher. CTRL = chymotrypsin-like protease; pcDNA3.1+/C-(K)-DYK = empty vector plasmid

6.2.2 CTRL variants show reduced activity or loss of function

To test the enzymatic properties of mutant CTRL, we collected supernatant and total lysate from HEK293T cells and activated CTRL with 125 ng porcine trypsin for 2 h. Since the enzyme cleaves polypeptides at phenylalanine and tyrosine residues (Szabó and Sahin-Tóth 2012), we used the synthetic substrates N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (AAPF) and N-succinyl-Ala-Ala-Pro-Tyr-p-nitroanilide (AAPY). When cleaving the amino acid residue, p-nitroaniline (p-NA) is released and is measurable at 405 nm as yellow color. The color intensity is proportional to the enzyme activity. However, AAPY showed less stable signals and lower intensity (supplementary section 9.4.2, supplementary Figure 9.2). Figure 6.3 shows the relative CTRL activity in supernatant (**A**) and lysate (**B**) on substrate AAPF. All measurements were set in relation to the CTRL wildtype activity (100 %). Overall, we found a similar activity pattern in supernatant and lysate. However, as CTRL is secreted, we expected less protein intracellularly. Consequently, the color intensity in the lysate samples was very low compared to the supernatant, which resulted in a higher standard deviation.

Five mutations showed similar activity to the wildtype in both supernatant and lysate: p.S15Y (112.8 % and 140.9 %, respectively), p.R106W (90.2 % and 83.1 %), p.S112N (91.0 % and 119.9 %), p.T150I (96.1 % and 89.7 %), and p.H173R (117.1 % and 107.6 %). Please note that the standard deviation of p.S15Y in the lysate is very high and consequently it cannot be assumed a gain-of-function mutation despite an activity of 140.9 %.

In contrast, other mutations showed an almost complete activity loss in the supernatant: p.G37E (0.3 %), p.C201Y (2.2 %), p.S208Y (1.1 %), p.G215R (0.7 %), p.L218Rfs*33 (0 %), and p.G230G (1.4 %). While most of them are in line with the Western blot results regarding secretion, p.G37E and p.S208Y showed a complete or nearly complete (<10 %) loss of activity despite normal secretion. Looking at the lysate, the same mutations led to an activity loss: p.G37E (7.3 %), p.C201Y (1.9 %), p.S208Y (5.9 %), p.L218Rfs*33 (0 %), and p.G230S (4.2 %). Only the activity of p.G215R differed in supernatant and lysate (0.7 % and 29.8 %, respectively).

Four variants show a reduced activity between 25-75 % in both supernatant and lysate: p.G20S (65.3 % and 50.5 %), p.G56S (51.7 % and 34.4 %), p.G61S (35.0 % and 32.4 %), and p.C220G (24.8 % and 39.9 %). Differences in the supernatant and lysate were observed in p.V177L, in which the activity in the supernatant was 61.1 %, while the activity in the lysate was 102.1 %, comparable to the wildtype. However, due to a relatively high standard deviation in the lysate samples, these differences are questionable. The same applies to p.L2S (supernatant: 88.2 %; lysate: 57.1 %).

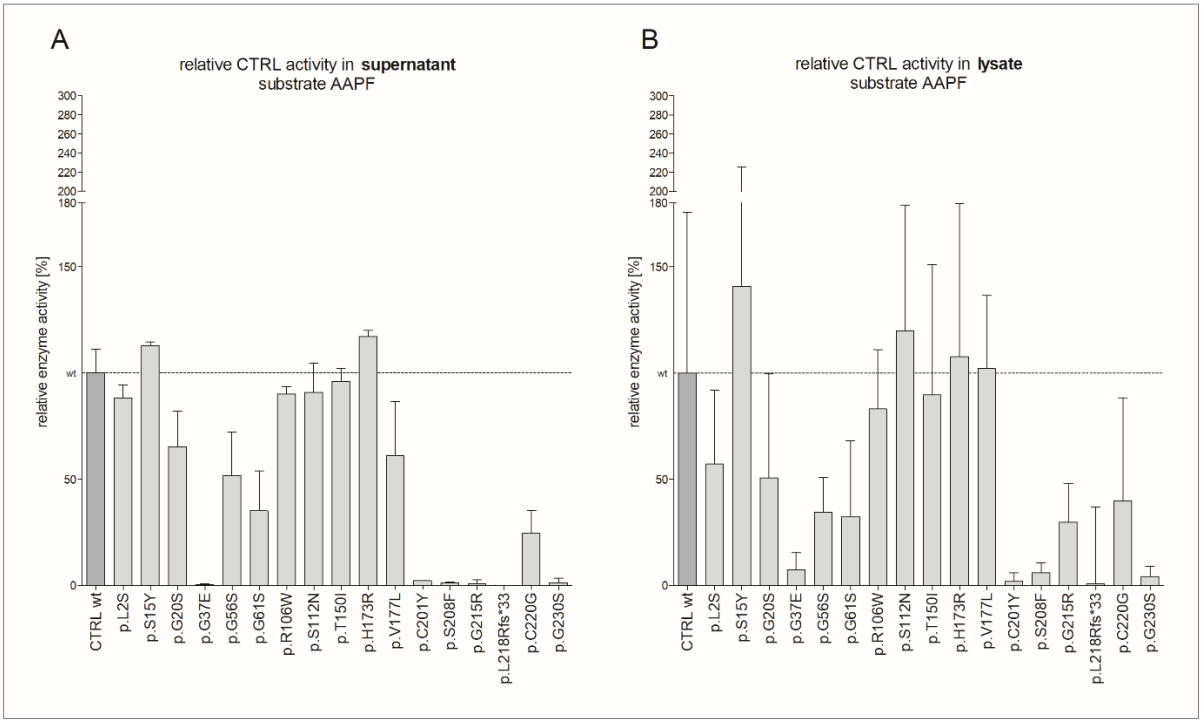


Figure 6.3: CTRL activity in supernatant and cell lysate samples using AAPF
Supernatant and total lysate were prepared from HEK293T cells after overexpression of CTRL wildtype (wt) or mutants. We activated the enzyme with 125 ng porcine trypsin for 2 h. We measured activity in technical triplicates on four biological replicates. Enzyme activity 8 min after addition of AAPF is shown. All measurements were set in relation to CTRL wt activity (100 %; dark grey bar). Relative CTRL activity in supernatant (**A**) and in lysate (**B**). AAPF = N-Suc-Ala-Ala-Pro-Phe-p-NA.

6.2.3 ER stress as possible consequence of CTRL mutations

As some CTRL mutants displayed reduced secretion, we hypothesized that intracellular accumulation of misfolded proteins might induce ER stress and consequently acinar cell damage and pancreatitis. Especially in mutations that showed some activity within the cell (lysate) but were not secreted.

Whenever ER homeostasis is disturbed for example by viral infections, environmental toxins or even mutant proteins, this results in accumulation of misfolded or unfolded proteins, so called ER stress (Oslowski and Urano 2011). Figure 6.4 shows the ER stress hypothesis schematically. The adaptive response to ER stress is the unfolded protein response (UPR) initiated by three master regulators: protein kinase R-like endoplasmic reticulum kinase (PERK), activating transcription factor 6 (ATF6), and inositol-requiring kinase 1 (IRE1α). In an unstressed state, the ER chaperone immunoglobulin binding protein (BiP) binds to the master regulators and keeps them inactive. In cell stress, BiP is relocated to the unfolded protein to facilitate folding. The master regulators are then active and induce expression of UPR or apoptotic genes. PERK phosphorylates eukaryotic initiation factor 2 (eIF2a), which inhibits further protein translation to protect the cell from apoptosis. In turn, eIF2a phosphorylates the activating transcription factors 4 (ATF4) to regulate several UPR target genes. ATF6 is cleaved in the Golgi apparatus and translocates to the nucleus, where it

activates UPR genes such as X-box binding protein 1 (XBP1). Through dimerization, IRE1α autophosphorylates and leads to splicing of XBP1, a leucin zipper transcription factor, to upregulate UPR target genes and trigger ER-associated protein degradation (ERAD). (Oslowski and Urano 2011)

To investigate the effects of mutated CTRL on ER stress, we analyzed *BiP* mRNA expression by qPCR as well as *XBP1* splicing. For this, we isolated RNA from transfected HEK293T cells and transcribed it into cDNA. As positive controls, we treated HEK293T cells overexpressing CTRL wildtype with either tunicamycin (TM), thapsigargin (TG), or dithiothreitol (DTT) 6 h prior to RNA isolation to provoke ER stress.

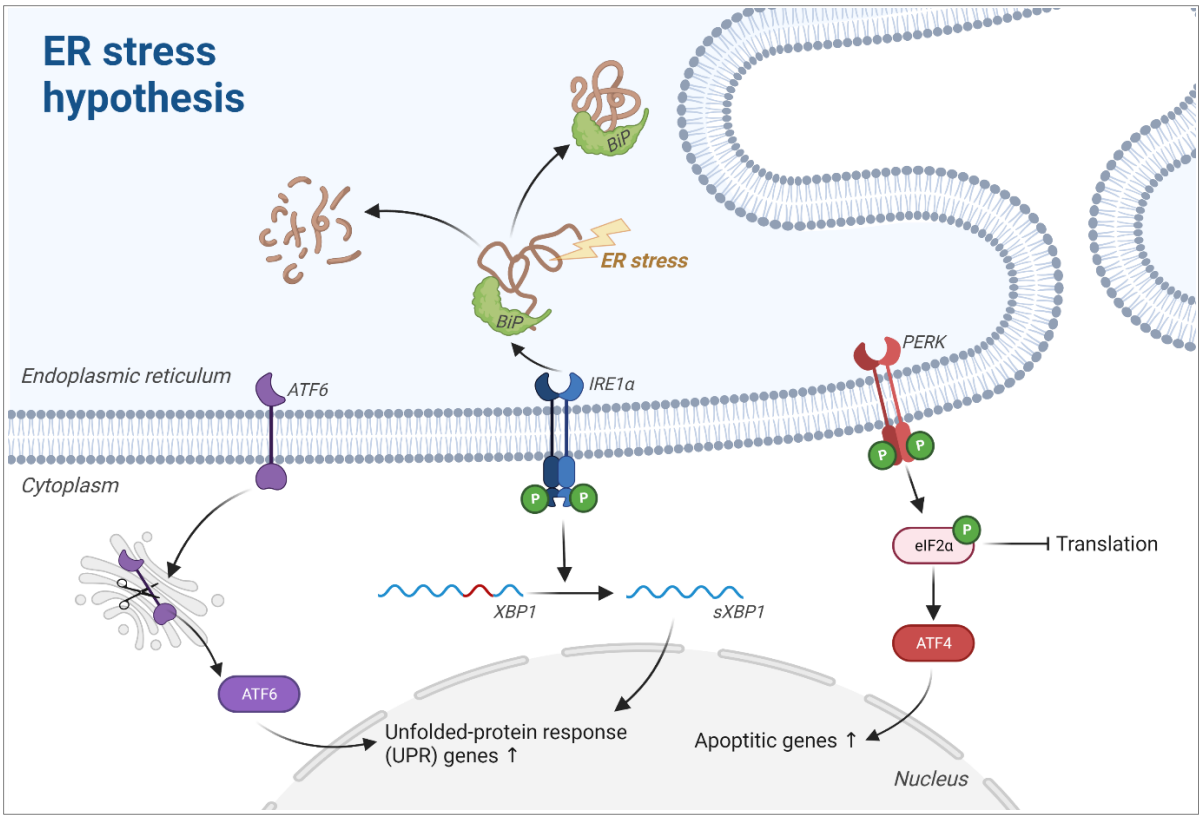


Figure 6.4: Schematic representation of the ER stress hypothesis
Mutation-induced misfolding of proteins lead to ER stress. Disturbed ER homeostasis results in accumulation of misfolded or unfolded proteins. The adaptive response is the unfolded protein response (UPR) initiated by three master regulators: ATF6 (purple), PERK (red) and IRE1 (blue). In an unstressed state, the ER chaperone BiP (green) binds to the master regulators to keep them inactive. Upon ER-stress, BiP relocalizes to the unfolded protein to facilitate folding. The master regulators are then active, leading to increased expression of UPR or apoptotic genes. PERK phosphorylates eIF2a, which inhibits further protein translation, and phosphorylates ATF4 to regulate several UPR target genes. ATF6 is cleaved in the Golgi apparatus and translocates to the nucleus to activate UPR genes such as XBP1. XBP1 is spliced by autophosphorylated IRE1α to upregulate UPR target genes and to provoke ERAD. ER = endoplasmic reticulum; ATF6 = activating transcription factor 6; IRE1α = inositol-requiring kinase 1; PERK = protein kinase R-like endoplasmic reticulum kinase; BiP = immunoglobulin binding protein; eIF2α = eukaryotic initiation factor 2; ATF4 = activating transcription factors 4; XBP1 = X-box binding protein 1; sXBP1 = spliced XBP1; P = phosphate; ERAD = ER-associated protein degradation. Figure created with *BioRender.com*.

Figure 6.5 shows the results of our ER stress analyses. *BiP* mRNA expression qPCRs were performed in technical triplicates on four biological replicates (**A**). The data shows that none

of the mutations increase the expression of *BiP* mRNA compared to the wildtype. Yet, the positive controls treated with TG and DTT show an increased expression with 3.7-fold and 4.0-fold increase, respectively. This indicates a successful induction of ER stress. In *XBP1* splicing PCR (**B**), spliced *XBP1* (416 bp) was present in all samples, even in untransfected cells and empty vector transfected cells. This might be due to a general stressed state of the cells during culturing, transfection, and harvest. No increased splicing was visible in most of the CTRL mutants compared to the wildtype. Though p.G230S shows a slight band in the spliced *XBP1*, this was only in the third biological replicate (n3). The positive controls show a slightly thicker band at 416 bp. In conclusion, the *BiP* mRNA expression data (**A**) are comparable to the results from the *XBP1* splicing PCR results (**B**), and ER stress is likely not a consequence of the investigated *CTRL* variants.

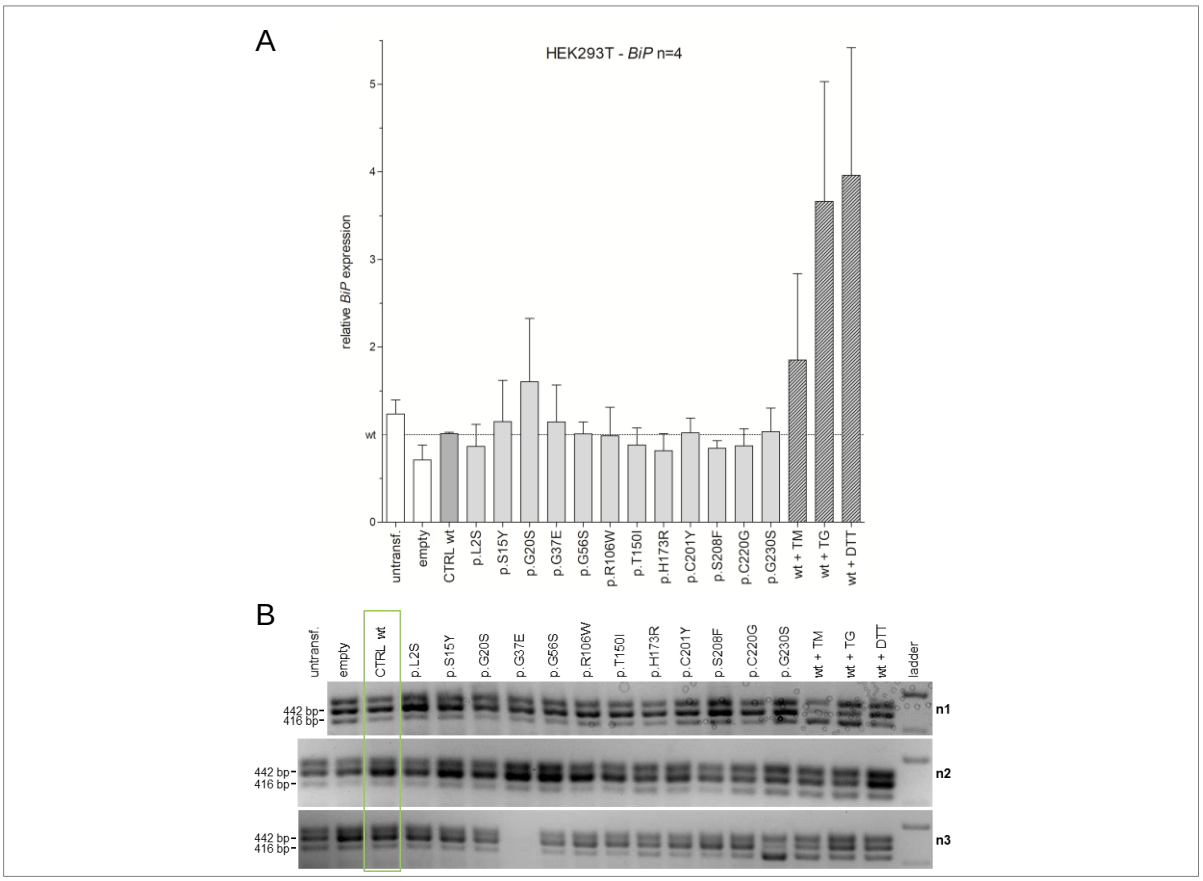


Figure 6.5: ER stress analysis results of CTRL wildtype and mutants
Three positive controls were included: CTRL wt treated with either tunicamycin (TM), thapsigargin (TG), or dithiothreitol (DTT) 6 h prior to RNA isolation. Untransfected cell cDNA (untransf.) and pcDNA3.1+/C-(K)-DYK transfected cells (empty) served as negative controls. (**A**) TaqMan assays on *BiP* mRNA expression in technical triplicates of biological n = 4. β-Actin and GAPDH served as internal controls. The horizontal line indicates baseline relative mRNA expression in CTRL wt (=1; dark grey bar). Bars in shaded grey show relative *BiP* expression in positive controls. (**B**) *XBP1* splicing PCR was conducted with three biological replicates (n1 – n3). Unspliced *XBP1* shows a PCR product size of 442 bp, while spliced *XBP1* has a PCR product size of 416 bp indicating ER stress. The third band represents a hybrid product of one spliced and one unspliced strand. All band were visually compared to CTRL wt (green frame). BIP = immunoglobulin binding protein; XBP1 = X-box binding protein 1

6.2.4 Summary of functional analyses of CTRL

The results of our functional analyses are summarized in Table 6.4 and Figure 6.6. Functionality was unchanged in 7/17 mutations (p.L2S, p.S15Y, p.R106W, p.S112N, p.T150I, p.H173R, and p.V177L). Five mutations showed normal secretion but reduced (p.G20S, p.G56S, p.G61S) or abolished activity (p.G37E, p.S208F). Notably, all secretion-reduced mutations are located at the C-terminal end and resulted in lost (p.C201Y, p.G215R, p.L218Rfs*33, p.G230S) or reduced activity (p.C220G), even within the cell. Mutation frequencies in patients and controls were tested by two-tailed Fisher's Exact test (Table 6.4).

Table 6.4: Fisher's Exact test results on CTRL mutations

Consequence	Mutation	CP (n=1,059) (%)	Controls (n=2,099) (%)	*p*-value	OR
Reduced / no secretion	p.C201Y p.G215R p.L218Rfs*33 p. C220G p.G230S	3 (0.3)	6 (0.3)	1.000	0.99
Activity <10 %	p.C201Y p.S208F p.G215R, p.L218Rfs*33 p.G230S	2 (0.2)	6 (0.3)	0.726	0.66
Activity 25-75 %	p.G20S p.G56S p.G61S p.V177L p.C220G	4 (0.4)	2 (0.1)	0.102	3.97

Mutation p.G37E was excluded in statistical analysis due to a separate control cohort. OR = odds ratio.

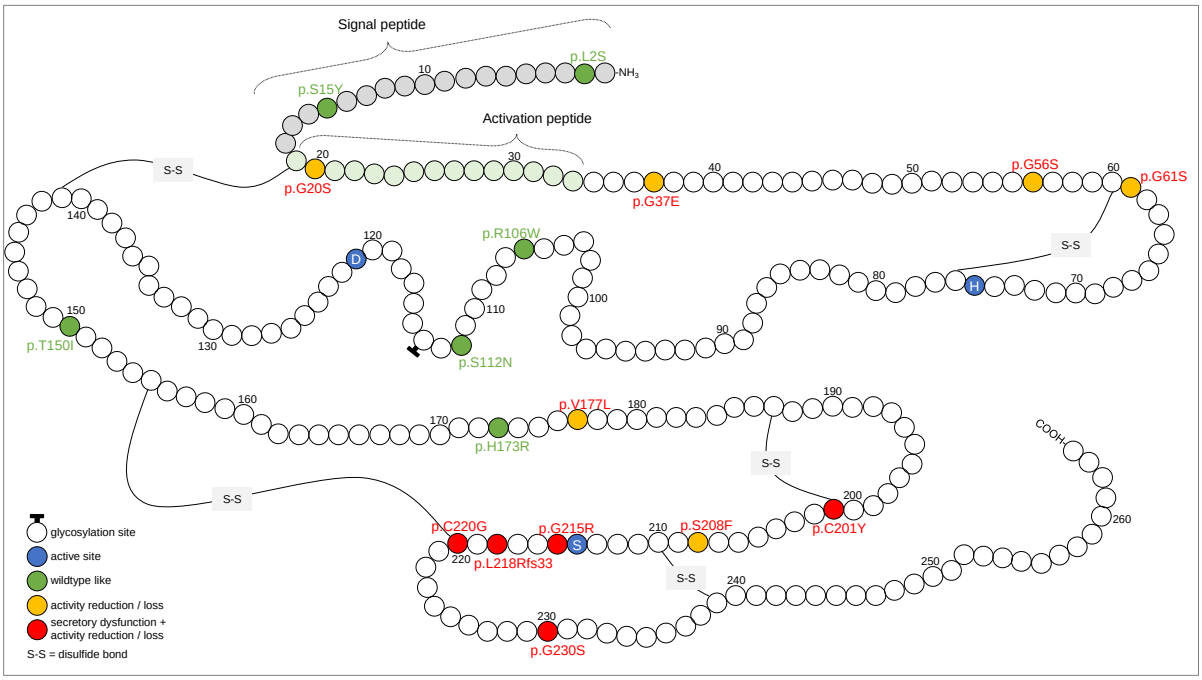


Figure 6.6: Summary of functional analyses of CTRL mutations
Grey = signal peptide; light green = activation peptide; blue = active site; S-S = disulfide bond; green = mutations with unaltered function; yellow = mutations with normal secretion but activity reduction / loss; red = mutations with secretory dysfunction and activity reduction / loss.

6.3 Transcription Factors

While digestive enzymes and their proteolytic cascade have long been the primary focus of pancreatitis research, a rather novel approach is the potential pathogenic role of transcription factors involved in pancreatic development and maintenance. We investigated NR5A2, RBPJL and HNF1a as factors involved in different potentially pathological pathways in the pancreas. Figure 6.7 depicts their individual domain structure and indicates the approximate location of mutations (arrows) found in patients and controls.

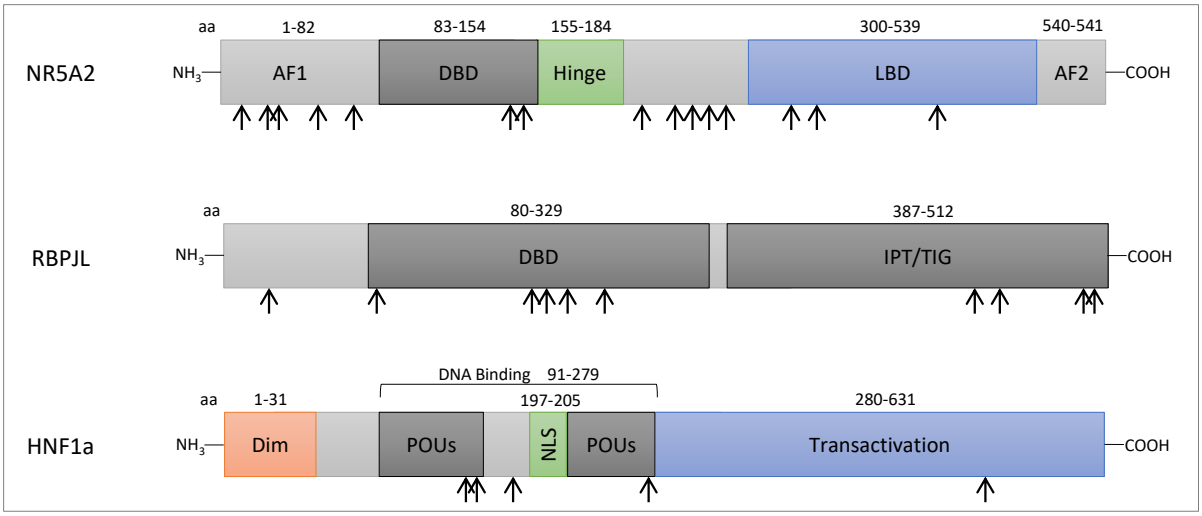


Figure 6.7: Schematic representation of the domain structures of NR5A2, RBPJL, and HNF1a
Arrows show the approximate location of mutations found in our WES cohort, which we included in functional analyses. For NR5A2, 15 mutations were investigated. For RBPJL, ten mutations were found and five mutations in HNF1a.
AF = transcriptional activation function; DBD = DNA binding domain; LBD = ligand binding domain; IPT/TIG = immunoglobulin like fold / involved in DNA binding; Dim = Dimerization; NLS = nuclear localization sequence; aa = amino acids; Hinge region = provides flexibility; POUs = well-conserved bipartite DNA binding domain

Starting this project, we obtained results from bioinformatics calculations of our WES data. When annotating a REVEL score ≥0.5 for *NR5A2*, *RBPJL* and *HNF1a*, several missense mutations were predicted deleterious, however, analysis indicates a significance of *NR5A2* only (p = 0.027, OR = 3.1) (Table 6.5).

Even though REVEL is a highly functional tool for predicting deleterious variants, it still shows limitations in the evaluation of our data set. First, REVEL only detects missense mutations. Several nonsense mutations in our patient cohort were not taken into account. Furthermore, the specificity and sensitivity of REVEL at a threshold of 0.5 – while still being distinguished – corresponds to 0.754 and 0.891, respectively (Ioannidis et al. 2016), leaving room for improvement. We also have to consider that gnomAD as reference database is not an ideal choice. While our patient cohort consists of predominant German individuals, the gnomAD cohort includes a completely different genomic composition of Europeans (non-finish European).

Table 6.5: Deleterious variant prediction in *NR5A2*, *RBPJL* and *HNF1a* using REVEL

Gene	**REVEL score**	**case count**			**control count**		
		het	hom	total	het	hom	total
NR5A2	≥0.5	5	0	5	87	0	87
	p-value	OR	Position (hg19)		Nucleotide change	AA change	
	0.027	3.10	1:200008803 1:200008924 1:200014674 1:200014697 1:200017542		C>T G>T G>A G>A G>C	p.R28* c.202+1 p.R142H p.G150R p.D236H	
Gene	REVEL score	case count			control count		
		het	hom	total	het	hom	total
RBPJL	≥0.5	5	0	5	193	0	193
	p-value	OR	Position (hg19)		Nucleotide change	AA change	
	0.296	1.39	20:43938325 20:43941014 20:43942132 20:43942153 20:43942710		G>A C>T T>C G>C C>T	p.E84K p.Q200* p.V215A p.R222P p.R265*	
Gene	REVEL score	case count			control count		
		het	hom	total	het	hom	total
HNF1a	≥0.5	18	0	18	845	0	845
	p-value	OR	Position (hg19)		Nucleotide change	AA change	
	0.318	1.15	12:121416663 12:121416828 12:121426663 12:121426790 12:121426811 12:121431343 12:121432086 12:121432115 12:121437110 12:121437322 12:121437391 12:121437401 12:121437410		G>A T>A G>A G>A C>T G>A G>A G>T A>G G>A C>G C>T G>A	p.G31D p.L86H p.M118I p.A161T p.R168C p.G183R p.R278Q p.G288W p.H514R p.G554R p.H577D p.P580L p.R583Q	

REVEL = rare exome variant ensemble learner; het = heterozygous; hom = homozygous; OR = odds ratio; AA = amino acid.

Therefore, we decided to analyze NR5A2, RBPJL and HNF1a functionally in regard to transcriptional activity in reporter-gene assays and endogenous target gene expression.

6.3.1 Verification of protein expression in Western blot

We overexpressed the wildtype (wt) transcription factors NR5A2, RBPJL and HNF1a in HEK293T cells and isolated total lysate for Western blot analysis (Figure 6.8). We loaded 20 µg lysate onto the gel and the proteins were detected with the monoclonal antibody against Flag-tag (**A**, red bands). A clear band is visible for HNF1a wt (67 kDa) and RBPJL wt (57 kDa). A faint second band for HNF1a is visible at ~150 kDa and the band for NR5A2 wt (61 kDa) displays as very weak signal. The housekeeper gene β-actin (42 kDa; green bands) served as internal loading control. To intensify the NR5A2 detection, a specific polyclonal

antibody directly against human NR5A2 was used to successfully detect NR5A2 wt (**B**). Here, we used 20 µg lysate as well.

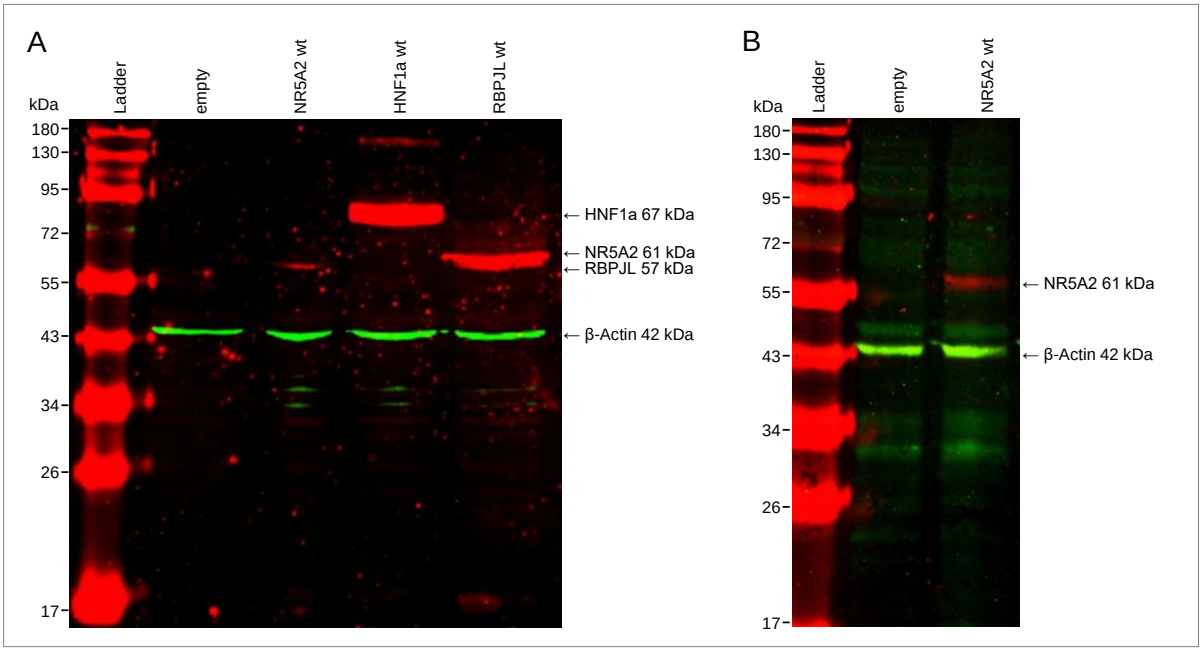


Figure 6.8: Western blot of wildtypes of NR5A2, HNF1A and RBPJL
We loaded 20 µg total lysate from HEK293T cells onto the gel and used pcDNA3.1+/C-(K)-DYK vector plasmid as negative control (empty). The housekeeper gene β-actin (42 kDa; green bands) served as internal loading control. (**A**) Protein detection using a monoclonal antibody against Flag-Tag (red band). (**B**) NR5A2 detection only: a specific polyclonal antibody against human NR5A2 was used (61 kDa, red band).
NR5A2 = nuclear receptor subfamily 5 group A member 2; HNF1a = hepatocyte nuclear factor 1 homeobox A; RBPJL = recombination signal binding protein for immunoglobulin kappa J region like

6.3.2 Molecular cloning of potential target gene promoter sequences

For the reporter gene assays, we cloned several promoter regions of potential target genes into a basic luciferase vector (pGL4.22). With the help of literature and the transcription factor binding site (TFBS) prediction tool JASPAR2020 (Castro-Mondragon et al. 2022), we decided on the length and region of each promoter to be cloned. Figure 6.9 shows the molecular cloning strategy representative of the *CPA2* promoter region.

First, we amplified the selected promoter region and introduced restriction enzyme cutting sites using specifically designed primers for mutagenesis. We digested both the amplified PCR product as well as the basic pGL4.22 vector with restriction enzymes. Through the sticky ends, we ligated the promoter region insert into the vector. The ligation samples were transformed into chemo-competent *E. coli* cells. An ampicillin resistance allowed selective growth of the bacteria to amplify the new vector plasmid. With this, we successfully cloned 14 different promoter regions with a length ranging from 515 bp (*CPA2* promoter) to 2,882 bp (*CYP17* promoter) as listed in Table 5.1. We verified the sequences with Sanger Sequencing at Eurofins Genomics.

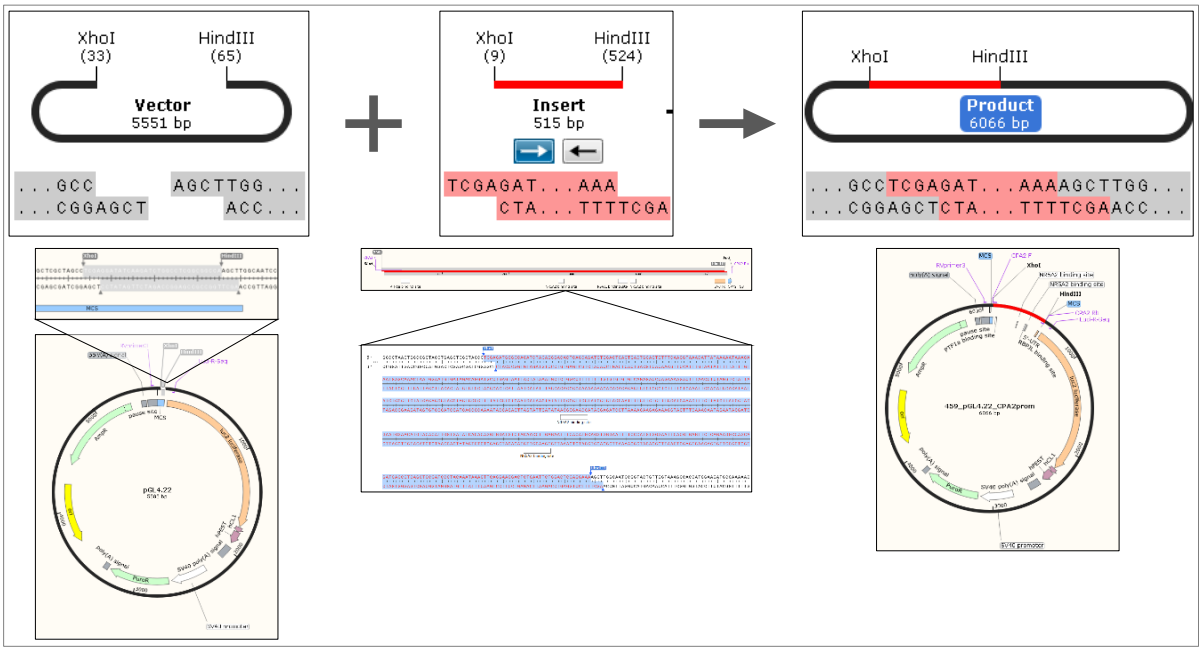


Figure 6.9: Molecular cloning of *CPA2* promoter region into pGL4.22
Scheme of molecular cloning using the *CPA2* promoter region as an example (-501/+20bp). Both, the empty pGL4.22 luciferase vector (left) and the amplified *CPA2* promoter region as insert (middle; red) were digested with *XhoI* and *HindIII* to create sticky ends. The final vector pGL4.22_CPA2prom (right) contains an inserted *CPA2* promoter region of 515 bp length located upstream of the coding region of the *luc2* luciferase gene.
Figure was created from SnapGene version 4.3.11.

Table 6.6: Promoter regions with potential TFBS successfully cloned into basic pGL4.22

Gene	Name	Promoter region	TFBS for	Cloned length [bp]
PRSS2	Serine protease 2	-1575/+18bp	RBPJL	1587
SPINK1	Pancreatic secretory trypsin inhibitor	-1154/+215bp	HNF1a/PTF1	1362
CEL	Carboxyl ester lipase	-551/+22bp	NR5A2	567
CPA2	Carboxypeptidase A2	-501/+20bp	NR5A2	515
CPA1	Carboxypeptidase A1	-2051/+27bp	NR5A2	2080
CELA3A	Chymotrypsin like elastase family member 3A	-1965/+23bp	NR5A2	1975
REG1A	Regenerating family member 1 alpha	-1136/+384bp	HNF1a; RBPJL	1514
GP2	Glycoprotein 2	-1402/+1084bp	RBPJL/PTF1	2480
CYP11B1	Steroid-11b-hydroxylase	-1061/+29bp	NR5A2	1084
CYP17	Steroid-17a-hydroxylase	-2856/+31bp	NR5A2	2882
INS	Insulin	-2314/+366bp	HNF1a	2687
IAPP	Islet amyloid polypeptidase	-2058/+494bp	HNF1a	2556
MTTP	Microsomal triglyceride transfer protein	-1613/+201bp	HNF1a	1818
CALR	Calreticulin	-1802/+93bp	NR5A2	1895

Exemplified explanation on promoter sequence -501/+20 bp = 501 base pairs in 5'-direction of 5'-UTR and 20 base pairs into 5'-UTR.

6.3.3 Luciferase Assay establishment with all promoters

To study the transcriptional activity of a mutant transcription factor, we overexpressed the factors in HEK293 cells either as a single transfection (only NR5A2 or only HNF1a) or in

combination with the PTF1 complex (PTF1a + RBPJL or PTF1a + RBPJL + E47). As RBPJL is part of the PTF1 complex, it was solely overexpressed with PTF1a with or without E47. Each transcription factor combination was then co-transfected with the previously cloned reporter gene plasmids. As an internal standard, a second luciferase vector (pRL-Ubi) was included. Results are shown in Figure 6.10. For each transcription factor, the best combinations were then selected for further luciferase assays. This resulted in a selection of combinations of transcription factors with or without complex together with a panel of promoter-luciferase vectors.

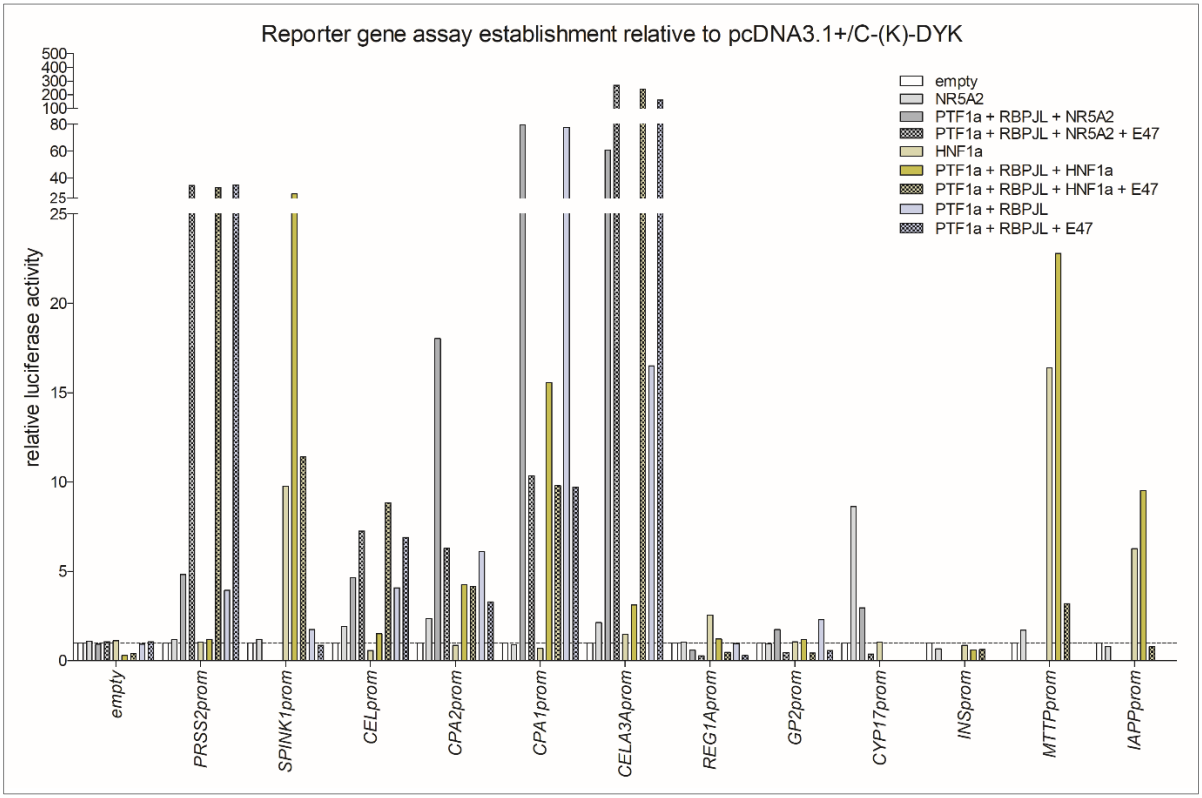


Figure 6.10: Luciferase assay to establish suitable transcription factor-promoter combinations
Luciferase activity was measured after transcription factors were overexpressed in HEK293 cells either as single transcription factor or in combination with the PTF1 complex (PTF1a + RBPJL or PTF1a + RBPJL + E47). The pcDNA3.1+/C-(K)-DYK vector (empty; white bars) was used as control. A co-transfection was performed with a reporter gene vector containing a target gene promoter (prom) region. Twelve promoters and a pGL4.22 control vector (empty; left block on x-axis) were investigated in a minimum of biological n = 2. Vectors of *CYP11B1* promoter and *CALR* promoter showed inconsistent values and therefore were excluded from further experiments.

Looking at the promoter of *PRSS2* (*PRSS2prom*), luciferase activity is increased in all combinations with E47, regardless of whether NR5A2 or HNF1a is present (dotted bars). Therefore, E47 seems to be the driving factor. In the *SPINK1prom* combinations, the highest luciferase activity of ~28-fold was achieved by the combination HNF1a in complex (PTF1a + RBPJL + HNF1a; dark yellow bar).

Among the digestive enzyme promoters (*CELprom*, *CPA2prom*, *CPA1prom* and *CELA3Aprom*), *CELprom* was rather induced by E47. The *CPA2prom* transfections led to the highest luciferase activity (18-fold) when co-transfected with NR5A2 in complex

(PTF1a + RBPJL + NR5A2; dark grey bar). As the activity with the complex (PTF1a + RBPJL; blue bar) led to a ~6-fold increase only, it is likely that NR5A2 is the driving factor here. Opposite results were found for *CPA1prom*, where the complex combination PTF1a + RBPJL led to an approximately 80-fold and 78-fold increase both with or without NR5A2. Consequently, NR5A2 is not responsible for enhanced gene expression. For *CELA3Aprom*, increased luciferase activity was observed in the presence of E47 (dotted bars). However, without E47, there were differences in the activity dependent on NR5A2. NR5A2 in complex led to a ~61-fold increase (dark grey bar), unlike the complex transfection without NR5A2 (~17-fold; light blue bar).

Representative for zymogen granule genes, we investigated the promoters for *REG1A* and *GP2*. Both, however, do not seem to be primarily targeted by any transcription factor combinations tested. *CYP17prom* as described by literature as target gene promoter for NR5A2 indeed results in a 9-fold increased activity when NR5A2 is transfected solely.

As a last panel, promoters of endocrine genes (*INSprom*, *MTTPprom*, *IAPPprom*) were included as potential target genes for HNF1a. We detected the greatest activity increase on the *MTTPprom*. When HNF1a was single transfected, the luciferase activity was ~16-fold, and in complex without E47 it increased to ~23-fold activity.

Based on our results, we completely excluded E47 from further experiments. To investigate NR5A2, we selected the combination with PTF1a and RBPJL for the *CPA2* and *CELA3A* promoters. The analysis of the *CYP17* promoter was conducted with solely NR5A2 without complex. The *CPA1prom* vector showed to be a target promoter for PTF1a and RBPJL. Furthermore, for the transcriptional activity of HNF1a, we selected the promoter vector *SPINK1prom* in combination with PTF1a, RBPJL and HNF1a and the *MTTPprom* for single transfection with HNF1a.

6.3.4 Reporter gene assay on transcriptional activity

To investigate the transcriptional activity of NR5A2, RBPJL and HNF1a, we performed reporter gene assays using the dual-luciferase system. We transfected HEK293 cells with the selected combinations of expression vectors and a firefly luciferase vector containing an upstream promoter region as previously established. A second luciferase vector containing the coding region for renilla luciferase (pRL_Ubi) allowed standardization of the measured values in the subsequent assays.

6.3.4.1 NR5A2 mutations in DNA binding site reduce transcriptional activity

For NR5A2, we investigated 15 variants found in our WES data of whom five were found in the control cohort only (p.G24R, p.R54Q, p.A195T, p.R267Q, p.G343E). The results of the

luciferase assays are shown in Figure 6.11. All measurements were referenced to the NR5A2 wildtype (wt = 1) and the empty vector pcDNA3.1+/C-(K)-DYK was used as negative control. For the experiments on two promoters, NR5A2 was co-transfected with PTF1a and RBPJL (**A**, *CPA2prom* and *CELA3Aprom*). For those, an additional negative control was included where PTF1a and RBPJL were transfected without NR5A2. This allowed us to exclude effects by PTF1a and RBPJL alone. For *CPA2prom*, the two mutations p.R142H and p.G150R showed a relative reduction in luciferase activity to 0.4-fold and 0.3-fold, respectively. This shows an activity at the level of PTF1a+RBPJL transfection and consequently suggests a loss-of-function in NR5A2. Both mutations are located in the DNA binding domain of the protein. In *CELA3Aprom*, solely p.G150R shows a reduced luciferase activity to 0.5-fold.

When NR5A2 was single transfected (**B**), its transcriptional activity on *CYP17prom* shows no changes in most of the mutations compared to the wildtype. The mutations in the DNA binding domain, p.R142H and p.G150R, however, resulted in an almost complete activity loss down to the level of the empty control (both 0.1-fold).

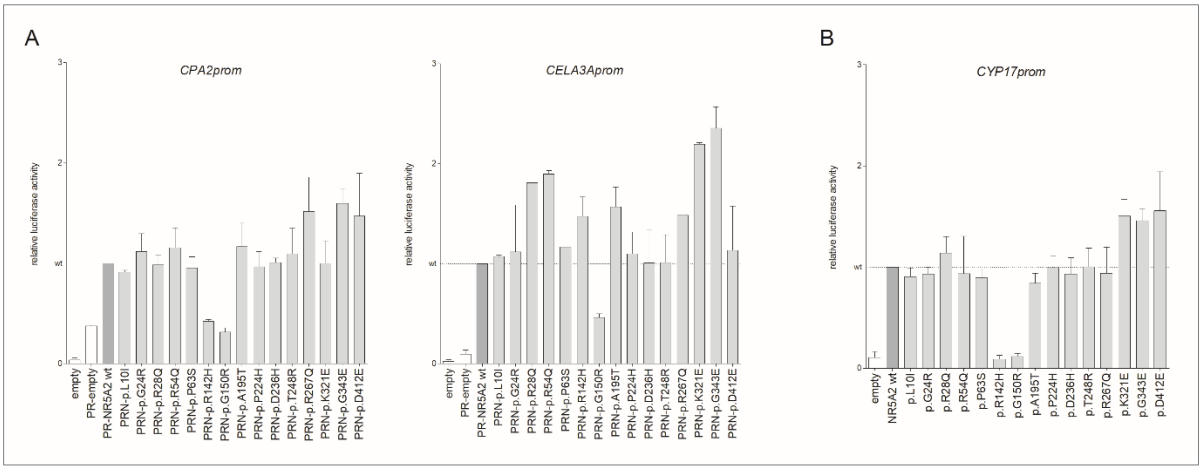


Figure 6.11: Luciferase activity of promoters targeted by NR5A2
Measurement of the relative luciferase activity after NR5A2 wildtype (wt) or mutants were overexpressed in HEK293 cells. Fifteen mutations were tested, and all measurements were referenced to the NR5A2 wt (= 1; dark grey bar). The empty vector pcDNA3.1+/C-(K)-DYK was used as negative control (empty; first white bar). Each measurement was done in a technical triplicate for a biological n = 3. (**A**) NR5A2 in complex with PTF1a + RBPJL was co-transfected to target *CPA2prom* or *CELA3Aprom* upstream the luciferase gene. To exclude effects by the complex, the duo PTF1a + RBPJL (PR-empty; second white bar) was additionally transfected as control. (**B**) Transcriptional activity of single expressed NR5A2 on *CYP17prom* upstream of the luciferase gene.
Some assays were performed by Hannah Frank as part of her Master thesis. PR = PTF1a + RBPJL; PRN = PTF1a + RBPJL + NR5A2; *CPA2prom* = promoter of *CPA2* in pGL4.22 luciferase vector (same for *CELA3Aprom* and *CYP17prom*).

6.3.4.2 RBPJL mutations reduce relative luciferase activity

We functionally investigated 10 *RBPJL* mutations, all of which were found in at least one patient. We measured transcriptional activity on *CPA1prom*. The results in Figure 6.12 show a relative luciferase activity reduction to ≤0.5-fold in three missense and two nonsense mutations: p.E84K (0.5), p.V215A (0.4), p.R222P (0.5), p.Q200* (0.4), and p.R265X (0.4). Here, again all measured values were set in relation to the wildtype transfection of RBPJL in

combination with PTF1a (= 1; dark grey bar) and pcDNA3.1+/C-(K)-DYK was used as empty control. Please note that as with NR5A2, the RBPJL mutations affecting promoter activity are all located in the potential DNA binding domain of the transcription factor.

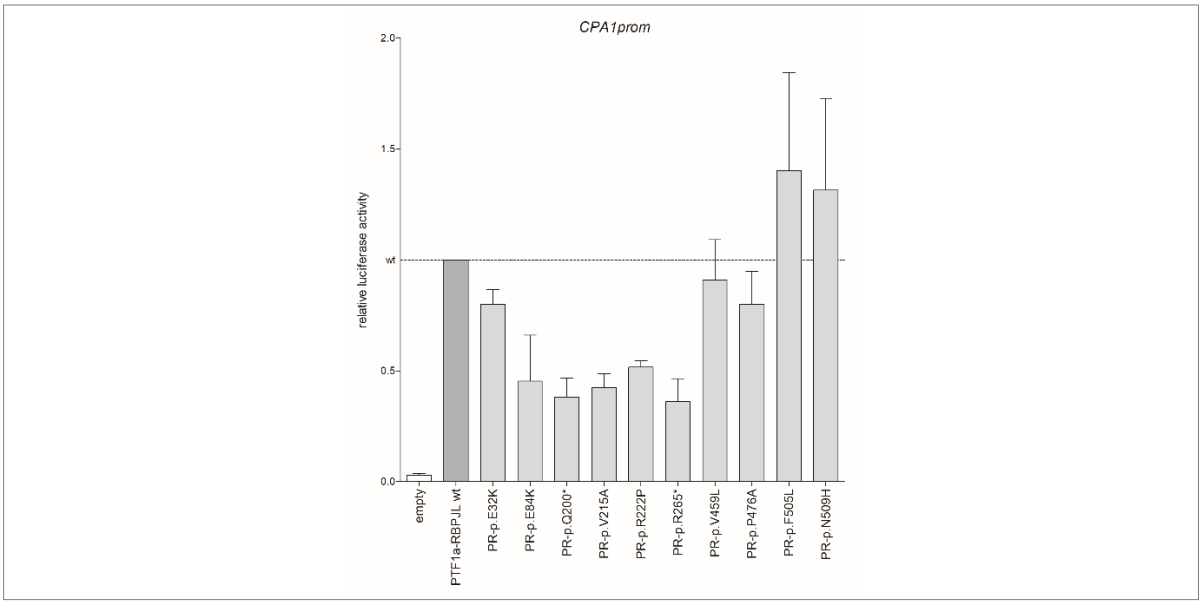


Figure 6.12: Luciferase activity of CPA1 promoter targeted by RBPJL
Relative luciferase activity after PTF1a and RBPJL wildtype (wt) or mutants were overexpressed in HEK293 cells co-transfected with the luciferase vector. We measured 10 RBPJL mutations as technical triplicates and set in reference to the RBPJL wt (= 1; dark grey bar). Biological replicates of n = 3. pcDNA3.1+/C-(K)-DYK served as negative control (empty; white bar). Some assays were performed by Hannah Frank during her Master thesis. PR = PTF1a + RBPJL; CPA1prom = promoter of CPA1 in pGL4.22 luciferase vector

6.3.4.3 HNF1a mutations do not affect promoter activity in luciferase assay

HNF1a mutations were tested with luciferase vectors harboring the *SPINK1* (*SPINK1prom*) or *MTTP* (*MTTPprom*) promoter fragment. We investigated 5 mutations: p.A161T, p.R168C, p.G183R, p.R278Q, and p.P580L (Figure 6.13). HNF1a was either transfected with PTF1a and RBPJL for *SPINK1prom* (**A**) or alone for *MTTPprom* (**B**). As we hypothesized disrupted transcriptional activity by HNF1a mutations, we expected reduced luciferase activities. However, none of the mutations showed an effect.

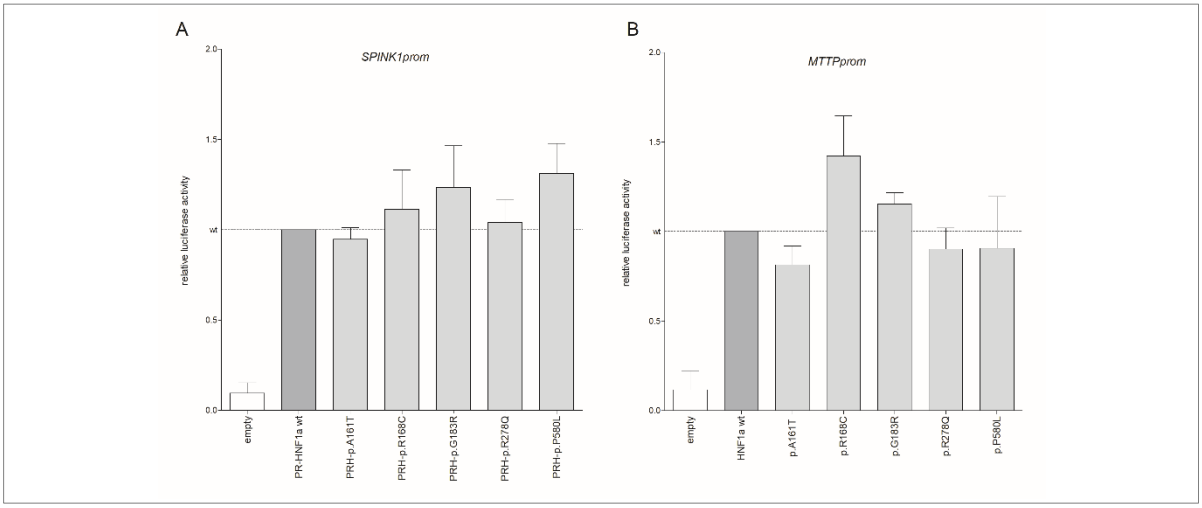


Figure 6.13: Luciferase activity of promoters targeted by HNF1a
Measurements of relative luciferase activity after HNF1a wildtype (wt) or mutants were overexpressed in HEK293 cells. The empty vector pcDNA3.1+/C-(K)-DYK served as negative control (white bar). All values were set in relation to the HNF1a wt (=1; dark grey bar). Each measurement was done in technical triplicates for a biological n = 3. (**A**) HNF1a in complex with PTF1a + RBPJL was co-transfected to target *SPINK1prom*. (**B**) Single expressed HNF1a on *MTTPprom* as target. Assay conduction was assisted by Hannah Frank during her Master thesis. PR = PTF1a + RBPJL; PRH = PTF1a + RBPJL + HNF1a; *SPINKprom* = promoter of SPINK1 in pGL4.22 luciferase vector (same for *MTTPprom*)

As we could not guarantee that the luciferase assay results have biological relevance or whether the experimental setup was flawed for HNF1a, we decided to include positive controls. The group of Prof. Anders Molven (University of Bergen) has published *HNF1a* loss-of-function mutations in MODY3 (Najmi et al. 2017; Althari et al. 2020) and generously provided us with two expression vectors containing the *HNF1a* loss-of-function mutations p.P112L and p.P447L as well as with the luciferase vector pGL3_RA. This vector harbors a 175 bp long fragment of the rat albumin promoter.

We additionally aimed to clone the HNF1a binding site for the human albumin (*ALB*) or for human β-fibrinogen (*FGB*) as triplicate into the basic pGL4.22 vector. This approach was based on the idea of Chi *et al* where they generated a vector with three small fragments of the HNF1a binding site in the *FGB* promoter (Chi et al. 2002). For this, the binding sites with a length of 75 bp for *ALB* or 169 bp for *FGB* were amplified with three different combinations of restriction enzyme cutting sites. While we were unable to ligate a triplicate of the fragments into the pGL4.22 vector, we still created a vector, which harbored two identical sequences for the binding site of HNF1a upstream the luciferase gene sequence (see Table 6.7).

Table 6.7: Promoter region with binding sites for HNF1a cloned into pGL4.22 vector

Gene	Name	Promoter region	TFBS for	Cloned length [bp]
ALB	Albumin	-129/+42 bp	HNF1a	2 x 75
FGB	β-fibrinogen	-120/-29 bp	HNF1a	2 x 169

TFBS = transcription factor binding site; exemplified explanation on promoter sequence -129/-42 bp = 129 base pairs in 5'-direction of 5'-UTR and 42 base pairs into 5'-UTR.

These newly cloned vectors were used for further luciferase assays to investigate *HNF1a* mutations. HNF1a wildtype or mutants were overexpressed in HEK293 cells and co-expressed with PTF1a and RBPJL as well as with the luciferase vectors of HNF1a target promoters. Luciferase activity measurements were conducted as one biological replicate in a technical triplicate (Figure 6.14). Experiments with the rat albumin promoter (**A**; p.GL3_RA) verifies the reductive effect of the deleterious p.P112L and p.P447L mutations, indicating a successful experimental setup. However, all five mutations showed similar luciferase activities than the HNF1a wildtype (wt = 1). Also, in the second approach with the double promoter regions of *ALB* and *FGB* (**B**), the results were similar. In conclusion, the investigated HNF1a mutations have no effect, leaving *HNF1a* unlikely as novel risk gene for CP.

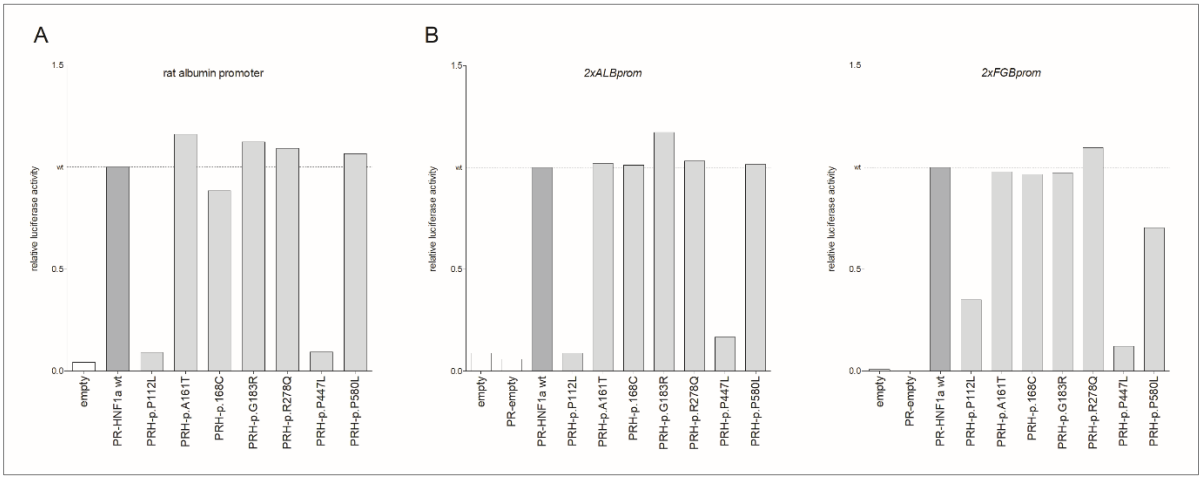


Figure 6.14: Luciferase activity of promoters targeted by HNF1a with positive controls included
HNF1a was co-expressed with PTF1a and RBPJL. Five mutations were tested. The empty vector pcDNA3.1+/C-(K)-DYK and the duo PTF1a + RBPJL served as negative controls (empty and PR-empty; white bars). As positive controls, two known MODY3 loss-of-function mutations (p.P112L, p.P447L) were analyzed. All values were set in relation to the HNF1a wt (=1; dark grey bar). We conducted one biological replicate with technical triplicates. (**A**) Assay with pGL3_RA with a 175 bp fragment of the rat albumin promoter. Vector was kindly provided by Prof. Anders Molven (University of Bergen). (**B**) Assay with luciferase vectors harboring a double fragment of *ALBprom* or *FGBprom.* PR = PTF1a + RBPJL; PRH = PTF1a + RBPJL + HNF1a; *2xALBprom* = two identical promoter regions of *ALB* in pGL4.22 luciferase vector (same for *2xFGBprom*)

6.3.5 Target gene expression analysis with qPCR

To confirm our results of the reporter gene assays, we analyzed the effects on endogenous gene expression. Upon overexpression of NR5A2 or RBPJL in Panc1 cells, we expected an altered mRNA expression of regulated target genes by the mutants in qPCRs. The panel for TaqMan assays consisted of different genes described in literature or by previous work done by Dr. Lara Unger in her PhD project (data not shown).

6.3.5.1 NR5A2

For NR5A2, the panel of target genes included transcription factors (*RBPJL*, *HNF1a*), pancreatic enzymes (*CPA1*, *CEL*, *PRSS1*) and inflammatory markers (*C3*, *IL18*, *IL32*). The combined results of four biological replicates are shown in Figure 6.15. The mRNA expression of *NR5A2* itself ensured us a successful overexpression, though there are some fluctuations between the mutations of NR5A2. The expression of RBPJL increased compared to the empty control (Figure 6.15; Panc1-*RBPJL*, white bar). However, this was not due to NR5A2, as the expression with PTF1a + RBPJL alone showed similar levels. Unfortunately, this was also true for target genes in the pancreatic enzyme as well as the inflammatory marker panel. Not only did we fail to show any differences in the mutant NR5A2 compared to the wildtype, but there was also no increase in mRNA expression when NR5A2 was taken out of the equation. In the panel of pancreatic enzymes, *CPA1* and *PRSS1* expression upon overexpression of PTF1a + RBPJL resulted in the same level of expression than with NR5A2 present. Therefore, we concluded that here NR5A2 is not the driving factor for mRNA expression. The same holds true for the inflammatory markers *C3* and *IL32*.

Additionally, the expression levels of *CEL* and of *IL18* both did not differ upon overexpression of the different transcription factor constellations and the empty control. Thus, these genes seem not to be regulated by PTF1a, RBPJL and NR5A2.

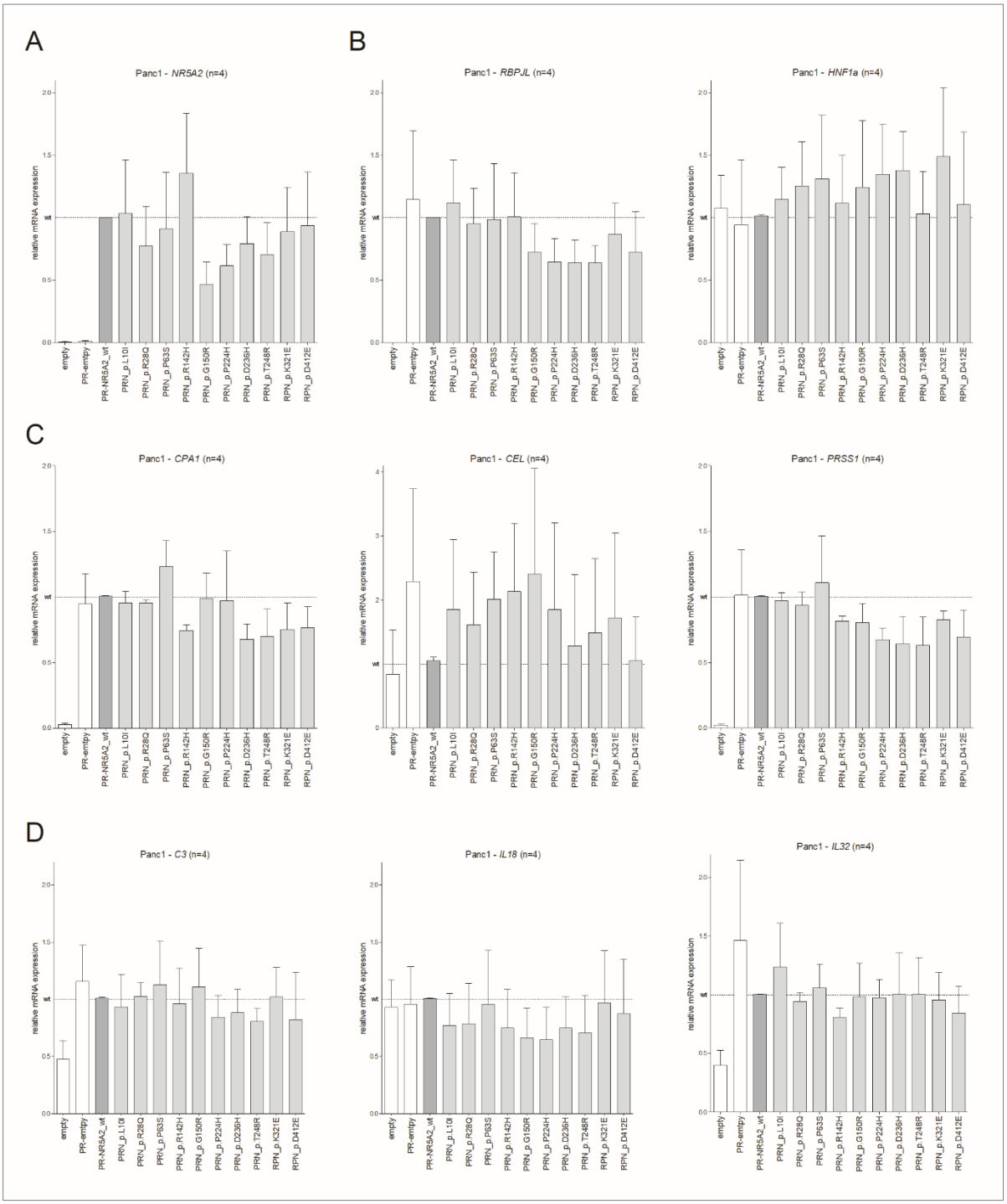


Figure 6.15: Target gene expression in Panc1 upon the overexpression of NR5A2 wt or mutant in complex
NR5A2 wildtype (wt) or mutants were overexpressed in Panc1 cells in complex with PTF1a and RBPJL (PR). RNA was isolated 72 h post-transfection and transcribed into cDNA. Relative target gene mRNA expression was then measured by qPCR using TaqMan assays and normalized to wt transfected Panc1 cells (= 1; horizontal line). GAPDH and β-actin expressions served as internal control. Vector pcDNA3.1+/C-(K)-DYK and the transfected duo PTF1a + RBPJL served as negative controls (empty and PR-empty; white bars). Each measurement was conducted as technical triplicates. The biological replicates are stated above each graph together with the target gene. (**A**) mRNA expression of NR5A2 itself. The panel of target genes is grouped into: (**B**) Transcription factors, (**C**) Pancreatic enzymes and (**D**) Inflammatory markers. Standard deviation is shown.
PR = PTF1a + RBPJL; PRN = PTF1a + RBPJL + NR5A2

Consequently, we could not verify the effects of NR5A2 mutations found in the reporter gene assays when analyzing endogenous target gene expression. This arose the questioning of our choice of cell line. We originally argued that Panc1 would have a closer phenotype to a pancreatic cell type. However, its origin from PDAC leaves it with a modal chromosomal number of 63 (Lieber et al. 1975) and altered metabolic and proteomic profiles (Watanabe et al. 2012; Kim et al. 2014). Thus, this cell line is far from an acinar cell. During cancer progression, acinar cells probably underwent acinar to ductal metaplasia (ADM). With the loss of acinar specific gene expression, the cells recondition towards a duct-like phenotype. This disqualifies pancreatic cancer cells as an acinar cell model. Consequently, we decided to change to the HEK293 cell line as used for the reporter gene assays.

We overexpressed NR5A2 wildtype or mutant in HEK293 cells (Figure 6.16). Again, the mRNA expression of *NR5A2* itself was successful for the wildtype and all mutants. Looking at the transcription factors, *RBPJL* was expressed independent of whether NR5A2 was present or not. *HNF1a* did not appear to be regulated by the transcription factor complexes as expression did not differ from the empty control.

As NR5A2 target genes, we chose a panel of six pancreatic enzymes. Indeed, three of them were now regulated by NR5A2, as expression clearly differed between PTF1a + RBPJL alone (Figure 6.16; white bars) and in combination with NR5A2 wt (Figure 6.16; dark grey bars). This is true for *CPA1*, *PRSS1* and *CELA3AB*. Additionally, we detected reduced expression of those target genes in the mutants p.R142H and p.G150R. Both variants reduce the mRNA expression of *CPA1* to the level of PTF1a + RBPJL control (0.3-fold). *PRSS1* expression was also downregulated to 0.2-fold. Thus, a complete loss-of-function of NR5A2 by these mutations occurred in both promoters. For *CELA3AB* expression, although levels did not completely reduce to baseline control levels (0.1-fold), there was still a reduction to 0.5-fold (p.R142H) and 0.3-fold (p.G150R), respectively. Overall, these results are consistent with the luciferase assay results.

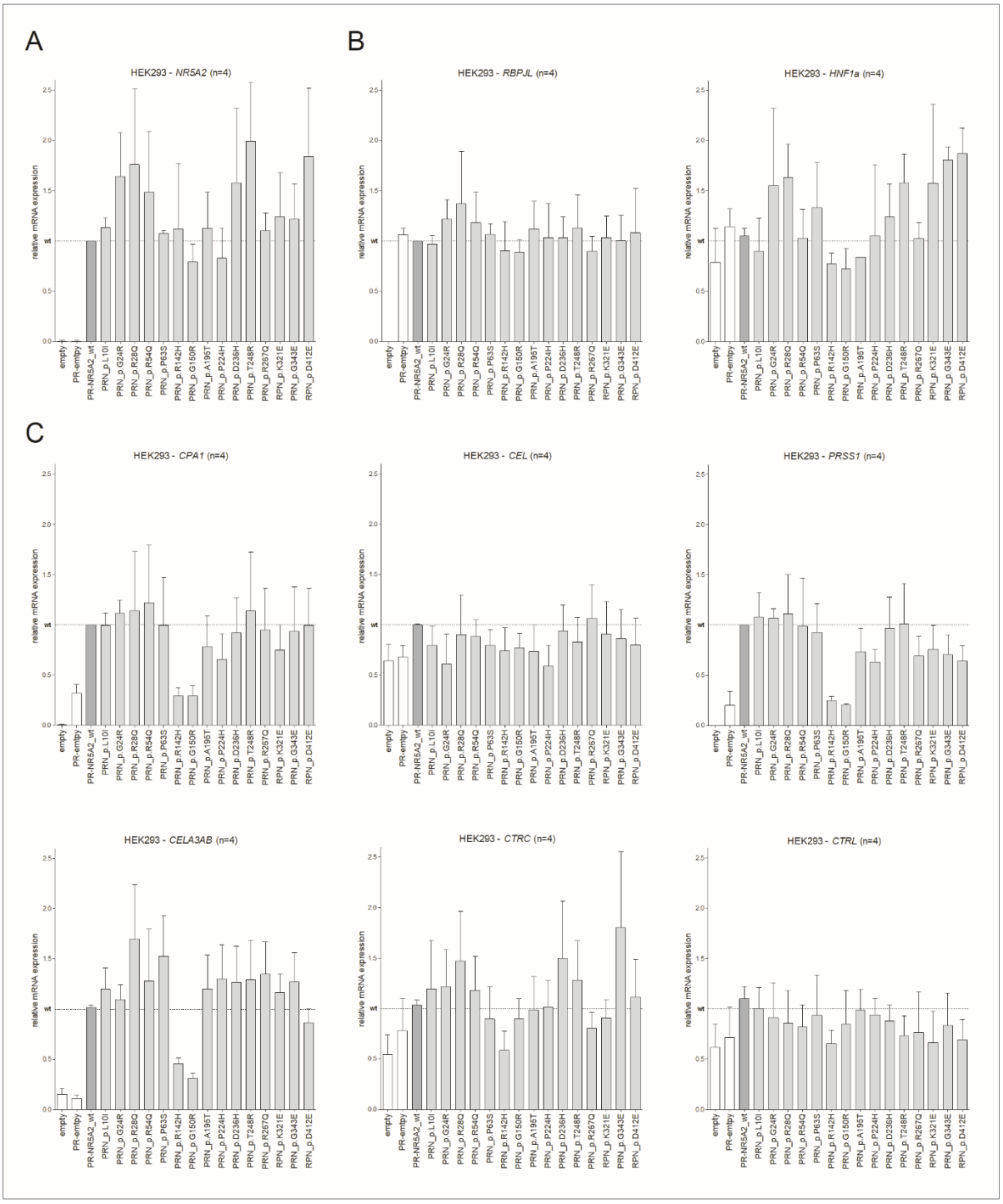


Figure 6.16: Target gene expression in HEK293 cells upon overexpression of NR5A2

NR5A2 wildtype (wt) or mutants were overexpressed in HEK293 cells in complex with PTF1a and RBPJL (PR). Relative target gene mRNA expression was measured by qPCR using TaqMan assays and normalized to wt transfected cells (= 1; horizontal line). GAPDH and β-actin expressions served as internal controls. The vector pcDNA3.1+/C-(K)-DYK and the transfected duo PTF1a + RBPJL served as negative controls (empty and PR-empty; white bars). We conducted each measurement as technical triplicates. The biological replicates are stated above each graph together with the target gene. (**A**) mRNA expression of NR5A2 itself. The panel of target genes is grouped into: (**B**) Transcription factors and (**C**) Pancreatic enzymes. Standard deviation is shown.

PR = PTF1a + RBPJL; PRN = PTF1a + RBPJL + NR5A2

6.3.5.2 RBPJL

The qPCR results show a transcriptional regulation of all tested target genes by RBPJL in complex with PTF1a. *RBPJL* expression itself was roughly on the same level in wildtype and in the mutants. As target genes, we selected two zymogen granule genes (*GP2*, *REG1A*) and three pancreatic enzymes (*CPA1*, *CEL*, *PRSS1*). In the zymogen granule genes, only a minor effect on *GP2* expression was found (0.5-fold in p.E84K and p.Q200*, 0.4-fold in p.V215A, and 0.6-fold in p.R222P) (Figure 6.17). For the expression of *REG1A*, this effect was only detectable in p.E84K down to 0.4-fold. In pancreatic enzymes, RBPJL mutations caused a stronger reduction. In all three genes, DNA binding domain mutations (p.E84K, p.Q200*, p.V215A, p.R222P, and p.R265*) reduced mRNA expression. Especially in *CPA1*, the levels are partly down to baseline levels with 0.1-fold (p.E84K), 0.2-fold (p.Q200*, p.V215A, p.R265*) and 0.3-fold (p.R222P). Note that p.P476A also reduced these genes to 0.5-fold (*CPA1*), 0.4-fold (*CEL*) and 0.5-fold (*PRSS1*).

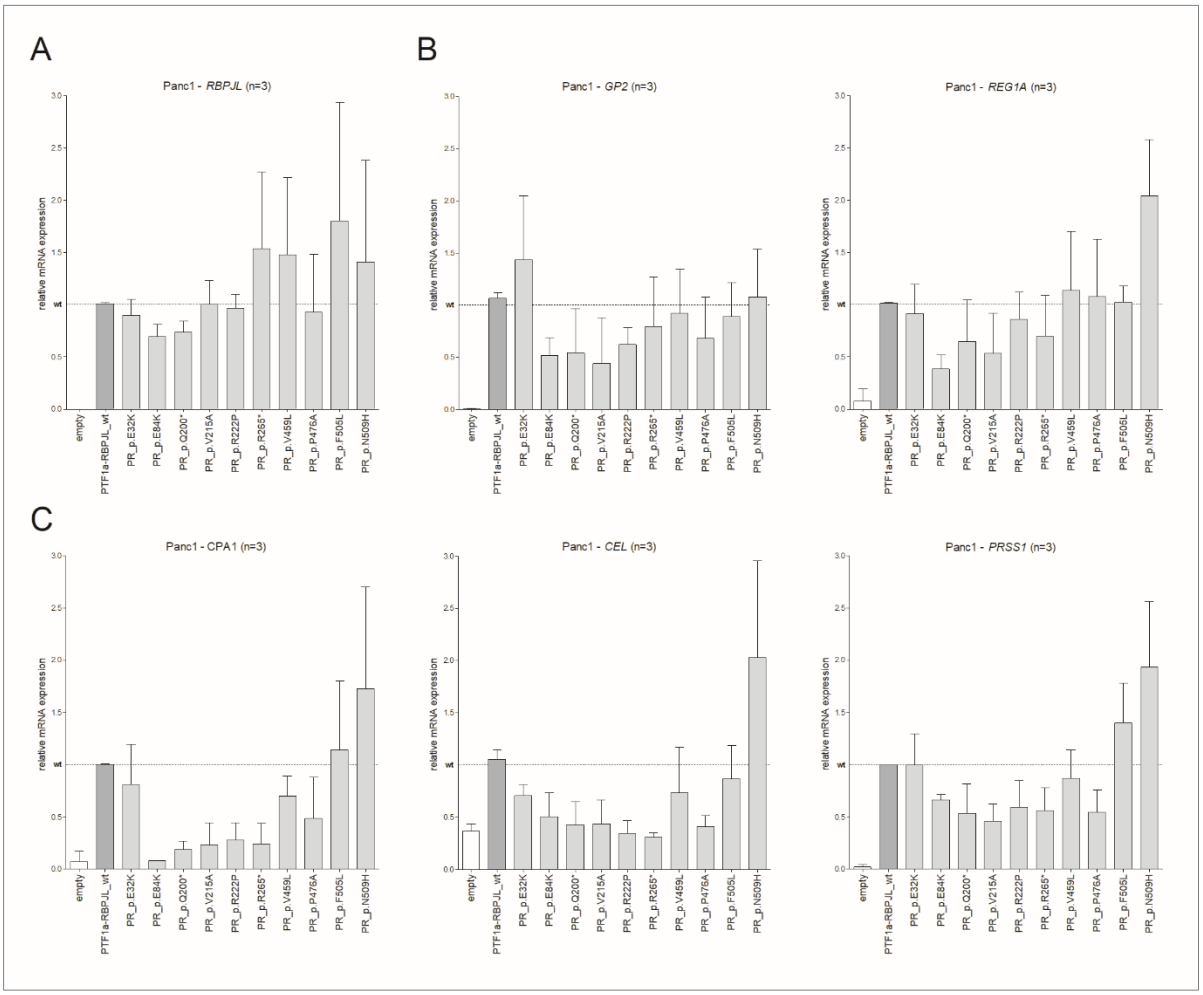


Figure 6.17: Target gene expression in Panc1 upon overexpression of RBPJL wt or mutants with PTF1a
Relative mRNA expression was measured by TaqMan qPCR and normalized to wt transfected cells (= 1; horizontal line) in technical triplicates. GAPDH and β-actin served as internal controls. pcDNA3.1+/C-(K)-DYK served as negative control (empty; white bars). The biological replicates are stated above each graph together with the target gene. (**A**) mRNA expression of RBPJL itself, (**B**) Zymogen granules and (**C**) Pancreatic enzymes. Standard deviation is shown. PR = PTF1a + RBPJL

6.3.6 Statistics

The initial WES analysis for the transcription factors *NR5A2*, *RBPJL* and *HNF1a* indicated a significance of *NR5A2* only (Table 6.5). We noticed a conformity between those calculations with our experimental results that we obtained. Especially the two mutations in the DNA binding domain, p.R142H and p.G150R, reduced NR5A2 function.

Based on our functional analysis, we calculated the mutation frequencies in affected individuals and in controls using a two-tailed Fisher's Exact test. Results for NR5A2 showed a significant effect of the missense mutations p.R142H and p.G150R and the nonsense mutations p.R28* and c.202+1 found in the patient cohort ($p = 0.013$; Table 6.8). While each mutant was found once in the patient cohort, none was found in controls.

Table 6.8: Fisher's Exact test results on NR5A2 mutations

Consequence	Mutation	CP (n=1,059) (%)	Controls (n=2,099) (%)	p-value	OR
Luciferase activity <50 %	p.R142H p.G150R	2 (0.2)	0 (0.0)	0.112	Inf
Endogenous gene expression >0.5	p.R142H p.G150R	2 (0.2)	0 (0.0)	0.112	Inf
Including non-sense mutations					
	p.R28* c.202+1 G>T p.R142H p.G150R	4 (0.4)	0 (0.0)	0.013	Inf

OR = odds ratio; Inf = infinite

Regarding *RBPJL*, functional analysis showed effects on the transcriptional activity in several mutations. Both, luciferase assays and mRNA expression of target genes in qPCR resulted in a reduction down to 50 % or lower. In the Fisher's Exact test, however, no significant differences were found ($p = 0.173$; Table 6.9). Even though mutations clearly showed functional effects, we dropped *RBPJL* as risk gene.

Table 6.9: Fisher's Exact test results on RBPJL mutations

Consequence	Mutation	CP (n=1,059) (%)	Controls (n=2,099) (%)	p-value	OR
Luciferase activity <50 %	p.E84K p.Q200* p.V215A p.R222P p.R265*	5 (0.5)	4 (0.2)	0.173	2.48
Endogenous gene expression >0.5	p.E84K p.Q200* p.V215A p.R222P p.R265*	5 (0.5)	4 (0.2)	0.173	2.48

OR = odds ratio

Thus, we focused on *NR5A2*. As the mutations p.R142H and p.G150R are both located in the DNA binding domain, we decided to investigate the protein-DNA binding capacity of these mutations using Electrophoretic Mobility Shift Assay (EMSA).

6.3.7 Protein-DNA binding capacity of NR5A2 in EMSA

Oligonucleotides with a length of 45 bp harbored the potential binding site of NR5A2 in the target genes *CPA2* and *CYP11A1*. Figure 6.18 shows the EMSA image of two biological replicates (n1 in line 1-8, n2 in line 9-18). The pcDNA3.1+/C-(K)-DYK vector served as control and in the second biological replicate, the mutation p.L10I was included as negative control. In both replicates, the band distribution pattern was similar. However, all visible bands are also shown in the empty control lanes (lane 1, 5, 9, and 14) and therefore are unlikely to be NR5A2 specific bindings. Furthermore, band intensity did not differ in the wildtype compared to mutant NR5A2 extracts.

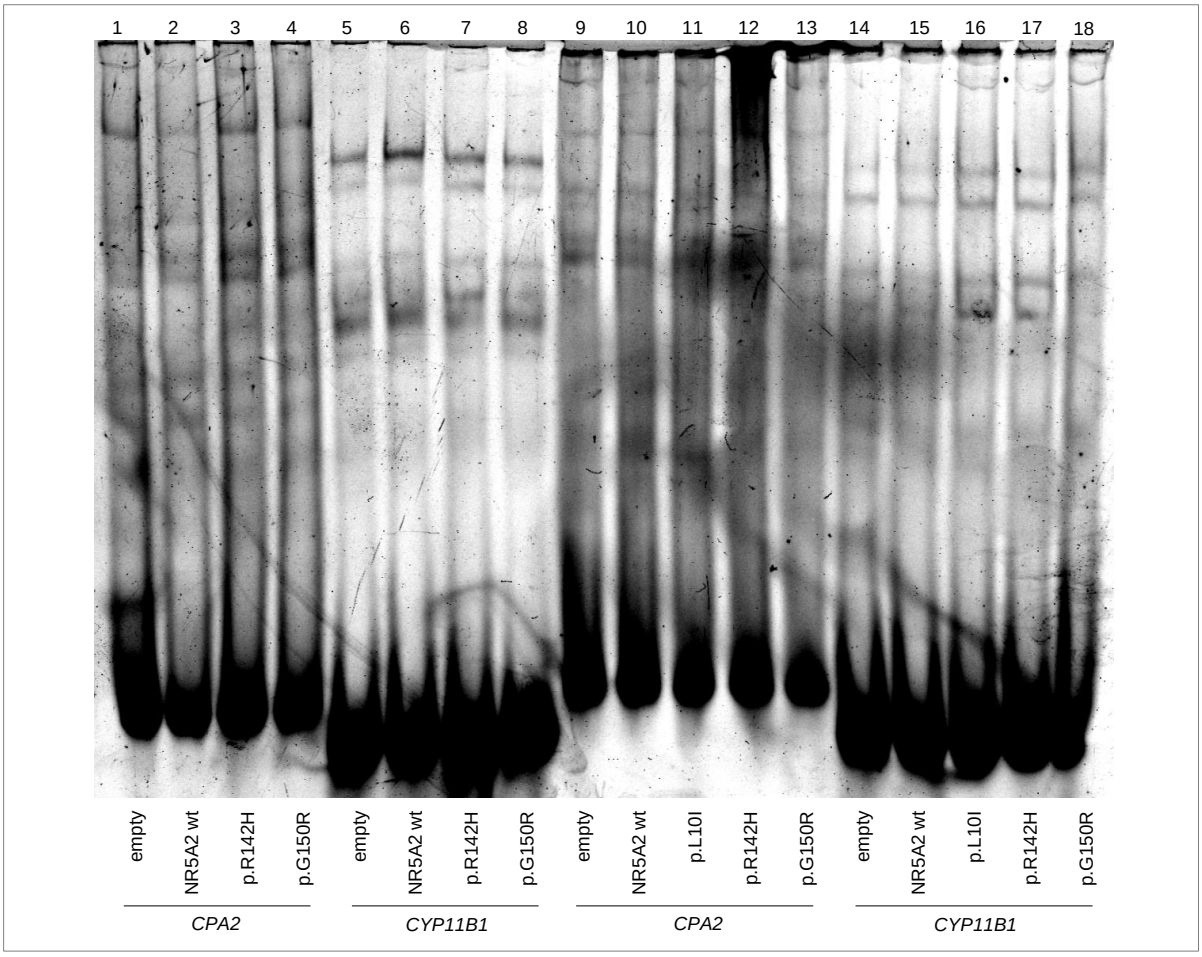


Figure 6.18: Nuclear protein binding of HEK293 cells upon overexpression of NR5A2
pcDNA3.1+/C-(K)-DYK (empty) was used as baseline control. Cy5-labelled oligonucleotides harbored a 45 bp promoter region of *CPA2* (lane 1-4 and 9-13) or *CYP11A1* (lane 5-8 and 14-18) as binding site for NR5A2. Two biological replicates are shown. n1 includes the mutations p.R142H and p.G150R (lane 1-8). n2 additionally includes p.L10I as potential negative control as it did not show any effect on transcription previously (lane 9-18). EMSA experiments were assisted by Beate Rauscher.

7 Discussion

Chronic pancreatitis (CP) is a relapsing or continuous inflammatory disease of the pancreas which can significantly impair both endocrine and exocrine functions (Witt et al. 2007). It has a strong genetic predisposition, especially in early-onset CP. Since the genetic origin of pancreatitis was first hypothesized in 1952 (Comfort and Steinberg 1952), several disease predisposing genes and loci have been identified. However, to this day there are still many CP cases with no identified underlying cause. Despite considerable efforts, the genetic basis of non-alcoholic CP (NACP) is only partially known. In our project, we aimed to identify novel genetic risk factors based on an extended candidate approach and the whole exome sequencing (WES) of 1,059 pediatric patients. We identified a panel of genes which are expressed in the pancreas and regulate its development and maintenance. We narrowed down this panel to further investigate five genes as focus for different potential pathological pathways: the water channel aquaporin 12 (*AQP12*), the digestive enzyme chymotrypsinogen-like protease (*CTRL*) and the transcription factors *NR5A2*, *RBPJL* and *HNF1a*. As overall results, the variants p.R142H and p.G150R in *NR5A2* might play a disease-causing role, while most other obtained results do not support an association with CP.

7.1 *AQP12* mutations are not associated with chronic pancreatitis

We investigated German NACP patients and controls regarding an association of *AQP12A* and *AQP12B* variants with the development of CP. We hypothesized that dysfunction of zymogen granules maturation or exocytosis promotes the development of CP. Under normal conditions, acinar cells store digestive enzymes as inactive precursors - so-called zymogens - in secretory granules. During maturation, the granule volume is reduced to concentrate the stored zymogens. Osmotic swelling of the granule then leads to exocytosis. Both processes require a coupled flow of water and ions (Thévenod 2002). The specific localization of AQP12 in the membrane of intracellular organelles of acinar cells suggests a role in maturation and exocytosis of zymogen granules (Itoh et al. 2005).

In the past, the physiological role of AQP12 was investigated by Ohta *et al.* as they generated pancreas-specific *Aqp12*$^{-/-}$ mice. Under normal conditions, there were no phenotypic differences between *Aqp12*$^{-/-}$ mice and their wildtype littermates regarding body weight, blood parameters, morphology, or pancreatic secretion. However, after overstimulation with caerulein, *Aqp12*$^{-/-}$ mice formed more and larger exocytotic vesicle, suggesting a role of AQP12 in early zymogen granule fusion. When water flow is disturbed, zymogen granules may accumulate, and a lower or negative intracellular pressure leads them to merge. This consequently forms larger and deformed vacuoles. Furthermore, more pancreatic organ damage such as more edema and necrosis was detected in *Aqp12*$^{-/-}$ mice after application of

caerulein (Ohta et al. 2009). Hence, we assume both, *AQP12A* and *AQP12B* as potential risk genes for the development of chronic pancreatitis. Our study, however, failed to show an association between *AQP12A* or *AQP12B* mutations and CP. Particularly, 2.1 % of patients and 2.8 % of controls were homozygous for an *AQP12B* nonsense variant. Further, the p.S152Tfs*24 variant, which was found in over 24 % of patients and controls, indicates a rather dispensability of this protein for exocrine pancreatic function.

Our negative findings do not generally rule out an involvement of aquaporins in the pathogenesis of CP. Pancreatic ductal fluid secretion was shown to be reduced in 4-week caerulein-treated mice *in vivo* and *in vitro* in isolated pancreatic duct fragments of 1-2-3-4-week caerulein-treated mouse pancreata. Balázs and colleagues showed increased epithelial mucus production in the small pancreatic ducts in the early phase of CP, which was paralleled by impaired pancreatic fluid secretion (Balázs et al. 2018). In addition, decreased secretion in CP may be associated with CFTR dysfunction (Hegyi et al. 2016). Several members of the AQP family have been reported to be expressed in the exocrine pancreas. While AQP8 and AQP12 are almost exclusively expressed in acinar cells, AQP1 and AQP5 are the major pancreatic ductal aquaporins (Burghardt et al. 2003). Thus, genetic defects in any of these other three aquaporins could affect CP risk.

A recently published study examined the role of AQP1 in pancreatitis. Bicarbonate and fluid secretion significantly decreased in Aqp1 deficient mice compared to wildtype mice. Noteworthy, acute pancreatitis (AP) was aggravated in *Aqp1*$^{-/-}$ mice after caerulein challenge. Serum amylase levels and the extent of pancreatic necrosis were significantly higher in knock-out compared to wildtype mice, while the extension of edema and inflammatory cell infiltration did not differ (Venglovecz et al. 2018). In humans, a lack of AQP1 leads to the absence of Colton blood group antigens and reduced osmotic water permeability of red blood cells. Preston and colleagues reported three individuals, two of whom were homozygous for different nonsense mutations, and the third had a missense mutation encoding a nonfunctioning AQP1 molecule. Surprisingly, none of the three suffer from any obvious clinical consequences, raising questions about the physiological importance of AQP1 and implying that other mechanisms might compensate for its absence (Preston et al. 1994).

Mutations have also been described in humans for the other ductally expressed aquaporin, AQP5, but these also do not result in a pancreatic phenotype. Missense mutations in this gene cause autosomal dominant diffuse non-epidermolytic palmoplantar keratoderma characterized by a white, spongy appearance of the affected skin areas upon contact with water (Blaydon et al. 2013).

So far, no mutations in *AQP8*, which is almost exclusively expressed in the pancreas in acinar cells, have been described in humans. In mice, knock-out of this gene does not result in an

apparent pancreatic phenotype (Yang et al. 2005). However, a caerulein challenge or another model for experimentally induced pancreatitis in *Aqp8*$^{-/-}$ mice has not been described so far. Taken together, human and mouse data do not currently support a significant role for these aquaporins in the pathogenesis of pancreatitis.

7.2 Mutations in *CTRL* do not show an association with chronic pancreatitis

Chymotrypsins are the second most abundant proteases in the pancreatic fluid. Their affinity to cleave all three isoforms of trypsinogen shows their role in the protective mechanism of intrapancreatic trypsinogen activation. Out of the four members of the chymotrypsin family, *CTRC*, *CTRB1* and *CTRB2* have already been associated with the development of CP (Rosendahl et al. 2008; Rosendahl et al. 2018). Therefore, we shifted our focus toward the fourth chymotrypsin. We investigated *CTRL* as potentially novel risk gene for CP. Though, its exact function as proteolytic enzyme is still not unraveled, we hypothesized that *CTRL* mutations might interfere with a proper protein-folding leading to ER stress. Thus, *CTRL* variants might be classified into the misfolding-dependent pathway of CP. Loss-of-function mutations in *CTRL* may underlie several possible mechanisms: reduced secretion, catalytic defects, or an increased degradation through trypsin (Beer et al. 2013). In our project, we identified several variants which diminish secretion of CTRL or impair its function. However, these effects were found for both mutations present in patients and mutations present in controls (Table 7.1).

Table 7.1: Summarized results on CTRL mutations

Exon	Position (hg19)	Nucleotide change	AA change	rs number	CADD	REVEL	CP (n=1,059) (%)	Controls (n=2,099) (%)	Secretion	Activity (S)	Activity (L)	*p*-value
1	16:67965751	c.5T>C	p.L2S	rs144681992	16.58	0.086	1 (0.9)	1 (0.05)	Yes	88.2	57.1	1.0
1	16:67965712	c.44C>A	p.S15Y	rs144605661	19.50	0.290	3 (0.3)	-	Yes	112.8	140.9	0.04
2	16:67965099	c.58G>A	p.G20S	rs140026167	22.4	0.266	2 (0.2)	-	Yes	65.3	50.5	0.11
2	16:67964893	c.166G>A	p.G56S	rs761948651	25.7	0.273	1 (0.9)	-	Yes	51.7	34.4	0.34
3	16:67964878	c.181G>A	p.G61S	rs141577010	34.0	0.673	-	1 (0.05)	Yes	35.0	32.4	1.0
5	16:67964631	c.316C>T	p.R106W	rs1188402196	20.5	0.220	1 (0.9)	-	Yes	90.2	83.1	0.34
5	16:67964475	c.335G>A	p.S112N	rs137951739	0.02	0.028	-	1 (0.05)	Yes	91.0	119.9	1.0
5	16: 67964361	c.449C>T	p.T150I	rs11552953	16.58	0.080	-	-	Yes	96.1	89.7	1.0
6	16: 67964203	c.518A>G	p.H173R	rs1134760	9.29	0.020	-	-	Yes	117.1	107.6	1.0
6	16:67964192	c.529G>T	p.V177L	rs776839938	22.8	0.064	-	1 (0.05)	Yes	61.1	102.1	1.0
6	16:67964119	c.602G>A	p.C201Y	rs201073485	28.9	0.923	1 (0.9)	-	No	2.2	1.9	0.34
6	16:67964098	c.623C>T	p.S208F	rs762981045	26.0	0.506	1 (0.9)	-	Yes	1.2	5.9	0.34
7	16:67963989	c.643G>A	p.G215R	rs780563969	32.0	0.714	-	1 (0.05)	No	0.7	29.8	1.0
7	16:67963977	c.653_654del	p.L218Rfs*33	-	-	-	-	3 (0.14)	No	-1.3	0.8	0.56
7	16:67963974	c.658T>G	p.C220G	rs147657411	27.3	0.599	2 (0.2)	-	Reduced	24.8	39.9	0.11
7	16:67963944	c.688G>A	p.G230S	rs774504385	31.0	0.801	-	2 (0.10)	No	1.4	4.2	0.55

AA = amino acid; CADD = combined annotation dependent depletion; REVEL = rare exome variant ensemble learner; CP = chronic pancreatitis; (S) = supernatant; (L) = lysate; P-values were calculated by two-tailed Fisher's Exact test with a 2x2 contingency table comparing carrier frequencies. Mutation p.G37E was excluded in statistical analysis due to a separate control cohort.

7.2.1 Secretory dysfunction in CTRL mutants does not lead to ER stress

All variants with impaired secretion are located at the C-terminal end of the protein, which suggests this domain to be important for the secretory function of CTRL. The nonsense variant p.L218Rfs*33 could not be detected in the Western blot either in the supernatant or in the lysate. Due to the serious amino acid sequence change, a premature stop codon is the consequence which shortens the protein by 14 amino acids. Despite this, it further leads to the disruption of two disulfide bonds which probably diminish protein folding. This indicates that either CTRL is not synthesized at all due to nonsense-mediated decay of its mRNA, or that the protein is rapidly degraded. CTRL has a total of five disulfide bonds, which are generally crucial for the tertiary structure of the protein. Their disruption by the variants p.C201Y and p.C220G severely changes CTRL structure and may cause a complete loss-of-function and therefore disturbing the secretion into the extracellular compartments (Braakman et al. 1992).

We did not detect the variants p.C201Y, p.G215R, and p.G230S in the supernatant, whilst a clear visible band in the cell lysate on the Western blot proves a successful production of the protein. This fostered the assumption that an intracellular retention due to misfolding in the ER lumen might be the underlying cause of impaired secretion (Kereszturi et al. 2009). However, we failed to detect ER stress in any *CTRL* variant based on two experimental approaches. Both the mRNA expression levels of *BiP* and *XBP1* splicing PCR did not reveal ER stress. Due to being a low abundance protein, its intracellular retention might not have a profound impact to appear in ER stress measurements. Similar results were obtained in misfolding SPINK1 variants. The mutations p.D50E, p.Y54H, and p.R67C led to intracellular retention and degradation (Király et al. 2007), however, without eliciting ER stress due to low expression levels (Boulling et al. 2007; Boulling et al. 2012).

7.2.2 Mutant position determines various causes of catalytic defects

Ten CTRL mutants showed reduced or lost activity in our assays. Depending on their position in the protein sequence, we hypothesize various underlying causes. Similar to the variant location to impair secretion, mutations which diminish activity are also predominantly located at the C-terminal end of the protein. A complete loss-of-function can be explained by severe structural changes of CTRL (Braakman et al. 1992). In the cell lysate samples, an activity loss (<10 %) was detected in the variants p.C201Y, p.S208F, p.L218Rfs*33, and p.G230S. Besides p.C201Y and p.C220G which disrupts the disulfide bonds p.C187-p.C201 and p.C155-p.C220, respectively and p.L218Rfs*33 which completely alters the amino acid sequence, the mutations p.G56S and p.G61S are located close to the disulfide bond p.C60-p.C76. Molecular modeling studies of the fibroblast growth factor receptor 2 (FGFR2) have shown that non-cysteine mutations p.W290G and p.T341P, which lie close to the

disulfide-bonded Cys-278 and Cys-342, respectively, disrupt the disulfide bond by causing conformational changes. This led to abnormal receptor dimerization and activation (Robertson et al. 1998). As our functional analysis revealed a reduced activity (25-75 %) in these mutants, we propose disrupted disulfide bonds through an abnormal conformation which impairs the enzyme's catalytic activity. The same applies for p.S208F which may interfere with the disulfide bond p.C210-p.C239.

As a proteolytic enzyme, CTRL contains active site residues characteristic for the serine protease family (Larsen et al. 1993). The histidine residue (p.H76, basic), the aspartic acid residue (p.D121, acidic) and the serine residue (p.S214, nucleophilic) form the catalytic triad of CTRL, which represents the nucleophile-base-acid pattern generally conserved for serine peptidases. Embedded in the active site of the enzyme, the residues interact to cleave the substrate (Polgár 2005). Even though, the residues are far apart on the primary protein structure, the 3D structure brings them together in a precise position. Thus, the surrounding sequence is of importance for a successful formation of the catalytic triad. Mutation p.G215R is located relatively close to the serine component. An amino acid change here could inhibit proper triad formation but also affect substrate binding and consequently lead to impaired hydrolysis.

Proteolytic impairment was also found in CTRL mutations that were rather N-terminally localized. The variants p.G20S and p.G37E both lead to reduced or lost activity, as they are located close to the cleaving site of the signal peptide (p.G18 ^{v} p.C19) and the activation peptide (p.R33 ^{v} p.I34). By destructing existing cleavage sites or exposing new ones, this could inhibit proper cleavage of the zymogen through trypsin (Rosendahl et al. 2008) and consequently interfere with its activation.

Even though the effects on secretory function and proteolytic capacity are serious in parts, these results were detectable in variants found in both patients and controls. Consequently, our data does not suggest an association with CP. Loss of CTRL might have only minor effects, as it makes up only ~10 % of total chymotrypsin content (Mosztbacher et al. 2020). Therefore, we suggest a somewhat backup mechanism by other chymotrypsin isoforms. Though with different preferences and affinities, all four isoforms of chymotrypsinogens, CTRB1, CTRB2, CTRC and CTRL, hydrolyze peptide bonds at tyrosine and phenylalanine residues (Szabó and Sahin-Tóth 2012). Therefore, they may substitute for each other's certain function when one fails due to genetic alterations.

7.2.3 Mice studies are needed to further unravel the function of CTRL

In a recent study, *Ctrl-KO* mice were created from C57BL/6N mice using CRISPR-Cas9 genome engineering to investigate the role of CTRL in caerulein-induced AP. The *Ctrl-KO* mice did not differ in phenotype or behavior from control mice and pancreatic morphology and histology were unchanged (Mosztbacher et al. 2020). While homozygous *Ctrl-KO* mice did not express any detectable Ctrl protein in the pancreas, the overall chymotrypsinogen level was reduced by only 10 %. Mosztbacher and colleagues thus concluded CTRL to be a rather low-abundance isoform of chymotrypsin (Mosztbacher et al. 2020). The treatment with caerulein for a secretagogue hyperstimulation-induced pancreatitis in both *Ctrl-KO* and control mice did not results in significant metabolic, morphologic, or proteolytic differences. This indicates that CTRL-dependent changes in the proteolytic activity does not impact the severity of caerulein-induced pancreatitis. However, it should be noted, that this mouse model was used to investigate a rather AP phenotype. Studies in the past have shown a clear difference in phenotype and experimental outcome whether an acute or chronic caerulein-induced pancreatitis model has been used. For example, in the case of a TRPV6 study regarding CP, the group of Masamune *et al* established a *TRPV6* knock-out mouse model for CP (Masamune et al. 2020). While AP models with single caerulein-treatment did not show any effect for TRPV6 (unpublished data by the research group of Prof. Hana Algül), $TRPV6^{mut/mut}$ mice developed severe CP with chronic caerulein administration than controls by increased levels of pancreatic enzymes, histological alterations, and fibrosis. Here, mice were injected with 0.1 µg caerulein/g mouse body weight each hour for 8 hours and 2 consecutive days to induce CP.

In addition, it should be noted that data from mouse experiments are not necessarily conferrable to humans, as there are discrepancies between mouse and human phenotypic outcomes. With respect to *CFTR*, mouse and human amino acid sequence share only a 78 % homology (Tata et al. 1991). Genetic variants lead to the absence of the CFTR channel in the exocrine pancreas (Riordan et al. 1989), which causes cystic fibrosis in humans, a fatal multiorgan disease. The most common disease-causing mutation in humans is p.F508del (Semaniakou et al. 2018). Mouse models harboring p.F508del, however, showed only a mild pancreatic phenotype (Semaniakou et al. 2018), which fails to mirror the severity of cystic fibrosis in humans.

In conclusion, genetic *CTRL* variations show to some extent severe interference in the protease function and secretion. However, these effects are not associated with a CP phenotype. Due to the missing understanding of its actual function, we are unable to identify possible compensating mechanisms to a loss-of-function in CTRL and it remains a matter of further research to this day.

7.3 Transcription factors

Organogenesis and maintenance of the pancreas is highly influenced by a complex transcriptional regulation. Several factors play an important role during different stages of development. Due to findings that some transcription factors contribute to pancreatic cancer risk (Petersen et al. 2010; Flandez et al. 2014; Naqvi et al. 2017), we hypothesized a contribution to CP as well. Thus, we investigated *NR5A2*, *RBPJL*, and *HNF1a* as representative players in different potential pathogenic pathways.

7.3.1 Mutations in the DNA binding domain of NR5A2 diminish its function

In pancreatitis, inflammatory processes injure the exocrine pancreas. During those attacks, damage recovery is initiated through the regeneration of acinar cells. It has been shown that NR5A2 plays a crucial role in the severity of acute inflammation as adult heterozygous *Nr5a2* knock-out mice have increased acinar-to-ductal metaplasia and impaired recovery after damage (Flandez et al. 2014). Decreased levels of Nr5a2 lead to an upregulation of inflammation promoting gene expression (Cobo et al. 2018). These findings strongly suggest that this transcription factor might also be crucial in the development of CP in humans. We analyzed 15 NR5A2 mutants regarding their transactivation potential on pancreas-specific target gene promoters.

While 13/15 mutations did not differ from wildtype on both the luciferase activity and the endogenous target gene mRNA expression, two variants, p.R142H and p.G150R, strongly reduced NR5A2 functions. In Western blot analysis, we were able to confirm protein synthesis of all NR5A2 mutants, thus could exclude that any effects in functional assays are based on the general lack of NR5A2.

A recent study on Nr5a2 haploinsufficiency (*Nr5a2*$^{+/-}$) suggests a so-called transcriptional switch, during which Nr5a2 interacts with activator protein 1 (AP-1) to occupy binding sites in inflammatory gene promoters (Cobo et al. 2018). ChIP-qPCR data showed reduced acinar gene expression while inflammatory gene expression was increased. This differential promoter occupancy by Nr5a2 was comparable to wildtype mice after caerulein treatment, indicating a basal inflammatory state (Cobo et al. 2018). Though we saw decreased levels of pancreatic enzymes (*CPA2*, *CELA3AB*), we failed to detect an upregulation of inflammatory markers (*C3*, *IL18*, *IL32*) in our mRNA expression assays. Consequently, we cannot confirm this switch. A different mechanism is thought to be the cause.

Both mutations affecting expression of target gene are located in the DNA binding domain of NR5A2 which is comprised of two zinc fingers [Lee et al. 2006; (Fayard et al. 2004) indicating an impaired DNA-binding capacity. In the panel of target genes regulated by NR5A2, we showed reduced mRNA expression levels of pancreatic enzymes CPA1, CELA3AB, and

PRSS1. Notably, a reduction in CELA3B is associated with CP (Tóth et al. 2022). A *CELA3B* splice-site variant in intron 2, c.129+1G>A, was significantly more common in German CP patients as well as in the in European replication cohort compared to controls. How reduced CELA3B function predisposes to pancreatitis, however, remains enigmatic. The same applies to *CELA3A*, in which a deletion of the entire gene is associated with the disease (unpublished data).

Subsequent EMSA experiments were conducted to investigate binding of mutant and wildtype NR5A2 on a 45 bp sequence in the *CPA2* and *CYP11B1* promoters. However, we missed to detect any differential effects which could be explained by falsely chosen oligonucleotides. EMSAs only show the transcription factor binding to an oligonucleotide which represents a very small fraction of the promoter sequence and thus, is quite error prone.

Furthermore, NR5A2 belongs to the family of nuclear receptors which are activated by ligand binding. Even though there are several suggestions such as phospholipids as potential ligands for NR5A2 (Krylova et al. 2005), the actual ligand binding is still enigmatic and might be an important piece of the puzzle of NR5A2 DNA-binding. Nuclear receptors that are activated by ligand binding and/or phosphorylation, undergo a conformational change, leading to the inhibition of co-repressor and the recruitment of co-activator complexes that influence the transcriptional activity (Fayard et al. 2004). A 45 bp sequence thus lacks regulatory regions which are necessary for co-factors of NR5A2 to bind.

In order to support our findings, *in vivo* research should be conducted. Our *in vitro* experiment in cell culture is a quite simplified model which is unable to reflect the natural conditions in CP patients. Mouse studies on Nr5a2 in the past have confirmed physiological effects. Heterozygous *Nr5a2*$^{+/-}$ mice already suffer from more severe AP with impaired damage recovery (Flandez et al. 2014). Further, *Nr5a2*-deficient mice lacked >90 % acini (Hale et al. 2014), with reduced concentrations of lipases and proteases (Holmstrom et al. 2011). In addition to acute models, chronic pancreatitis mouse models carrying the p.R142H or p.G150R variant could therefore give further insight. Even though mouse data do not mirror the human nature (Semaniakou et al. 2018), it still presents valuable data on a systemic impact of genetic variants on an organism as a whole.

7.3.2 Attenuating effects of RBPJL mutations are not associated with CP

The PTF1 complex is crucial to acinar gene regulation and cell identity. As part of the complex, RBPJL is responsible for high transcriptional activation. Thus, mutations are hypothesized to disrupt DNA binding affinity, PTF1 complex formation, protein stability or transcriptional activity (Nair et al. 2018). Findings in murine pancreatic acinar cells after transfection with siRNA targeting *Rbpjl* confirmed its regulation of digestive enzymes as it led to a significantly decreased mRNA expression of target genes like *Cel*, *Cpa1* and *Cpa2* (Nair

et al. 2018). Indeed, we were able to detect a reduction in transcriptional activity and endogenous target gene expression in five of the investigated *RBPJL* mutations. Targeting the *CPA1* promoter, relative luciferase activity was reduced by 50 %. Further, we measured a similar reduction of endogenous *CPA1* mRNA expression in Panc1.

RNA-seq studies in *Rbpjl* knock-out mice suggested zymogen granule proteins as targets for RBPJL in the PTF1 complex (Masui et al. 2010). Thus, we included *GP2* and *REG1A* in the endogenous mRNA expression assays. We confirmed a regulation by RBPJL wildtype with PTF1a. In comparison to the empty control, a distinguished mRNA expression level of both *GP2* and *REG1A* were detectable. *GP2* is a specific cell surface marker of human pancreatic progenitors that is co-expressed with PTF1a (Cogger et al. 2017). It is cleaved from the zymogen granule membrane to be secreted into the pancreatic juice, thus, might play a role in protein plug formation in chronic pancreatitis (Freedman et al. 1993). However, the *RBPJL* variants only showed minor reductive effects on the mRNA expression. Regulation by RBPJL thus cannot be confirmed with certainty. Further, we cannot rule out PTF1a as the driving factor besides RBPJL, as we only transfected both as complex.

Albeit we were able to demonstrate severe effects of RBPJL mutations on transcriptional activity in pancreatic genes, these findings cannot be associated with CP. Already the initial burden test of our WES data on the gnomAD database resulted in no significant prediction of deleterious mutations ($p = 0.296$; OR 1.39). Further, the Fisher Exact test verified that the diminishing effects of RBPJL are not significantly different between the patient and control cohorts. Consequently, we rule out *RBPJL* as risk gene and suggest a compensational mechanism for the lack of RBPJL. Studies in *Rbpjl* knock-out ($Rbpjl^{ko/ko}$) mice show that the replacement of *Rbpj* by *Rbpjl* in the PTF1 complex occurs in the final stages of acinar differentiation (Masui et al. 2010). Although the pancreas in $Rbpjl^{ko/ko}$ mice is smaller than in wildtype mice, it develops normally and shows no functional differences regarding islet tissue organization and hormone levels. Consequences of *Rbpjl* inactivation are restricted to the exocrine compartment only. RNA-seq data revealed decreased levels of 20 different digestive enzymes between 50 and 98 % (Masui et al. 2010). Nevertheless, Masui and colleagues suggest with ChIP data a partly compensation of *Rbpjl* by *Rbpj* in the PTF1 complex. Whereas *Rbpj* was not detected on selected promoters of secretory enzymes in wildtype or heterozygous ($Rbpjl^{+/ko}$) mice, it was present in homozygous knock-out mice.

Rbpjl in the adult pancreas harbors its transcriptional action nearly exclusively through the PTF1 complex (Masui et al. 2010). Transcriptome sequencing and chromatin immunoprecipitation sequencing demonstrated that PTF1a plays an important role in the expression of genes associated with the production of secretory digestive enzymes by direct gene targeting or indirect regulation through other transcription factors like RBPJL (Hoang et al. 2016; Jakubison et al. 2018). *PTF1a* knock-out mice showed an impaired acinar cell

homeostasis and differentiation which is directly connected to the pathology of pancreatitis and pancreatic adenocarcinoma (Hoang et al. 2016). Thus, mutations found in *PTF1a* may also be relevant to investigate in the future.

7.3.3 HNF1a variants do not alter the transactivation of target promoters

HNF1a is a hepatocyte nuclear factor known to be expressed in the liver and the pancreas to regulate the growth and function of islet β cells. Already in 1996, *HNF1a* was identified as the most common risk gene for MODY3 (Yamagata et al. 1996). Further expression studies in knock-out mice showed its role in the maintenance of normal β cell function, as mutations lead to a defective insulin secretion, fatty liver, and renal dysfunction (Pontoglio et al. 1998). Apart from the endocrine compartment, HNF1a also regulates differentiated cell phenotypes in acinar cells (Molero et al. 2012). Further, in a GWAS, *HNF1a* was identified as a risk gene for pancreatic cancer (Petersen et al. 2010; Pierce and Ahsan 2011). Thus, hypothetically genetic variants in *HNF1a* might also contribute to a CP phenotype.

Boulling *et al.* found a synergistic effect between PTF1a, RBPJL and HNF1a on the expression of *SPINK1* (Boulling et al. 2017), which was already confirmed in our research group by Dr. Lara Unger during her PhD project with the induction of *SPINK1* mRNA expression in Panc1 (data not published). We conducted reporter gene assays to investigate the activation of the *SPINK1* promoter and demonstrated a ~28-fold increased transactivation in combination with PTF1a and RBPJL. As an endocrine target gene, we additionally included *MTTP* (microsomal triglyceride transfer protein). Dysfunctional binding of HNF1a to the *MTTP* promoter suggests an affected insulin responsiveness (Au et al. 2008). Indeed, we could confirm a regulation of *MTTP* by HNF1a (~16-fold). Unfortunately, we were unable to detect any effects on the transactivation by HNF1a variants compared to the wildtype.

We first questioned the experimental design. To assure a successful luciferase assay, we included known disease-causing *HNF1a* variants as well as luciferase vectors published by Anders Molven and colleagues (Najmi et al. 2017; Althari et al. 2020). We additionally aimed to increase transactivation by cloning promoter regions with repeated binding sites for HNF1a as demonstrated in a previous study (Chi et al. 2002). The inclusion of the MODY3 variants p.P112L and p.P447L proves the experimental design to be correct, as they led to an abolished or strongly decreased relative luciferase activity.

There is also the strong possibility, that the presence of a co-factor would have further increased the relative luciferase activity, leading to a detectable relative reduction. HNF1a might not fully develop its transactivational function when the mutant cannot interact with its co-factor. For DNA-binding, HNF1a forms a homo- or heterodimer with other factors such as the structurally related HNF1β transcription factor (Mendel et al. 1991). As HNF1a is highly expressed in hepatocytes, it might be obvious to select a liver cell line such as HepG2 as a

more appropriate *in vitro* model. However, one should keep in mind that HEK293 was chosen as cell line, because it lacks endogenous expression of HNF1a [https://www.proteinatlas.org, (Uhlén et al. 2015). Endogenous expressed HNF1a might interfere with the measured effects of overexpressed HNF1a in the dual luciferase assay.

Due to the fact, that merely the positive control mutants of HNF1a reduced activity while the five mutations from the patient cohort left transactivation unchanged, we conclude these variants to be innocuous.

7.4 Chronic pancreatitis as a multigenic disease

Though our data on *CTRL* and *RBPJL* showed strong effects on their functionality, no direct association with CP can be concluded. We failed to explain a pancreatitis like phenotype, as loss-of-function mutations such as *CTRL* p.L218Rfs*33 were also found in the control cohort. A study on *CPA2* variants presented similar results, as none of the mutations in *CPA2* – not even the loss-of-function variants – seem to be disease-causing (Nakano et al. 2015). We have to keep in mind that CP is a complex multigenic disease and a single mutation might not be sufficient to cause a pancreatitis phenotype. Affected individuals often carry additional mutations in several disease-associated genes (Rosendahl et al. 2013). For example, in 901 German ICP subjects harboring one *CTRC* alteration, 5.8 % also carried a heterozygous *PRSS1* variant and 15.3 % also showed the *SPINK1* p.N34S mutations (121 heterozygous, 17 homozygous) (Rosendahl et al. 2008). Further, 6 out of 30 (20 %) patients with a functionally defective *TRPV6* variant were also heterozygous for *SPINK1* p.N34S (Masamune et al. 2020).

7.5 Outlook: the mystery of genetic risk factors in CP remains

In our project, we investigated five genes as focus for different potential pathological pathways for CP. While an association of *AQP12*, *CTRL*, *RBPJL*, and *HNF1a* could be excluded, we were able to identify *NR5A2* as potential novel genetic risk gene. Still, associated variants in this gene account for only a small percentage of pancreatitis cases, as merely 4 out of 1,059 patients (0.4 %) carry an effective *NR5A2* mutation. Further research is needed in order to support our findings, predominantly *in vivo* experiments in a chronic pancreatitis mouse model can give us further insight on the systemic consequences of certain genetic variants.

Moreover, a reliable control cohort comparable to the gene pool of our patient cohort is preferable rather than the European non-finish gnomAD database for implicating prediction tools on deleterious variants. Thus, additional bioinformatics calculations are needed to develop a good basis for experimental investigation.

As a final thought, we should be aware of the existing possibility that ICP cases might not necessarily underlie yet undiscovered genetic risk factors. Rather mutations in regulatory non-coding regions of known risk factors, which have not been studied yet, should be focused on. In the past, GWAS data on ACP already identified non-coding risk loci (Rosendahl et al. 2018), which also confirmed intronic alterations in high linkage disequilibrium with *SPINK1* p.N34S. The sequencing of coding regions only includes flanking splice site and core promotor regions at most. Thus, whole genome sequencing data should identify noncoding mutations predicted to be disease-causing as another approach to decipher the ongoing mystery of ICP cases.

8 References

Althari S, Najmi LA, Bennett AJ, Aukrust I, Rundle JK, Colclough K, et al. Unsupervised Clustering of Missense Variants in HNF1A Using Multidimensional Functional Data Aids Clinical Interpretation. The American Journal of Human Genetics. 2020;107(4):670–82. doi:10.1016/j.ajhg.2020.08.016.

Amundadottir LT. Pancreatic Cancer Genetics. Int J Biol Sci. 2016;12(3):314–25. doi:10.7150/ijbs.15001.

Annicotte J-S, Fayard E, Swift GH, Selander L, Edlund H, Tanaka T, et al. Pancreatic-duodenal homeobox 1 regulates expression of liver receptor homolog 1 during pancreas development. Mol Cell Biol. 2003;23(19):6713–24. doi:10.1128/mcb.23.19.6713-6724.2003.

Arvaniti E, Vakrakou A, Kaltezioti V, Stergiopoulos A, Prakoura N, Politis PK, Charonis A. Nuclear receptor NR5A2 is involved in the calreticulin gene regulation during renal fibrosis. Biochim Biophys Acta. 2016;1862(9):1774–85. doi:10.1016/j.bbadis.2016.06.013.

Au W-S, Lu L, Yeung C-M, Liu C-C, Wong OG, Lai L, et al. Hepatocyte nuclear factor 1 binding element within the promoter of microsomal triglyceride transfer protein (MTTP) gene is crucial for MTTP basal expression and insulin responsiveness. J Mol Endocrinol. 2008;41(4):229–38. doi:10.1677/JME-08-0080.

Balázs A, Hegyi P, Sahin-Tóth M. Pathogenic cellular role of the p.L104P human cationic trypsinogen variant in chronic pancreatitis. Am J Physiol Gastrointest Liver Physiol. 2016;310(7):G477-G486. doi:10.1152/ajpgi.00444.2015.

Balázs A, Balla Z, Kui B, Maléth J, Rakonczay Z, Duerr J, et al. Ductal Mucus Obstruction and Reduced Fluid Secretion Are Early Defects in Chronic Pancreatitis. Front Physiol. 2018;9:632. doi:10.3389/fphys.2018.00632.

Barrio R, Bellanné-Chantelot C, Moreno JC, Morel V, Calle H, Alonso M, Mustieles C. Nine novel mutations in maturity-onset diabetes of the young (MODY) candidate genes in 22 Spanish families. J Clin Endocrinol Metab. 2002;87(6):2532–9. doi:10.1210/jcem.87.6.8530.

Beer S, Zhou J, Szabó A, Keiles S, Chandak GR, Witt H, Sahin-Tóth M. Comprehensive functional analysis of chymotrypsin C (CTRC) variants reveals distinct loss-of-function mechanisms associated with pancreatitis risk. Gut. 2013;62(11):1616–24. doi:10.1136/gutjnl-2012-303090.

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;26(1):117–30. doi:10.1128/MCB.26.1.117-130.2006.

Bjørkhaug L, Sagen JV, Thorsby P, Søvik O, Molven A, Njølstad PR. Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in Norway. J Clin Endocrinol Metab. 2003;88(2):920–31. doi:10.1210/jc.2002-020945.

Blaydon DC, Lind LK, Plagnol V, Linton KJ, Smith FJD, Wilson NJ, et al. Mutations in AQP5, encoding a water-channel protein, cause autosomal-dominant diffuse nonepidermolytic palmoplantar keratoderma. Am J Hum Genet. 2013;93(2):330–5. doi:10.1016/j.ajhg.2013.06.008.

Boulling A, Le Maréchal C, Trouvé P, Raguénès O, Chen J-M, Férec C. Functional analysis of pancreatitis-associated missense mutations in the pancreatic secretory trypsin inhibitor (SPINK1) gene. Eur J Hum Genet. 2007;15(9):936–42. doi:10.1038/sj.ejhg.5201873.

Boulling A, Witt H, Chandak GR, Masson E, Paliwal S, Bhaskar S, et al. Assessing the pathological relevance of SPINK1 promoter variants. European Journal of Human Genetics. 2011;19(10):1066–73. doi:10.1038/ejhg.2011.79.

Boulling A, Keiles S, Masson E, Chen J-M, Férec C. Functional analysis of eight missense mutations in the SPINK1 gene. Pancreas. 2012;41(2):329–30. doi:10.1097/MPA.0b013e3182277b83.

Boulling A, Masson E, Zou W-B, Paliwal S, Wu H, Issarapu P, et al. Identification of a functional enhancer variant within the chronic pancreatitis-associated SPINK1 c.101AG (p.Asn34Ser)-containing haplotype. Hum Mutat. 2017;38(8):1014–24. doi:10.1002/humu.23269.

Braakman I, Helenius J, Helenius A. Manipulating disulfide bond formation and protein folding in the endoplasmic reticulum. EMBO J. 1992;11(5):1717–22. doi:10.1002/j.1460-2075.1992.tb05223.x.

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;72:248–54. doi:10.1006/abio.1976.9999.

Burghardt B, Elkjaer M-L, Know T-H, Rácz GZ, Varga G, Steward MC, Nielsen S. Distribution of aquaporin water channels AQP1 and AQP5 in the ductal system of the human pancreas. Gut. 2003;52(7):1008–16. doi:10.1136/gut.52.7.1008.

Carrère J, Figarella C, Guy O, Thouvenot JP. Human pancreatic chymotrypsinogen A: a non-competitive enzyme immunoassay, and molecular forms in serum and amniotic fluid. Biochim Biophys Acta. 1986;883(1):46–53. doi:10.1016/0304-4165(86)90133-9.

Cassidy BM, Zino S, Fjeld K, Molven A, Lowe ME, Xiao X. Single nucleotide polymorphisms in CEL-HYB1 increase risk for chronic pancreatitis through proteotoxic misfolding. Hum Mutat. 2020;41(11):1967–78. doi:10.1002/humu.24105.

Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acid Research. 2022;50(D1):D165-D173. doi:10.1093/nar/gkab1113.

Chaudhary J, Skinner MK. Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells. Mol Endocrinol. 1999;13(5):774–86. doi:10.1210/mend.13.5.0271.

Chi Y-I, Frantz JD, Oh B-C, Hansen L, Dhe-Paganon S, Shoelson SE. Diabetes Mutations Delineate an Atypical POU Domain in HNF-1α. Molecular Cell. 2002;10(5):1129–37. doi:10.1016/S1097-2765(02)00704-9.

Cobo I, Martinelli P, Flández M, Bakiri L, Zhang M, Carrillo-de-Santa-Pau E, et al. Transcriptional regulation by NR5A2 links differentiation and inflammation in the pancreas. Nature. 2018;554(7693):533–7. doi:10.1038/nature25751.

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;9(6):2464–76. doi:10.1128/mcb.9.6.2464-2476.1989.

Cogger KF, Sinha A, Sarangi F, McGaugh EC, Saunders D, Dorrell C, et al. Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. Nat Commun. 2017;8(1):331. doi:10.1038/s41467-017-00561-0.

Cohn JA, Friedman KJ, Noone PG, Knowles MR, Silverman LM, Jowell PS. Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. The New England Journal of Medicine. 1998;339(10):653–8. doi:10.1056/NEJM199809033391002.

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

Edlund H. Pancreatic organogenesis - developmental mechanisms and implications for therapy. Nat Rev Genet. 2002;3(7):524–32. doi:10.1038/nrg841.

Ehebauer M, Hayward P, Martinez-Arias A. Notch signaling pathway. Sci STKE. 2006;2006(364):cm7. doi:10.1126/stke.3642006cm7.

Etemad B, Whitcomb DC. Chronic pancreatitis: diagnosis, classification, and new genetic developments. Gastroenterology. 2001;120(3):682–707. doi:10.1053/gast.2001.22586.

Ewers M, Canaff L, Weh A, Masson E, Eiseler K, Chen J-M, et al. The three common polymorphisms p.A986S, p.R990G and p.Q1011E in the calcium sensing receptor (CASR) are not associated with chronic pancreatitis. Pancreatology. 2021;21(7):1299–304. doi:10.1016/j.pan.2021.08.008.

Fayard E, Auwerx J, Schoonjans K. LRH-1: an orphan nuclear receptor involved in development, metabolism and steroidogenesis. Trends Cell Biol. 2004;14(5):250–60. doi:10.1016/j.tcb.2004.03.008.

Felderbauer P, Hoffmann P, Einwächter H, Bulut K, Ansorge N, Schmitz F, Schmidt WE. A novel mutation of the calcium sensing receptor gene is associated with chronic pancreatitis in a family with heterozygous SPINK1 mutations. BMC Gastroenterol. 2003;3:34. doi:10.1186/1471-230X-3-34.

Fjeld K, Weiss FU, Lasher D, Rosendahl J, Chen J-M, Johansson BB, et al. A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis. Nat Genet. 2015;47(5):518–22. doi:10.1038/ng.3249.

Flandez M, Cendrowski J, Cañamero M, Salas A, Del Pozo N, Schoonjans K, Real FX. Nr5a2 heterozygosity sensitises to, and cooperates with, inflammation in KRas(G12V)-driven pancreatic tumourigenesis. Gut. 2014;63(4):647–55. doi:10.1136/gutjnl-2012-304381.

Freedman SD, Sakamoto K, Venu RP. GP2, the homologue to the renal cast protein uromodulin, is a major component of intraductal plugs in chronic pancreatitis. J Clin Invest. 1993;92(1):83–90. doi:10.1172/JCI116602.

Green J, Naot D, Cooper G. Hepatocyte nuclear factor 1 negatively regulates amylin gene expression. Biochem Biophys Res Commun. 2003;310(2):464–9. doi:10.1016/j.bbrc.2003.09.046.

Guo MH, Plummer L, Chan Y-M, Hirschhorn JN, Lippincott MF. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. The American Journal of Human Genetics. 2018;103(4):522–34. doi:10.1016/j.ajhg.2018.08.016.

Hale MA, Swift GH, Hoang CQ, Deering TG, Masui T, Lee Y-K, et al. The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. Development. 2014;141(16):3123–33. doi:10.1242/dev.109405.

Hanley NA, Rainey WE, Wilson DI, Ball SG, Parker KL. Expression Profiles of SF-1, DAX1, and CYP17 in the Human Fetal Adrenal Gland: Potential Interactions in Gene Regulation. Molecular endocrinology. 2001;15(1):57–68. doi:10.1210/mend.15.1.0585.

Harries LW. Alternate mRNA processing of the hepatocyte nuclear factor genes and its role in monogenic diabetes. Expert Rev Endocrinol Metab. 2006;1(6):715–26. doi:10.1586/17446651.1.6.715.

Hegyi P, Wilschanski M, Muallem S, Lukacs GL, Sahin-Tóth M, Uc A, et al. CFTR: A New Horizon in the Pathomechanism and Treatment of Pancreatitis. Rev Physiol Biochem Pharmacol. 2016;170:37–66. doi:10.1007/112_2015_5002.

Hegyi E, Sahin-Tóth M. Genetic Risk in Chronic Pancreatitis: The Trypsin-Dependent Pathway. Digestive diseases and sciences. 2017;62(7):1692–701. doi:10.1007/s10620-017-4601-3.

Hegyi E, Sahin-Tóth M. Human CPA1 mutation causes digestive enzyme misfolding and chronic pancreatitis in mice. Gut. 2019;68(2):301–12. doi:10.1136/gutjnl-2018-315994.

Hoang CQ, Hale MA, Azevedo-Pouly AC, Elsässer HP, Deering TG, Willet SG, et al. Transcriptional Maintenance of Pancreatic Acinar Identity, Differentiation, and Homeostasis by PTF1A. Mol Cell Biol. 2016;36(24):3033–47. doi:10.1128/MCB.00358-16.

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;25(16):1674–9. doi:10.1101/gad.16860911.

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics. 2016;99(4):877–85. doi:10.1016/j.ajhg.2016.08.016.

Ishibashi K, Tanaka Y, Morishita Y. The role of mammalian superaquaporins inside the cell. Biochim Biophys Acta. 2014;1840(5):1507–12. doi:10.1016/j.bbagen.2013.10.039.

Itoh T, Rai T, Kuwahara M, Ko SBH, Uchida S, Sasaki S, Ishibashi K. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. Biochem Biophys Res Commun. 2005;330(3):832–8. doi:10.1016/j.bbrc.2005.03.046.

Jakubison BL, Schweickert PG, Moser SE, Yang Y, Gao H, Scully K, et al. Induced PTF1a expression in pancreatic ductal adenocarcinoma cells activates acinar gene networks, reduces tumorigenic properties, and sensitizes cells to gemcitabine treatment. Mol Oncol. 2018;12(7):1104–24. doi:10.1002/1878-0261.12314.

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;291(25):12897–905. doi:10.1074/jbc.M116.725374.

Jin K, Xiang M. Transcription factor Ptf1a in development, diseases and reprogramming. Cell Mol Life Sci. 2019;76(5):921–40. doi:10.1007/s00018-018-2972-z.

Jung JS, Preston GM, Smith BL, Guggino WB, Agre P. Molecular structure of the water channel through aquaporin CHIP. The hourglass model. J Biol Chem. 1994;269(20):14648–54.

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;30(4):575–82. doi:10.1002/humu.20853.

Kim Y, Han D, Min H, Jin J, Yi EC, Kim Y. Comparative proteomic profiling of pancreatic ductal adenocarcinoma cell lines. Mol Cells. 2014;37(12):888–98. doi:10.14348/molcells.2014.0207.

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

Kopan R, Ilagan MXG. The canonical Notch signaling pathway: unfolding the activation mechanism. Cell. 2009;137(2):216–33. doi:10.1016/j.cell.2009.03.045.

Krapp A, Knöfler M, Ledermann B, Bürki K, Berney C, Zoerkler N, et al. 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;12(23):3752–63. doi:10.1101/gad.12.23.3752.

Krüger B, Albrecht E, Lerch MM. The Role of Intracellular Calcium Signaling in Premature Protease Activation and the Onset of Pancreatitis. American journal of pathology. 2000;157(1):43–50. doi:10.1016/S0002-9440(10)64515-4.

Krylova IN, Sablin EP, Moore J, Xu RX, Waitt GM, MacKay JA, et al. Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. Cell. 2005;120(3):343–55. doi:10.1016/j.cell.2005.01.024.

Larsen F, Solheim J, Kristensen T, Kolstø A-B, Prydz H. A tight cluster of five unrelated human genes on chromosome 16q22.1. Hum Mol Genet. 1993;2(10):1589–95. doi:10.1093/hmg/2.10.1589.

Lasher D, Szabó A, Masamune A, Chen J-M, Xiao X, Whitcomb DC, et al. Protease-Sensitive Pancreatic Lipase (PNLIP) Variants Are Associated With Early Onset Chronic Pancreatitis. American Journal of Gastroenterology. 2019;114(6):974–83. doi:10.14309/ajg.0000000000000051.

Lee Y-K, Schmidt DR, Cummins CL, Choi M, Peng L, Zhang Y, et al. Liver receptor homolog-1 regulates bile acid homeostasis but is not essential for feedback regulation of bile acid synthesis. Mol Endocrinol. 2008;22(6):1345–56. doi:10.1210/me.2007-0565.

Lee MG, Ohana E, Park HW, Yang D, Muallem S. Molecular Mechanism of Pancreatic and Salivary Glands Fluid and HCO3- Secretion. Physiol Rev. 2012;92(1):39–74. doi:10.1152/physrev.00011.2011.

Li D, Duell EJ, Yu K, Risch HA, Olson SH, Kooperberg C, et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. Carcinogenesis. 2012;33(7):1384–90. doi:10.1093/carcin/bgs151.

Lieber M, Mazzetta J, Nelson-Rees W, Kaplan M, Todaro G. Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. Int J Cancer. 1975;15(5):741–7. doi:10.1002/ijc.2910150505.

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;25(4):402–8. doi:10.1006/meth.2001.1262.

Lowenfels AB, Maisonneuve P, Cavallini G, Ammann RW, Lankisch PG, Andersen JR, et al. Pancreatitis and the Risk of Pancreatic Cancer. The New England Journal of Medicine. 1993;328(20):1433–7.

MacDonald RJ, Swift GH, Real FX. Transcriptional Control of Acinar Development and Homeostasis. Prog Mol Biol Transl Sci. 2010;97:1–40. doi:10.1016/S1877-1173(10)97001-4.

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50(W1):W276-279. doi:10.1093/nar/gkac240.

Masamune A, Kotani H, Sörgel FL, Chen J-M, Hamada S, Sakaguchi R, et al. Variants That Affect Function of Calcium Channel TRPV6 Are Associated With Early-Onset Chronic Pancreatitis. Gastroenterology. 2020;158(6):1626–1641. doi:10.1053/j.gastro.2020.01.005.

Masui T, Swift GH, Deering T, Shen C, Coats WS, Long Q, et al. Replacement of Rbpj with Rbpjl in the PTF1 complex controls the final maturation of pancreatic acinar cells. Gastroenterology. 2010;139(1):270–80. doi:10.1053/j.gastro.2010.04.003.

Mendel DB, Hansen LP, Graves MK, Conley PB, Crabtree GR. HNF-1 alpha and HNF-1 beta (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. Genes Dev. 1991;5(6):1042–56. doi:10.1101/gad.5.6.1042.

Méndez-Giménez L, Ezquerro S, da Silva IV, Soveral G, Frühbeck G, Rodríguez A. Pancreatic Aquaporin-7: A Novel Target for Anti-diabetic Drugs? Frontiers in Chemistry. 2018;6(99):1–10. doi:10.3389/fchem.2018.00099.

Molero X, Vaquero EC, Flández M, González AM, Ortiz MÁ, Cibrián-Uhalte E, et al. Gene expression dynamics after murine pancreatitis unveils novel roles for Hnf1α in acinar cell homeostasis. Gut. 2012;61(8):1187–96. doi:10.1136/gutjnl-2011-300360.

Moore PC, Cortez JT, Chamberlain CE, Alba D, Berger AC, Quandt Z, et al. Elastase 3B mutation links to familial pancreatitis with diabetes and pancreatic adenocarcinoma. J Clin Invest. 2019;129(11):4676–81. doi:10.1172/JCI129961.

Mosztbacher D, Jancsó Z, Sahin-Tóth M. Loss of chymotrypsin-like protease (CTRL) alters intrapancreatic protease activation but not pancreatitis severity in mice. Sci Rep. 2020;10(1):11731. doi:10.1038/s41598-020-68616-9.

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, et al. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell. 1989;58(3):537–44. doi:10.1016/0092-8674(89)90434-0.

Nair AK, Sutherland JR, Traurig M, Piaggi P, Chen P, Kobes S, et al. Functional and association analysis of an Amerindian-derived population-specific p.(Thr280Met) variant in RBPJL, a component of the PTF1 complex. Eur J Hum Genet. 2018;26(2):238–46. doi:10.1038/s41431-017-0062-6.

Najmi LA, Aukrust I, Flannick J, Molnes J, Burtt N, Molven A, et al. Functional Investigations of HNF1A Identify Rare Variants as Risk Factors for Type 2 Diabetes in the General Population. Diabetes. 2017;66(2):335–46. doi:10.2337/db16-0460.

Nakano E, Geisz A, Masamune A, Niihori T, Hamada S, Kume K, et al. Variants in pancreatic carboxypeptidase genes CPA2 and CPB1 are not associated with chronic pancreatitis. Am J Physiol Gastrointest Liver Physiol. 2015;309(8):G688-94. doi:10.1152/ajpgi.00241.2015.

Naqvi AAT, Hasan GM, Hassan MI. Investigating the role of transcription factors of pancreas development in pancreatic cancer. Pancreatology. 2017;18(2):184–90. doi:10.1016/j.pan.2017.12.013.

Ohta E, Itoh T, Nemoto T, Kumagai J, Ko SBH, Ishibashi K, et al. Pancreas-specific aquaporin 12 null mice showed increased susceptibility to caerulein-induced acute pancreatitis. Am J Physiol, Cell Physiol. 2009;297(6):C1368–C1378. doi:10.1152/ajpcell.00117.2009.

Okita K, Yang Q, Yamagata K, Hangenfeldt KA, Miyagawa J-i, Kajimoto Y, et al. Human Insulin Gene Is a Target Gene of Hepatocyte Nuclear Factor-1a (HNF-1alpha) and HNF-1beta. Biochem Biophys Res Commun. 1999;263(2):566–9. doi:10.1006/bbrc.1999.1412.

Ortlund EA, Lee Y, Solomon IH, Hager JM, Safi R, Choi Y, et al. Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. Nat Struct Mol Biol. 2005;12(4):357–63. doi:10.1038/nsmb910.

Oslowski CM, Urano F. Measuring ER stress and the unfolded protein response using mammalian tissue culture system. Meth Enzymol. 2011;490:71–92. doi:10.1016/B978-0-12-385114-7.00004-0.

Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010;42(3):224–8. doi:10.1038/ng.522.

Pierce BL, Ahsan H. Genome-wide "pleiotropy scan" identifies HNF1A region as a novel pancreatic cancer susceptibility locus. Cancer Res. 2011;71(13):4352–8. doi:10.1158/0008-5472.CAN-11-0124.

Polgár L. The catalytic triad of serine peptidases. Cell Mol Life Sci. 2005;62(19-20):2161–72. doi:10.1007/s00018-005-5160-x.

Pontoglio M, Sreenan S, Roe M, Pugh W, Ostrega D, Doyen A, et al. Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice. J Clin Invest. 1998;101(10):2215–22. doi:10.1172/JCI2548.

Preston GM, Carroll TP, Guggino WB, Agre P. Appearance of water channels in Xenopus oocytes expressing red cell CHIP28 protein. Science. 1992;256(5055):385–7. doi:10.1126/science.256.5055.385.

Preston GM, Smith BL, Zeidel ML, Moulds JJ, Agre P. Mutations in aquaporin-1 in phenotypically normal humans without functional CHIP water channels. Science. 1994;265(5178):1585–7. doi:10.1126/science.7521540.

Rausa FM, Galarneau L, Bélanger L, Costa RH. The nuclear receptor fetoprotein transcription factor is coexpressed with its target gene HNF-3beta in the developing murine liver, intestine and pancreas. Mech Dev. 1999;89(1-2):185–8. doi:10.1016/s0925-4773(99)00209-9.

Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-Splice - improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Med. 2021;13(1):31. doi:10.1186/s13073-021-00835-9.

Reseland JE, Larsen F, Solheim J, Eriksen JA, Hanssen LE, Prydz H. A novel human chymotrypsin-like digestive enzyme. J Biol Chem. 1997;272(12):8099–104. doi:10.1074/jbc.272.12.8099.

Rinderknecht H. Activation of Pancreatic Zymogens: Normal Activation, Premature Intrapancreatic Activation, Protective Mechanisms Against Inappropriate Activation. Digestive diseases and sciences. 1986;31(3):314–21. doi:10.1007/BF01318124.

Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989;245(4922):1066–73. doi:10.1126/science.2475911.

Robertson SC, Meyer AN, Hart KC, Galvin BD, Webster MK, Donoghue DJ. Activating mutations in the extracellular domain of the fibroblast growth factor receptor 2 function by disruption of the disulfide bond in the third immunoglobulin-like domain. Proceedings of the National Academy of Sciences. 1998;95(8):4567–72. doi:10.1073/pnas.95.8.4567.

Rosendahl J, Witt H. Pathogenese der chronischen Pankreatitis. Internist. 2021;62(10):1007–14. doi:10.1007/s00108-021-01150-6.

Rosendahl J, Witt H, Szmola R, Bhatia E, Ózsvári B, Landt O, et al. Chymotrypsin C (CTRC) alterations that diminish activity or secretion are associated with chronic pancreatitis. Nat Genet. 2008;40(1):78–82. doi:10.1038/ng.2007.44.

Rosendahl J, Landt O, Bernadova J, Kovacs P, Teich N, Bödeker H, et al. CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated? Gut. 2013;62(4):582–92. doi:10.1136/gutjnl-2011-300645.

Rosendahl J, Kirsten H, Hegyi E, Kovacs P, Weiss FU, Laumen H, et al. Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. Gut. 2018;67(10):1855–63. doi:10.1136/gutjnl-2017-314454.

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;3(10):1613–24. doi:10.1101/gad.3.10.1613.

Rygiel AM, Unger LS, Sörgel FL, Masson E, Matsumoto R, Ewers M, et al. Variants in the pancreatic CUB and zona pellucida-like domains 1 (CUZD1) gene in early-onset chronic pancreatitis - A possible new susceptibility gene. Pancreatology. 2022;22(5):564–71. doi:10.1016/j.pan.2022.04.015.

Sahin-Tóth M. Genetic risk in chronic pancreatitis: the misfolding-dependent pathway. Current opinion in gastroenterology. 2017;33(5):390-395. doi:10.1097/MOG.0000000000000380.

Sankaran SJ, Xiao AY, Wu LM, Windsor JA, Forsmark CE, Petrov MS. Frequency of progression from acute to chronic pancreatitis and risk factors: a meta-analysis. Gastroenterology. 2015;149(6):1490-1500.e1. doi:10.1053/j.gastro.2015.07.066.

Schneider A, Löhr JM, Singer MV. The M-ANNHEIM classification of chronic pancreatitis: introduction of a unifying classification system based on a review of previous classifications of the disease. J Gastroenterol. 2007;42(2):101–19. doi:10.1007/s00535-006-1945-4.

Semaniakou A, Croll RP, Chappe V. Animal Models in the Pathophysiology of Cystic Fibrosis. Front Pharmacol. 2018;9:1475. doi:10.3389/fphar.2018.01475.

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. The New England Journal of Medicine. 1998;339(10):645–52.

Shelton CA, Whitcomb DC. Genetics and Treatment Options for Recurrent Acute and Chronic Pancreatitis. Curr Treat Options Gastroenterol. 2014;12(3):359–71. doi:10.1007/s11938-014-0022-y.

Sirianni R, Seely JB, Attia G, Stocco DM, Carr BR, Pezzi V, Rainey WE. Liver receptor homologue-1 is expressed in human steroidogenic tissues and activates transcription of genes encoding steroidogenic enzymes. Journal of Endocrinology. 2002;174(3):R13-R17. doi:10.1677/joe.0.174r013.

Szabó A, Sahin-Tóth M. Determinants of chymotrypsin C cleavage specificity in the calcium-binding loop of human cationic trypsinogen. FEBS J. 2012;279(23):4283–92. doi:10.1111/febs.12018.

Szmola R, Sahin-Tóth M. Chymotrypsin C (caldecrin) promotes degradation of human cationic trypsin: Identity with Rinderknecht's enzyme Y. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(27):11227–32. doi:10.1073/pnas.0703714104.

Tang X-Y, Zou W-B, Masson E, Hu L-H, Ferec C, Chen J-M, et al. The CTRB1-CTRB2 risk allele for chronic pancreatitis discovered in European populations does not contribute to disease risk variation in the Chinese population due to near allele fixation. Gut. 2018;67(7):1368–9. doi:10.1136/gutjnl-2017-315180.

Tata F, Stanier P, Wicking C, Halford S, Kruyer H, Lench NJ, et al. Cloning the mouse homolog of the human cystic fibrosis transmembrane conductance regulator gene. Genomics. 1991;10(2):301–7. doi:10.1016/0888-7543(91)90312-3.

Thévenod F. Ion channels in secretory granules of the pancreas and their role in exocytosis and release of secretory proteins. Am J Physiol, Cell Physiol. 2002;283(3):C651-72. doi:10.1152/ajpcell.00600.2001.

Tóth A, Demcsák A, Zankl F, Oracz G, Unger LS, Bugert P, et al. Loss-of-function variant in chymotrypsin like elastase 3B (CELA3B) is associated with non-alcoholic chronic pancreatitis. Pancreatology. 2022;22(6):713–8. doi:10.1016/j.pan.2022.06.258.

Tugwood J, Issemann I, Green S. LRH-1: a nuclear hormone receptor active in the absence of exogenous ligands. Gene bank accession number M81385 in the activation of fushi tarazu. Science. 1991;252:848–51.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419. doi:10.1126/science.1260419.

Venglovecz V, Pallagi P, Kemény LV, Balázs A, Balla Z, Becskeházi E, et al. The Importance of Aquaporin 1 in Pancreatitis and Its Relation to the CFTR Cl- Channel. Front Physiol. 2018;9:854. doi:10.3389/fphys.2018.00854.

Wang ZN, Bassett M, Rainey WE. Liver receptor homologue-1 is expressed in the adrenal and can regulate transcription of 11b-hydroxylase. J Mol Endocrinol. 2001;27(2):255–8. doi:10.1677/jme.0.0270255.

Wartenberg P, Lux F, Busch K, Fecher-Trost C, Flockerzi V, Krasteva-Christ G, et al. A TRPV6 expression atlas for the mouse. Cell Calcium. 2021;100:102481. doi:10.1016/j.ceca.2021.102481.

Watanabe M, Sheriff S, Lewis KB, Cho J, Tinch SL, Balasubramaniam A, Kennedy MA. Metabolic Profiling Comparison of Human Pancreatic Ductal Epithelial Cells and Three Pancreatic Cancer Cell Lines using NMR Based Metabonomics. Journal of Molecular Biomarkers & Diagnosis. 2012;3(2):S3-002. doi:10.4172/2155-9929.s3-002.

Whitcomb DC, Lowe ME. Human Pancreatic Digestive Enzymes. Digestive diseases and sciences. 2007;52:1–17. doi:10.1007/s10620-006-9589-z.

Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nat Genet. 1996;14:141-145. doi:10.1038/ng1096-141.

Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, et al. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. Nat Genet. 2012;44(12):1349–54. doi:10.1038/ng.2466.

Wiebe PO, Kormish JD, Roper VT, Fujitani Y, Alston NI, Zaret KS, et al. Ptf1a binds to and activates area III, a highly conserved region of the Pdx1 promoter that mediates early pancreas-wide Pdx1 expression. Mol Cell Biol. 2007;27(11):4093–104. doi:10.1128/MCB.01978-06.

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

Witt H, Luck W, Hennies HC, Claßen M, Kage A, Laß U, et al. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet. 2000;25(2):213–6. doi:10.1038/76088.

Witt H, Sahin-Tóth M, Landt O, Chen J-M, Kähne T, Drenth JPH, et al. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. Nat Genet. 2006;38(6):668–73. doi:10.1038/ng1797.

Witt H, Apte MV, Keim V, Wilson JS. Chronic Pancreatitis: Challenges and Advances in Pathogenesis, Genetics, Diagnosis, and Therapy. Gastroenterology. 2007;132(4):1557–73. doi:10.1053/j.gastro.2007.03.001.

Witt H, Beer S, Rosendahl J, Chen J-M, Chandak GR, Masamune A, et al. Variants in CPA1 are strongly associated with early-onset chronic pancreatitis. Nat Genet. 2013;45(10):1216–20. doi:10.1038/ng.2730.

Xiao AY, Tan MLY, Wu LM, Asrani VM, Windsor JA, Yadav D, Petrov MS. Global incidence and mortality of pancreatic diseases: a systematic review, meta-analysis, and meta-regression of population-based cohort studies. The Lancet Gastroenterology & Hepatology. 2016;1(1):45–55. doi:10.1016/S2468-1253(16)30004-8.

Yadav D, Lowenfels AB. The Epidemiology of Pancreatitis and Pancreatic Cancer. Gastroenterology. 2013;144(6):1252–61. doi:10.1053/j.gastro.2013.01.068.

Yamagata K, Oda N, Kaisaki P, Menzel S, Furuta H, Vaxillaire M, et al. Mutations in the hepatocyte nuclear factor-1a gene in maturity-onset diabetes of the young (MODY3). Nature. 1996;384:455–8.

Yang B, Song Y, Zhao D, Verkman AS. Phenotype analysis of aquaporin-8 null mice. Am J Physiol, Cell Physiol. 2005;288(5):C1161-70. doi:10.1152/ajpcell.00564.2004.

Zhang C-K, Lin W, Cai Y-N, Xu P-L, Dong H, Li M, et al. Characterization of the genomic structure and tissue-specific promoter of the human nuclear receptor NR5A2 (hB1F) gene. Gene. 2001;273(2):239–49. doi:10.1016/S0378-1119(01)00586-8.

8.1 List of publications and manuscripts

Published

Eiseler, Katharina; Dropmann, Lea Marie; Bugert, Peter; Ewers, Maren; Witt, Heiko (2022): Genetic analysis of the aquaporin water channels *AQP12A* and *AQP12B* in patients with chronic pancreatitis. In: Pancreatology 22 (8), S. 1079-1083. DOI: 10.1016/j.pan.2022.09.240

Ewers, Maren; Canaff, Lucie; Weh, Antonia; Masson, Emmanuelle; **Eiseler, Katharina**; Chen, Jian-Min et al. (2021): The three common polymorphisms p.A986S, p.R990G and p.Q1011E in the calcium sensing receptor (CASR) are not associated with chronic pancreatitis. In: Pancreatology 21 (7), S. 1299-1304. DOI: 10.1016/j.pan.2021.08.008.

In preparation

Genetic and functional analysis of chymotrypsin-like protease (CTRL) in chronic pancreatitis. **Eiseler, Katharina**; Neppl, Lea; Schmidt, Andreas W.; Rauscher, Beate; Ewers, Maren; Masson, Emmanuelle; Chen, Jian-Min; Férec, Claude; Grammatikopoulos, Tassos; Greenhalf, William; Halloran, Christopher; Neoptolemos, John; Sturm, Marc; Rosendahl, Jonas; Laumen, Helmut; Witt, Heiko.

9 Supplementary Data

9.1 Primers

9.1.1 Primers for Sanger sequencing

For *AQP12*, we performed mutation screening by Sanger sequencing. For this, we designed primers for PCR and sequencing complementary to intronic sequences flanking the *AQP12A* and *AQP12B* exons based on the published nucleotide sequences (ENSG00000184945 and ENSG00000185176, respectively) (Table 9.1). Oligonucleotides were synthesized by TIB MOLBIOL (Berlin, Germany) in a GSF purification grade.

Table 9.1: Primers for *AQP12A* and *AQP12B* exon amplification and Sanger sequencing

Gene	Name	Sequence 5'→3'
AQP12B	1FA	CATAGGCAGGGACGCAGCCA
AQP12B	1FB	GTGCTTGGGTGCAGACCCAG
AQP12B	1FC**	GGTCCCACGTGCTGTACCCC
AQP12B	1FD	AGAGGGAAAGGGCTGTGAGG
AQP12A	1FA	CATAGGCAGGGACACAGCCC
AQP12A	1FB	GTGCTTGGGTGCAGACTCAC
AQP12A	1FC**	AGTCCCACATGCTGTACCAT
AQP12A	1FD	AGAAGGAAAGGGCTGCGGGC
AQP12AB	2RA**	ACCTGGCTGCAGCTTGCTCG
AQP12A	3RC**	GGCTGCAGGCACTGGTGGCA
AQP12B	3RC**	GGCTGCAGGCACTGATGGGG
AQP12AB	3FB**	CCCATGGCTGGCATACCCAG
AQP12AB	S11F*	TGGTGTTCCCAGAGGCCCAG
AQP12AB	S2F*	TCCCAGGCCTGCACACAGTG
AQP12AB	3FA*	CTGGGCCTCAGGGTTCTTGG

(*) primer used for Sanger sequencing; (**) primary used primer sequences for PCR

We ordered primers from TIB MOLBIOL for the verification of mutations found in WES regarding all other genes of interest, *CTRL* (GenBank # AC009077.7; NM_001907.2), *NR5A2* (NM_205860.2), *RBPJL* (NM_001281449.1), and *HNF1a* (NM_000545.6), in a GSF purification grade.

Table 9.2: Primers for mutation verification in *CTRL* with Sanger sequencing

Gene	Name	Sequence 5'→3'	Exon
CTRL	P1FA	CAAAGTTGACCGCAGTGAGC	1 - 3
CTRL	P3R	CTCCGTGTCTGCAGCCCAGG	1 - 3
CTRL	P4FA	CTTCCTGCCTCCTCCCCTGG	4 - 7
CTRL	P7RA	GTCTCTGGGGACCTGAAGCC	4 - 7
CTRL	S1F*	CTGTGATCCCAGCCACCAGG	1
CTRL	S23F*	GTGCTGTACCCTCTCGACAG	2, 3
CTRL	S45F*	TGCCTCCTCCCCTGGCTGTC	4, 5
CTRL	S67F*	TCAGGCCAAAGCTCAGGGTG	6, 7

(*) primer used for Sanger sequencing

Table 9.3: Primers for mutation verification in *NR5A2* with Sanger sequencing

Gene	Name	Sequence 5'→3'
NR5A2	1F	GTGCTGCATAGCACCTCTCC
NR5A2	1R	GGGTATTAGAATGACACCAG
NR5A2	S1F*	GAGTCATGTGATGAGTCAAAGCC
NR5A2	2F	TAACCCTGACCTCCTCCTCG
NR5A2	2R	TTAGCCTGAAGCTGCCCTGC
NR5A2	S2F*	ACAGGGCTCAGAGGTCCTGC
NR5A2	4F	GGTAGCCAAATCTCCCCAGG
NR5A2	4R	GTGTCTAATGCATCTGCTCCCC
NR5A2	S4F*	CCTGGCACAGTTTGGGAAGG
NR5A2	5F	CGACACACACTTGGAGGAAGG
NR5A2	5R	GCTGCCACAGTCCTGAAGTCAAGG
NR5A2	S5F*	CACACATACCACTTCTTGTCG
NR5A2	7F	GTGCCTCAAACTACTTGCATGGGG
NR5A2	7R	CCATTCCCCAACGTATCCACTG
NR5A2	S7F*	GTGAGACCCCACATCTCTATTC

(*) primer used for Sanger sequencing

Table 9.4: Primers for mutation verification in *RBPJL* with Sanger sequencing

Gene	Name	Sequence 5'→3'
RBPJL	23F	GGTCCAGAGAGGGTAGTGTC
RBPJL	23R	AGCAGTACCCACAGAGCCTG
RBPJL	S2F*	GGAAAGGACCCTGCAGAAGTGG
RBPJL	S3F*	GGCTTCGTTGAGCACCTTGC
RBPJL	7F	AAGCAGGACTGCAGCAGGTC
RBPJL	7R	TGGAGAGGCCATTTGAGGCC
RBPJL	S7F*	GGACAGGGCACTGATAGCAC
RBPJL	10F	TCTAACGCAGAAGCCCACGC
RBPJL	12R	TCCCAAACTTGGCTCCTCGG
RBPJL	S12R*	AAGCAGGGCCTAGAACCCAG

(*) primer used for Sanger sequencing

Table 9.5: Primers for mutation verification in *HNF1a* with Sanger sequencing

Gene	Name	Sequence 5'→3'
HNF1a	2F	GTATCCATAGGCCTGTGTCC
HNF1a	2R	GCAGGGACCCATTTCATTCATC
HNF1a	S2F*	GTTGACAAGGTTCCAGCACC
HNF1a	34F	GCTGGGTCTCTTCCTCTAGC
HNF1a	34R	AGGTGACTGCTGTCACTGGG
HNF1a	S3F*	TCCTCTGGTTCAGCGCCATG
HNF1a	S4F*	GGCAGAGCTCAGCTTCTCAG
HNF1a	89F	CGCTTAGTACGTAGCACCCG
HNF1a	89R	GTTCAACCCTGAGCCAAGCC
HNF1a	S8F*	CCAGTCTGGCTGTTCAGCAG
HNF1a	S9R*	CCACAGTGACGGACAGCAAC

(*) primer used for Sanger sequencing

9.1.2 Primers for molecular cloning

We designed primers manually and verified their properties and binding capacity using the SnapGene Viewer Software. Primers were then ordered from TIB MOLBIOL. All primers used for cloning were purified by HPLC, while those primers solely used for control sequencing were purified by GSF.

Table 9.6: Primers used for molecular cloning of target gene promoters

Gene	Name	Sequence 5'→3'
CPA2	F	GAGTGAGGCTCGAGATGGGCCACATCTACAGGGACACTC
CPA2	Rb	TGGCCAAAGCTTTTCCTGGAGTCCAGAATTCAGAGTCC
CEL	F	CTGAGCCTCGAGTCCTCCCTCCCAGGACACATCAGG
CEL	R	CATGGTAAGCTTCAGCCTCTGGGTGGCCCTCCC
CYP11B1	F	GGTATAACTCGAGTGTGAATCCATCTGGTCATGGACTC
CYP11B1	R	GCCATTCCAAAGCTTCCTCCACCCTGTTCAGCTGCAATCC
CYP17	F	AGAATCCTCGAGCACCTCGGATTGCTTCTTTACGG
CYP17	R	GGCAGGCAAGCTTGACAGCAGTGGAGTAGAAGAGCTGTG
CALR	F	GGAAGCCTCGAGCCCACAGGTGCCTGACACAGGAACAG
CALR	R	GGGCCGAAGCTTCGGCAACGCGCGGGCCCTTTAAAAC
SPINK1	Fa	GTCACTTGCTAGCCTCAACAAAAAGGCAATGGAAGCTCC
SPINK1	Ra	TGTTACCTCGAGGGCTGAAGTTCTGCGTCCAGAGGTC
CELA3A	F	CAGACACTCGAGAGTGGAATCTGGGCTTATCGGACC
CELA3A	Ra	CGGAGCAAGCTTAGTTTTGTGATGATAGGAAAAGGGCAC
REG1A	Fb	ATGACAGGCGCTAGCCACCACACCAGGCCCATCATC
REG1A	Ra	AGCTGGCTCGAGCCATGCTGAGCTGCAATGAATCTC
GP2	F *NheI* mut	GTTGTAGCTAGCTGGTTCAGATTGGAATGGAGAAGCC
GP2	R *XhoI* mut	TGAGGCCTCGAGGTCACTTTGCTGTATGCAGACTTCCC
PRSS2	Fa *NheI* mut	CAGCCTGCTAGCGATTAGAGAGGGTACCGGGGAGAGAGGG
PRSS2	R *XhoI* mut	AGTAGATTCCTCGAGGTAGAGTGTGCCTGATTGCTGGTGG
INS	Fa *NheI* mut	AAGGTGGCTAGCCACGTGACAGTGGGAGGGCCTCTCTCAG
INS	R HindIII mut	AGGGCCAAGCTTGAAGGACAGTGATCTGGGAGACAGGCAG
CPA1	F *NheI* mut	CCTGGCGCTAGCCTATCATAAGACCTAAGTTAAGCTGGGC
CPA1	Ra *HindIII* mut	CGCATGAAGCTTCCGGGAGGGAAGGTCAGACTGAGGTCG
IAPP	F *NheI* mut	GAACTTGCTAGCATGTCACTGGCCCTGGTATGCTTTCTAC
IAPP	Ra *XhoI* mut	CTTGCAGCTCGAGGATGCCCATTGCTTCTCAAATTTTCTGC
MTTP	F *NheI* mut	CGAATAGCTAGCTACCCGAAGGGCACGTATCATGCACCACATAGCC
MTTP	R *XhoI* mut	GAAGAATCTCGAGGACCAGCAATCCTCAACTGCATCCAGTGCCCAGC

Table 9.7: Primers used for control sequencing and/or colony PCR of cloned vectors

Gene	Name	Sequence 5'→3'
-	RVprimer3	CTAGCAAAATAGGCTGTCCC
-	Luci-R-Seq	CTTCGAGTGGGTAGAATGGCG
CYP17	Fa Seq	ATTGTGCCTGGAGCTGAGGC
CYP17	Fb Seq	GGTTCTCTGCTGTCGAACCC
CYP17	Ra Seq	CCAACAGTACCGGATTGCCAAGC
StAR	R Seq	TGGCTTACGCCTGTAACCCC
PRSS1	R Seq	ACAACCTGGCCAGAGCAGTG

Gene	Name	Sequence 5'→3'
CELA3A	R Seq	GGGTTCAAGCGATTCTCCTGCC
REG1A	R Seq	GACCTACCTCAAGGGGTAGG
GP2	R Seq	CAAGCCACCTTCTCTACTCCC
GP2	F Seq	GCCTCTTCTCAGTCTCATGCC
GP2	Fa Seq	GCACAATGCTGAGACAGCTGG
GP2	Fb Seq	GGTTAAGAGTGGCCACTGGG
GP2	Ra Seq	CTGAAGGCCACTGAGCTTGG
INS	R Seq	AGCTATCTGCAGGACCCCTC
INS	F Seq	CACACTCCTCTCAAGGAGGC
INS	Fa Seq	AAGGCAGGGTGGGAACTAGG
CPA1	F Seq	AGCCTCTAGATGGTGCTGGG
CPA1	R Seq	GTGAGGACTGCACAGTGTCC
IAPP	R Seq	GAGGTAGGAGGCAGGACTAG
IAPP	F Seq	CTAGGGGGCGTGTCTTATGG
IAPP	Fa Seq	AGTACTGTGTGAGCTAAGCC
MTTP	F Seq	ATTCCCACAGGCGGCCTAAG
MTTP	R Seq	CCTTGTCCCTGGAAACTCCC

Table 9.8: Primers used for molecular cloning of a triple promoter fragment

Gene	Name	Sequence 5'→3'
ALB	Famp	CCAGATGGTAAATATACACAAGGG
ALB	F1 *NheI* mut	CCAGATGGTAAATATACACAAGGGATTTAGTGCTAGCATTTTTTGGCAAGAATATTATG
ALB	R1 *XhoI* mut	ACCCACCTCGAGGTGCCAAAGGCGTGTGGGGTTGACAGAAGAG
ALB	F2 *XhoI* mut	GGGATTTAGTCTCGAGATTTTTTGGCAAGAATATTATGAATTTTGTAATCGG
ALB	R2 *BglII* mut	ACCCACAGATCTGTGCCAAAGGCGTGTGGGGTTGACAGAAG
ALB	F3 *BglII* mut	GGGATTTAGTAGATCTATTTTTTGGCAAGAATATTATGAATTTTGTAATCGG
ALB	R3 *HindIII* mut	GAAGGGAAATAAAGGTTACCCACAAGCTTGTGCCAAAGGCGTGTGGGGTTGAC
ALB	Ramp	AGGGAAATAAAGGTTACC
FGB	Famp	GCTTTGCTGGGAAGATGTTG
FGB	F1 *NheI* mut	GTAAATAAGCTTTGCTGGGAAGATGTTGCTTGCTAGCTAAAATGGTTCAGCCAACAAGTGAACC
FGB	R1 *XhoI* mut	CAATCCTATCTCGAGCTGAGTCCTCTGCTAGG
FGB	F2 *XhoI* mut	TTGCTTCTCGAGTAAAATGGTTCAGCCAACAAGTGAACC
FGB	R2 *ApaI* mut	CAATCCTATGGGCCCCTGAGTCCTCTGCTAGG
FGB	F3 *ApaI* mut	TTGCTTGGGCCCTAAAATGGTTCAGCCAACAAGTGAACC
FGB	R3 *HindIII* mut	GTAGACTTAACTGAGAGATCTTCAATCCTATAAGCTTCTGAGTCCTCTGCTAGG
FGB	Ramp	CTGAGAGATCTTCAATCC

9.1.3 PCR-mutagenesis primers

Primers were designed manually, and their properties and binding capacity were verified using the SnapGene Viewer Software. Primers were synthesized by TIB MOLBIOL and purified by HPLC. The primers pcDNA3.1-2F and pcDNA3.1-4R were solely used for control sequencing and purified by GSF only.

Table 9.9: Primer for vector mutagenesis

Name	**Sequence 5'→3'**
pcDNA3.1 *EcoRI* F	GCCACCGAATTCTGCAGATATCC
pcDNA3.1 *PacI* R	TCGAGGTTAATTAACGGGTTTATCACTTATCG
pcDNA3.1 *PacI* F	TAAACCCGTTAATTAACCTCGACTGTGCC
pcDNA3.1 *BpiI* R	ATGCCTGCTATTGTCTTCCCAATCCTCCCC
pcDNA3.1 *MunI* F	GACCGACAATTGCATGAAGAATCTGC
pcDNA3.1 Flag NR5A2 F	GCCACCATGGATTACAAGGATGACGACGATAAGTCTTCTAATTCAGATAC
pcDNA3.1 Flag NR5A2 R	GAAGACTTATCGTCGTCATCCTTGTAATCCATGGTGGCGGATCCG
pcDNA3.1 *HindIII* F	AAACTTAAGCTTGGTACCGAGCTCGGATCC
pcDNA3.1 NR5A2 *PacI* R	TCGAGGTTAATTAACGGGTTTATTATGCTCTTTTGGCATGCAACATTTC
pcDNA3.1 RBPJL *PacI* R	TCGAGGTTAATTAACGGGTTTACTAAGTCTGGATGAAGAGGTGGAAGTTG
pcDNA3.1 HNF1a *PacI* R	TCGAGGTTAATTAACGGGTTTATTACTGGGAGGAAGAGGCCATCTGGGTG
pcDNA Flag CTRL F	CGGATCCGCCACCATGGATTACAAGGATGACGACGATAAGTTGCTGCTCAGCCTGACCC
pcDNA Flag CTRL R	GGGTCAGGCTGAGCAGCAACTTATCGTCGTCATCCTTGTAATCCATGGTGGCGGATCCG
CTRL L218R *PacI* R	TCGAGGTTAATTAACGGGTTTATTAGTTGTAGGCTATGACCTGG
Sequencing primer	Sequence 5'→3'
pcDNA3.1-2F	CGCAAATGGGCGGTAGGCGTG
pcDNA3.1-4R	CTACTCAGACAATGCGATGC

9.1.4 Primers for XBP1 splicing PCR

Primer sequences are adopted from a former protocol by Dr. Lara Unger as part of her Master Thesis in 2018 and ordered from TIB MOLBIOL in GSF purification grade. The primer combination amplifies both the spliced and un-spliced XBP1 as different product sizes.

Table 9.10: Primers for XBP1 splicing PCR

Name	**Sequence 5'→3'**
XBP1 F	CCTTGTAGTTGAGAACCAGG
XBP1 R	GGGCTTGGTATATATGTGG

9.2 TaqMan assays for qPCR

All primers and probes for TaqMan assays were designed and obtained from TIB MOLBIOL. For each assay, we tested two forward and two reverse primers for their optimal combination. Based on the cycle threshold (C_t) value and the slope of the results, we selected the best combination. Most establishment work was conducted by Dr. Lara Unger, except the TaqMan assays for *BiP*, *ALB*, *HNF4a*, *AKAP7*, *HOMER2*, *IL32*, *C3*, and *CYP7A1*. The hydrolysis probe is labeled with a donor (6FAM-) and an acceptor (—BBQ). The primers were ordered in a GSF purification grade, the probes in HPLC purification grade.

Table 9.11: TaqMan assay for ER stress analysis

Name	Orientation	Sequence 5'→3'
BiP F	forward	AAATTTGAAGAGCTCAACATGGATCT
BiP A	reverse	GGTTTATGCCACGGGATGGT
BiP P	probe	6FAM-CCGGTCTACTATGAAGCCCGTCCAGAAAGTG-—BBQ

Table 9.12: TaqMan assays for housekeepers

Name	Orientation	Sequence 5'→3'
ACTB F	forward	AGCCTCGCCTTTGCCGA
ACTB R	reverse	CTGGTGCCTGGGGCG
ACTB TM	probe	6FAM-CCGCCGCCCGTCCACACCCGCC—BBQ
GAPDH fwd	forward	GAAGGTGAAGGTCGGAGTC
GAPDH rev	reverse	GAAGATGGTGATGGGATTTC
GAPDH TM	probe	6FAM-CAAGCTTCCCGTTCTCAGCC—BBQ

Table 9.13: TaqMan assays for transcription factors

Name	Orientation	Sequence 5'→3'
NR5A2 S	forward	AATGTCTAAGTGTTGGAATGAAGCTA
NR5A2 A	reverse	GCCCAAACTTATTCCTTCCTC
NR5A2 P	probe	6FAM-ACGCATTCGGTCGGCCCTTACA—BBQ
HNF1a F	forward	CTCCCCCAGTAAGGTCCACGGTG
HNF1a R	reverse	CCCCCAGCTGCTGAGACCAG
HNF1a P	probe	6FAM-CTATGGACAGCCTGCGACCAGTG—BBQ
RBPJL S	forward	CTACCTCCCATGATCATCCG
RBPJL A	reverse	GCCTGGAAACTGGAATGC
RBPJL P	probe	6FAM-CAGTGTGCGCTCCTTGATGTGGA—BBQ

Table 9.14: TaqMan assays for transcription factor target genes

Name	Orientation	Sequence 5'→3'
hCPA1 Fa	forward	CTGGTTTGCAAAGAAGATCAC
hCPA1 A	reverse	TTGGTGACGATCTCCAGGAA
hCPA1 TM	probe	6FAM-CAGCTTTCACCGCCATTCTCGAC—BBQ
CELA3AB Q34F	forward	GCCGGCCACTGCATCTCGAG
CELA3AB A	reverse	GATGTCATTGCCACAGGC
CELA3AB TM	probe	6FAM-CACACGAGCGGTTCCAGAGTGG—BBQ
CEL F	forward	GCCAATCTGCCAGGTAACT

Name	Orientation	Sequence 5'→3'
CEL A	reverse	GATATTCCTCTTCACCCAAGCAA
CEL P	probe	6FAM-CTTCGGGATCAGCACATGGC—BBQ
SPINK1 Q12F	forward	CTGTTGAGTCTATCTGGTAACAC
SPINK1 B	reverse	GTTCATTGTAACATTTGGCCTC
SPINK1 P1	probe	6FAM-CTTCCCAGGGAGTCAGCTCCAG—BBQ
hPRSS1 F	forward	ACTGCTACAAGTCCCGCA
hPRSS1 A	reverse	GAGTCTTCCTGTCGTATTGG
hPRSS1 TM	probe	6FAM-CTGGGAGAGCACAACATCGAAGTCC—BBQ
GP2 F	forward	CCATCCACTGTGGAGGACAA
GP2 A	reverse	GCAAACTGTGGACATCAGAACTA
GP2 P	probe	6FAM- CGAGGAGGAGTGCCTTGCCCT—BBQ
REG1A F	forward	CTGACCTCAAGCACAGGATTC
REG1A R	reverse	GGGTAGGTAAAACTATTGAAGATTAACAA
REG1A P	probe	6FAM-CCTTGTGAAGACAAGTTCTCCTTTGTCTGC—BBQ
ALB S	forward	TGCAGAAGACTATCTATCCGTGG
ALB A	reverse	TGACTCTGTCACTTACTGGCG
ALB P	probe	6FAM-CCTGAACCAGTTATGTGTGTTGCATGAG—BBQ
HNF4a S	forward	CATGTACTCCTGCAGATTTAGCC
HNF4a A	reverse	CAGTAGCGGCACTGGTTC
HNF4a P	probe	6FAM-AGTGCGTGGTGGACAAAGACAAG—BBQ
AKAP7 S	forward	GCTGGTGATGCAATTATTAAATGAAGATG
AKAP7 R	reverse	ATCCAACCTGATTTCCAAAAGTAC
AKAP7 P	probe	6FAM-CCCTGGAGGAGTTCTTCTATGAATGGTTTC—BBQ
HOMER2 S	forward	TCGAGACCTCAAGTAATCATTCC
HOMER2 R	reverse	CTTCAGGTGTGTGTTGGCTG
HOMER2 P	probe	6FAM-CCAGTGTCAACGGGACGGAC—BBQ
C3 F	forward	GCGAATGGACAAAGTCGG
C3 R	reverse	GCCCAGGGAGATGAAACG
C3 P	probe	6FAM-CGCAGCTCCTTGGGGTACTT—BBQ
IL32ex3,4 F	forward	CATGTGCTTCCCGAAGGTC
IL32ex4 R	reverse	CATTCGGGCCTTCAGCTT
IL32ex4 P	probe	6FAM-TTCATGTCATCA+GA+GA+GG—BBQ
CCL2 S	forward	GCTCAGCCAGATGCAATC
CCL2 R	reverse	CCACAATGGTCTTGAAGATCACA
CCL2 TM	probe	6FAM-CACCTGCTGTTATAACTTCACCAATAGGAA—BBQ
IL18 F	forward	CAGTCTACACAGCTTCGGGA
IL18 R	reverse	CACAAAGTTGATGCAATTGTCTTCT
IL18 TM	probe	6FAM-AGATCGCTTCCTCTCGCAACAAACTA—BBQ
CYP7A1 S	forward	AAAGTCAGCTTGGAAGGCA
CYP7A1 A	reverse	TCCTTGATTATACTATCTAATACTGGCA
CYP7A1 P	probe	6FAM-TTGAGTCAAGCAGAACTGAATGACCTG—BBQ

hCPA1 = human carboxypeptidase A1; CELA3AB = chymotrypsin-like elastase family member 3A and 3B; CEL = carboxyl ester lipase; SPINK1 = serine protease inhibitor kazal-type 1; hPRSS1 = human serine protease 1; GP2 = glycoprotein 2; REG1A = regenerating family member 1A; ALB = albumin; HNF4a = hepatocyte nuclear factor 4 alpha; AKAP7 = A kinase anchoring protein 7; HOMER2 = homer scaffold protein 2; C3 = complement 3; IL32 = interleukin 32; CCL2 = C-C motif chemokine ligand 2; IL18 = interleukin 18; CYP7A1 = cholesterol 7 alpha hydroxylase.

9.3 Synonymous *AQP12A* and *AQP12B* variants

Data shows synonymous and intronic variants in both genes that we detected besides non-synonymous variants. Again, for these variants there was no statistically significant difference between patients and controls (p>0.05) in any of the statistical calculations. For the intronic and synonymous variants, a less stringent CADD score (score ≥10) was used, however, no variant was predicted to be deleterious.

Table 9.15: Synonymous and intronic *AQP12A* variants in German CP patients and controls

Exon/ Intron	Position (hg38)	Nucleotide change	AA change	rs number	CADD	CP (n=292) n (%)	Controls (n=143) n (%)	*p*-value
1	2:240692042	c.123+6T>C	-	rs780725873	-	2 (0.7)	3 (2.1)	0.34
1	2:240692056	c.124-18G>T	-	rs556220702	-	1 (0.3)	0	1.0
1	2:240692059	c.124-15C>G c.124-15G>G	-	rs4328629	-	123 (42.1) 89 (30.5)	58 (40.6) 49 (34.3)	0.71
1	2:240692066	c.124-8G>T	-	rs4374371	-	0	1 (0.7)	0.33
2	2:240692082	c.132G>A	p.T44=	rs200769008	3.52	1 (0.3)	0	1.0
2	2:240692199	c.249T>C	p.T83=	rs11889147	0.94	1 (0.3)	3 (2.1)	0.11
2	2:240692248	c.298C>T	p.L100= (het) p.L100= (hom)	rs76463197	5.93	98 (33.6) 26 (8.9)	44 (30.8) 11 (7.7)	0.74
2	2:240692469	c.519C>G	p.P173= (het) p.P173= (hom)	rs12465023	0.92	130 (44.5) 44 (15.1)	57 (39.9) 25 (17.5)	0.62
2	2:240692484	c.534C>T	p.P178= (het) p.P178= (hom)	rs58669952	0.46	131 (44.9) 43 (14.7)	58 (40.6) 25 (17.5)	0.61

P-values were calculated by two-tailed Fisher's Exact test with 2x3 contingency table comparing genotype frequencies and with a 2x2 contingency table comparing carrier frequencies. CP = chronic pancreatitis; AA = amino acid; CADD = combined annotation dependent depletion; REVEL = rare exome variant ensemble learner; het = heterozygous; hom = homozygous.

Table 9.16: Synonymous and intronic *AQP12B* variants in German CP patients and controls

Exon/ Intron	Position (hg38)	Nucleotide change	AA change	rs number	CADD	CP (n=292) n (%)	Controls (n=143) n (%)	*p*-value
1	2:240682856_40682857	c.-19_-20AT>TC	- (het) - (hom)	rs1559457676	-	83 (28.4) 10 (3.4)	49 (34.3) 6 (4.2)	0.38
1	2:240682846	c.-9C>T	-	rs375213961	1.28	1 (0.3)	0	1.0
1	2:240682778	c.60A>G	p.A20= (het) p.A20= (hom)	rs184458523	3.94	77 (26.4) 7 (2.4)	40 (28.0) 4 (2.8)	0.83
1	2:240682709	c.129C>T	p.G43= (het) p.G43= (hom)	rs192451353	1.13	23 (7.9) 2 (0.7)	14 (9.8) 1 (0.7)	0.79
1	2:240682694	c.144G>C	p.G48=	rs370144346	4.71	1 (0.3)	0	1.0
1	2:240682664	c.174C>T	p.V58=	rs200034703	0.57	3 (1.0)	5 (3.5)	0.12
1	2:240682652	c.186C>T	p.P62=	rs4676387	6.26	0	1 (0.7)	0.33
1	2:240682592	c.246G>C	p.A82= (het) p.A82= (hom)	rs192736899	2.07	21 (7.2) 2 (0.7)	14 (9.8) 1 (0.7)	0.60
1	2:240682589	c.249C>T	p.H83=	rs373980987	2.31	1 (0.3)	0	1.0
1	2:240682501	c.337C>T	p.L113=	rs561469792	8.81	1 (0.3)	0	1.0
1	2:240682394	c.444C>T	p.A148= (het) p.A148= (hom)	rs80038749	8.91	10 (3.4) 0	8 (5.6) 1 (0.7)	0.13
3	2:240676632	c.837G>A	p.P279=	rs1293940937	2.22	0	2 (1.4)	0.11
3	2:240676520	*25G>A	-	rs752144507	1.24	1 (0.3)	0	1.0

P-values were calculated by two-tailed Fisher's Exact test with 2x3 contingency table comparing genotype frequencies and with a 2x2 contingency table comparing carrier frequencies. CP = chronic pancreatitis; AA = amino acid; CADD = combined annotation dependent depletion; REVEL = rare exome variant ensemble learner; het = heterozygous; hom = homozygous.

9.4 Chymotrypsin-like protease

9.4.1 Western blot analyses of CTRL mutants

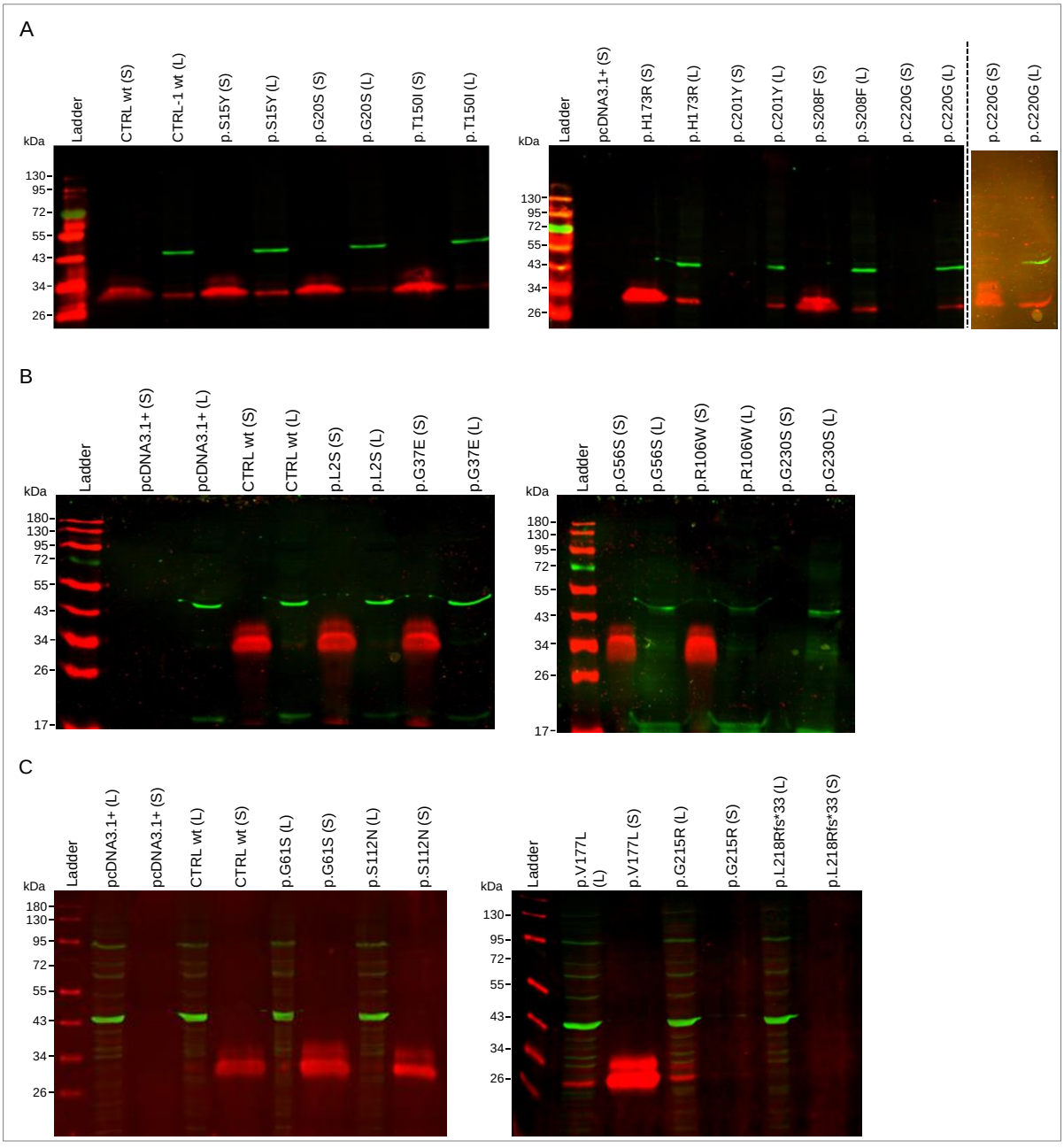


Figure 9.1: Original Western blot of HEK293T supernatant (S) and total lysate (L) after CTRL overexpression
Antibody against Flag-Tag detected the overexpressed CTRL (28 kDa; red bands). The housekeeper gene β-actin (42 kDa; green bands) served as internal loading control. We loaded supernatant protein precipitated from 400 µl media and 40 µg total lysate. The figures are representative for a total of four biological replicates. (**A**) Mutation set 1 found in CP patients by WES. p.C220G shows reduced secretion in some biological replicates. (**B**) Mutation set 2 found in CP patients (p.L2S, p.G56S, p.R106W) as well as controls (p.G37E, p.G230S). (**C**) Mutation set 3 found in the control cohort.
CTRL = chymotrypsin-like protease; pcDNA3.1+/C-(K)-DYK = empty vector plasmid

9.4.2 Activity assay of CTRL mutations with substrate AAPY

We performed CTRL activity assays CTRL simultaneously with two different synthetic substrates, as the enzyme cleaves polypeptides at tyrosine and phenylalanine residues (Szabó and Sahin-Tóth 2012). The synthetic substrate N-succinyl-Ala-Ala-Pro-Tyr-p-nitroanilide (AAPY) resulted in lower signal intensities than AAPF (with a phenylalanine residue). Therefore, we continued with AAPF only. The results with AAPY are shown in Figure 9.2.

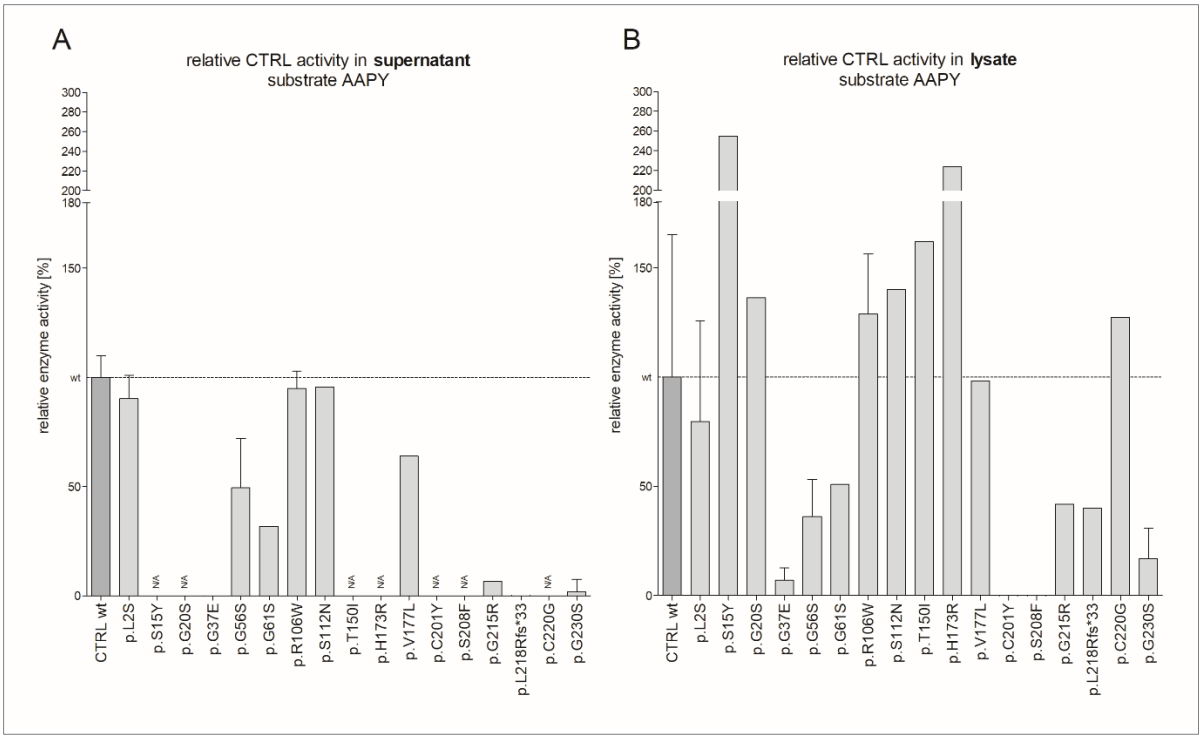


Figure 9.2: CTRL activity in supernatant and lysate from HEK293T cells using AAPY
We activated the enzyme with 125 ng trypsin for 2 h. We performed activity assays in technical triplicates on minimum one biological replicate (n=1). The figure shows enzyme activity 8 min after the substrate AAPY was added. All measurements were set in relation to CTRL wt activity (100 %; dark grey bar). (**A**) Relative CTRL activity in supernatant. (**B**) Relative CTRL activity in lysate.
AAPY = N-Suc-Ala-Ala-Pro-Tyr-p-NA; N/A = not determined

9.4.3 Activation assay of CTRL mutations p.G20S and p.G56S

As the CTRL mutations p.G20S and p.G56S displayed reduced activity in the AAPF assay, we were wondering whether this is a consequence of a delayed activation of CTRL or a reduced activity overall. For this, we performed an activation with 125 ng porcine trypsin for a total of 2 h as for the activity assay but repeatedly took samples at ten different incubation time points to measure the proteolytic capacity of CTRL. This allowed us to generate an activity graph of CTRL over time, as shown in Figure 9.3. We measured activity 5 min, 10 min, 20 min, 30 min, 40 min, 50 min, 60 min, 80 min, 100 min, and 120 min after activation. While p.G56S has a constant reduced activity compared to CTRL wt over the course of 2 h, p.G20S shows an initially higher activity, which drops below the CTRL wt after 30 min in the supernatant and after 15 min in the lysate.

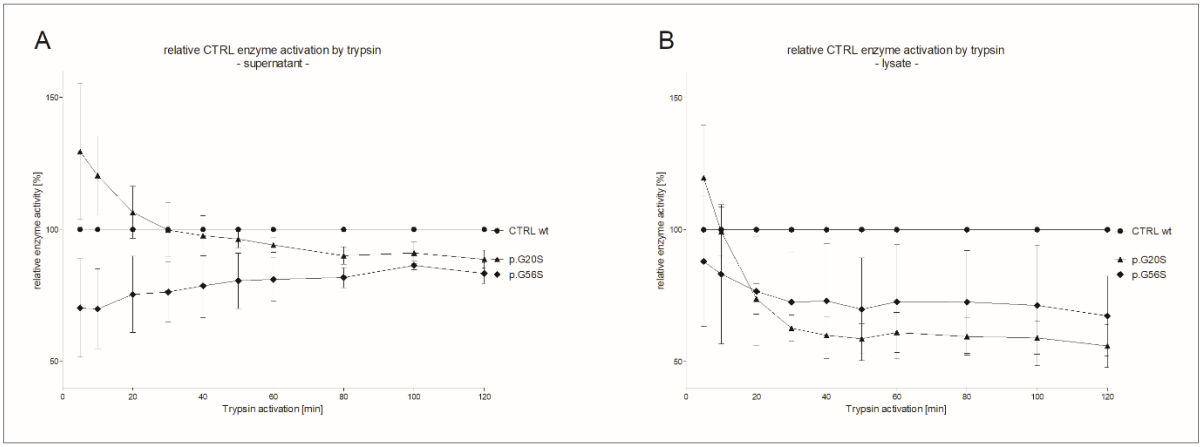


Figure 9.3: CTRL activity in supernatant and lysate samples over time
Supernatant and total lysate were prepared from HEK293T cells after overexpression of CTRL wildtype (wt), p.G20S or p.G56S. We activated the enzyme with 125 ng porcine trypsin for 2 h and took samples at ten different incubation time points. The figure shows enzyme activity 8 min after the substrate AAPF (N-Suc-Ala-Ala-Pro-Phe-p-NA) was added to the sample. All measurements were set in relation to CTRL wt activity (100 %; horizontal line). (**A**) Relative CTRL activity in supernatant. Activation assay was performed in technical triplicates on four biological replicates (n=4). (**B**) Relative CTRL activity in lysate. Activation assay was performed in technical triplicates on three biological replicates (n=3).

9.5 Transcription factors

9.5.1 Flag-tag vs. untagged

During the project, the question arose whether the Flag-tag impedes the activity of the transcription factors. Therefore, we overexpressed the three transcription factors NR5A2, RBPJL and HNF1a either with or without the Flag-tag in combination with the PTF1 complex (PTF1a + RBPJL) in three different cell lines: HEK293T, HEK293 and Panc1. After 72 h, RNA was isolated and transcribed into cDNA. For each transcription factor, we conducted a selected panel of TaqMan assays for potential target genes. Additionally, we performed the TaqMan assay for the transcription factor itself to verify its expression. The results are shown in Figure 9.4. In the case of NR5A2 (**A**), the target genes *CPA1* and *CELA3AB* showed an increased expression when NR5A2 in combination with the PTF1a complex is present compared to the empty control and untransfected samples (white bars). No differences between the N-terminal Flag-tag and the untagged transcription factor complex was visible in all three cell lines. Looking at RBPJL (**B**), the target genes *CPA1* and *GP2* were both expressed though *GP2* expression was rather low in all cell lines. Nevertheless, the results of the C-terminal Flag-tag and the untagged transcription factor did not differ. The same applies to HNF1a (**C**). mRNA expression of its target gene *SPINK1* was increased compared to the empty control or the untransfected samples and did not result in different expression levels with or without the C-terminal Flag-tag.

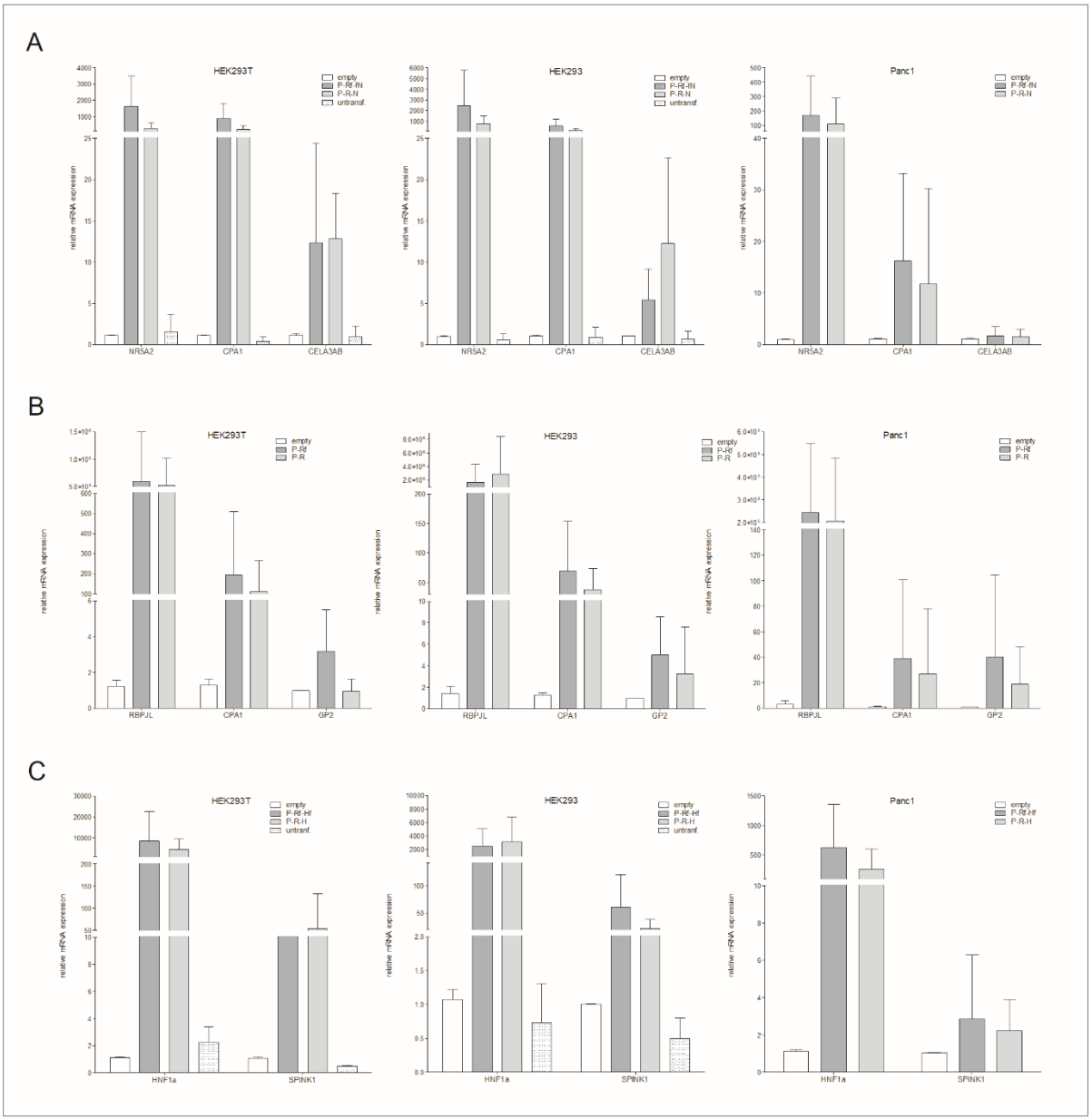


Figure 9.4: qPCR results of transcription factors target gene expression – Flag-tag vs. untagged
Overexpression with either Flag-tag (dark grey bar) or without tag (light grey bar) in three cell lines: HEK293T (left); HEK293 (middle); Panc1 (right). pcDNA3.1+/C-(K)-DYK served as negative control (empty, white bar) as well as samples from untransfected cells (white patterned bars). We performed TaqMan assays on selected target genes with three technical replicates of a biological n = 4. GAPDH and β-actin served as internal controls. Some experiments were conducted by Marlis Kugler as part of her Bachelor thesis. (**A**) NR5A2: mRNA expression of *NR5A2*, *CPA1* and *CELA3AB* (N-terminal Flag-tag). (**B**) RBPJL: mRNA expression of *RBPJL*, *CPA1* and *GP2* (C-terminal Flag-tag). (**C**) HNF1a: mRNA expression of *HNF1a* and *SPINK1* (C-terminal Flag-tag).
P-Rf-fN = PTF1a + RBPJL-flag + Flag-NR5A2; P-R-N = PTF1a + RBPJL + NR5A2; P-Rf-Hf = PTF1a + RBPJL-flag + HNF1a-flag; P-R-H = PTF1a + RBPJL + HNF1a; P-Rf = PTF1a + RBPJL-flag; P-R = PTF1a + RBPJL; untransf. = untransfected.

The results show that the Flag-tag does not affect the transcriptional activity of the transcription factors and thus does not distort the results generated with tagged proteins. Nevertheless, except for Western blot analyses, we used the untagged factors for the experiments to generate conditions closer to biological.

9.5.2 Western blot analysis of all NR5A2 mutations

We performed Western blot as described in section 5.5 in Methods to verify the successful overexpression of NR5A2 wildtype (wt) and mutants. Total lysate was prepared from HEK293 cells, and a total of 40 µg were loaded onto the gel. For the detection on the Western blot membrane, a specific polyclonal antibody directed against human NR5A2 was used (rabbit; 1:1,000). The housekeeper gene β-actin was used as internal loading control (goat; 1:2,500). Figure 9.5 shows NR5A2 wt and all 15 mutations at the correct height of roughly 61 kDa (red). The β-actin band (green) suggest an even signal intensity and consequently an equal loading of the samples.

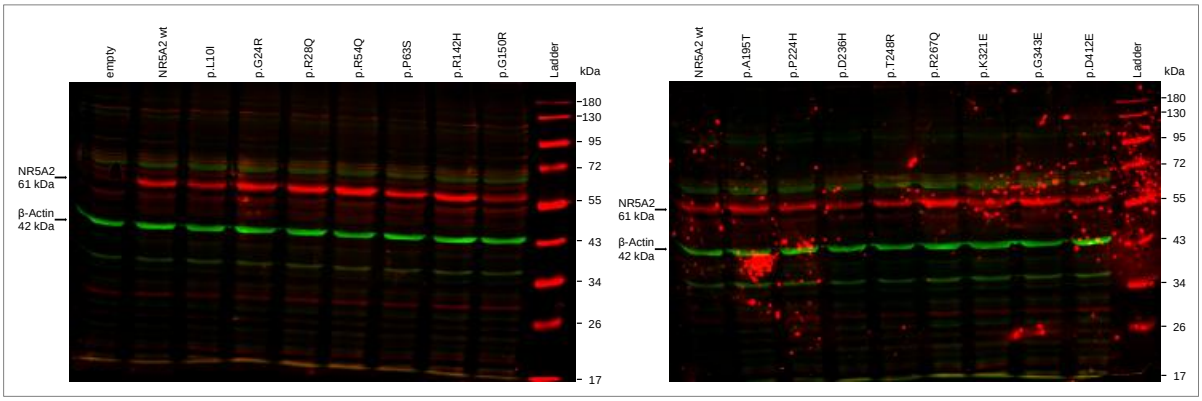


Figure 9.5: Western blot analysis of all NR5A2 mutations
Total lysate was prepared from HEK293 cells. A specific antibody against human NR5A2 was used in 1:1,000 ratio to detect the overexpressed NR5A2 wildtype (wt) and the 15 investigated mutants (61 kDa; red bands). For better comparison, the NR5A2 wt was loaded on each gel separately. The housekeeper gene β-actin (42 kDa; green bands) was used as internal loading control (ratio 1:2,500). The gel was loaded with 40 µg total lysate. pcDNA3.1+/C-(K)-DYK vector (empty) was used as control plasmid. NR5A2 = nuclear receptor subfamily 5 group A member 2

9.6 Company provided protocols

- Wizard® SV Gel and PCR Clean-Up System (Promega, A9282)
- RNeasy® Mini Kit (Qiagen, 74106)
- QuantiTect® Reverse Transcription Kit (Qiagen, 205313)
- Nuclear Extract Kit (Active Motif Europe, 40010)

Wizard® SV Gel and PCR Clean-Up System

INSTRUCTIONS FOR USE OF PRODUCTS A9280, A9281, A9282, AND A9285.

Quick PROTOCOL

DNA Purification by Centrifugation

Gel Slice and PCR Product Preparation

A. Dissolving the Gel Slice

1. Following electrophoresis, excise DNA band from gel and place gel slice in a 1.5ml microcentrifuge tube.
2. Add 10µl Membrane Binding Solution per 10mg of gel slice. Vortex and incubate at 50–65°C until gel slice is completely dissolved.

B. Processing PCR Amplifications

1. Add an equal volume of Membrane Binding Solution to the PCR amplification.

Binding of DNA

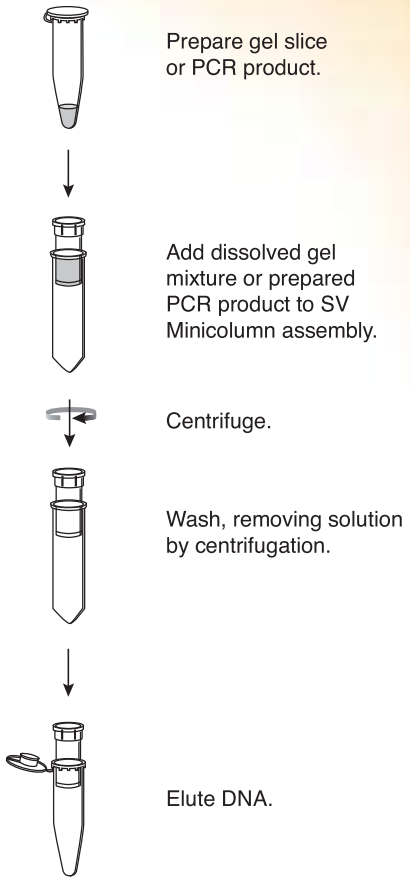
1. Insert SV Minicolumn into Collection Tube.
2. Transfer dissolved gel mixture or prepared PCR product to the Minicolumn assembly. Incubate at room temperature for 1 minute.
3. Centrifuge at 16,000 × *g* for 1 minute. Discard flowthrough and reinsert Minicolumn into Collection Tube.

Washing

4. Add 700µl Membrane Wash Solution (ethanol added). Centrifuge at 16,000 × *g* for 1 minute. Discard flowthrough and reinsert Minicolumn into Collection Tube.
5. Repeat Step 4 with 500µl Membrane Wash Solution. Centrifuge at 16,000 × *g* for 5 minutes.
6. Empty the Collection Tube and recentrifuge the column assembly for 1 minute with the microcentrifuge lid open (or off) to allow evaporation of any residual ethanol.

Elution

7. Carefully transfer Minicolumn to a clean 1.5ml microcentrifuge tube.
8. Add 50µl of Nuclease-Free Water to the Minicolumn. Incubate at room temperature for 1 minute. Centrifuge at 16,000 × *g* for 1 minute.
9. Discard Minicolumn and store DNA at 4°C or –20°C.



3760MA07_2A

Additional protocol information is available in Technical Bulletin #TB308, available online at:
www.promega.com

ORDERING/TECHNICAL INFORMATION:
www.promega.com • Phone 608-274-4330 or 800-356-9526 • Fax 608-277-2601

Printed in USA. Revised 11/09
Part #9FB072

Protocol: Purification of Total RNA from Animal Cells Using Spin Technology

This protocol requires the RNeasy Mini Kit.

Determining the correct amount of starting material

It is essential to use the correct amount of starting material in order to obtain optimal RNA yield and purity. The minimum amount is generally 100 cells, while the maximum amount depends on:

- The RNA content of the cell type
- The RNA binding capacity of the RNeasy spin column (100 µg RNA)
- The volume of Buffer RLT required for efficient lysis (the maximum volume of Buffer RLT that can be used limits the maximum amount of starting material to 1 x 10^7 cells)

RNA content can vary greatly between cell types. The following examples illustrate how to determine the maximum amount of starting material:

- COS cells have high RNA content (approximately 35 µg RNA per 10^6 cells). Do not use more than 3 x 10^6 cells, otherwise the RNA binding capacity of the RNeasy spin column will be exceeded.
- HeLa cells have average RNA content (approximately 15 µg RNA per 10^6 cells). Do not use more than 7 x 10^6 cells, otherwise the RNA binding capacity of the RNeasy spin column will be exceeded.
- NIH/3T3 cells have low RNA content (approximately 10 µg RNA per 10^6 cells). The maximum amount of starting material (1 x 10^7 cells) can be used.

If processing a cell type not listed in Table 2 (page 21) and if there is no information about its RNA content, we recommend starting with no more than 3–4 x 10^6 cells. Depending on RNA yield and purity, it may be possible to increase the cell number in subsequent preparations.

Do not overload the RNeasy spin column, as this will significantly reduce RNA yield and purity.

Counting cells is the most accurate way to quantitate the amount of starting material. As a guide, the number of HeLa cells obtained in various culture vessels after confluent growth is given in Table 4.

Table 4. Growth area and number of HeLa cells in various culture vessels

Cell-culture vessel	Growth area (cm^2)*	Number of cells†
Multiwell-plates		
96-well	0.32–0.6	4–5 x 10^4
48-well	1	1 x 10^5
24-well	2	2.5 x 10^5
12-well	4	5 x 10^5
6-well	9.5	1 x 10^6
Dishes		
35 mm	8	1 x 10^6
60 mm	21	2.5 x 10^6
100 mm	56	7 x 10^6
145–150 mm	145	2 x 10^7
Flasks		
40–50 ml	25	3 x 10^6
250–300 ml	75	1 x 10^7
650–750 ml	162–175	2 x 10^7

* Per well, if multiwell plates are used; varies slightly depending on the supplier.
† Cell numbers are given for HeLa cells (approximate length = 15 µm), assuming confluent growth. Cell numbers will vary for different kinds of animal cells, which vary in length from 10 to 30 µm.

Important points before starting

- If using the RNeasy Kit for the first time, read "Important Notes" (page 19).
- If working with RNA for the first time, read Appendix A (page 73).

- Cell pellets can be stored at –90 to –65°C for later use or used directly in the procedure. Determine the number of cells before freezing. Frozen cell pellets should be thawed slightly so that they can be dislodged by flicking the tube in step 2. Homogenized cell lysates from step 3 can be stored at –90 to –65°C for several months. Frozen lysates should be incubated at 37°C in a water bath until completely thawed and salts are dissolved. Avoid prolonged incubation, which may compromise RNA integrity. If any insoluble material is visible, centrifuge for 5 min at 3000–5000 x *g*. Transfer supernatant to a new RNAse-free glass or polypropylene tube, and continue with step 4. Buffer RLT may form a precipitate upon storage. If necessary, redissolve by warming, and then place at room temperature.
- Buffer RLT and Buffer RW1 contain a guanidine salt and are therefore not compatible with disinfecting reagents containing bleach. See page 7 for safety information.
- Perform all steps of the procedure at room temperature. During the procedure, work quickly.
- Perform all centrifugation steps at 20–25°C in a standard microcentrifuge. Ensure that the centrifuge does not cool below 20°C.

Things to do before starting

- If purifying RNA from cell lines rich in RNAses, we recommend adding β-mercaptoethanol (β-ME) to Buffer RLT before use. Add 10 µl β-ME per 1 ml Buffer RLT. Dispense in a fume hood and wear appropriate protective clothing. Buffer RLT containing β-ME can be stored at room temperature for up to 1 month.
- Buffer RPE is supplied as a concentrate. Before using for the first time, add 4 volumes of ethanol (96–100%) as indicated on the bottle to obtain a working solution.
- If performing optional on-column DNase digestion, prepare DNase I stock solution as described in Appendix D (page 82).

Procedure

1. Harvest cells according to step 1a or 1b.

1a. Cells grown in suspension (do not use more than 1 x 10^7 cells):
Determine the number of cells. Pellet the appropriate number of cells by centrifuging for 5 min at 300 x *g* in a centrifuge tube (not supplied). Carefully remove all supernatant by aspiration, and proceed to step 2.
Note: Incomplete removal of cell-culture medium will inhibit lysis and dilute the lysate, affecting the conditions for binding of RNA to the RNeasy membrane. Both effects may reduce RNA yield.

1b. Cells grown in a monolayer (do not use more than 1 x 10^7 cells):
Cells can be either lysed directly in the cell-culture vessel (up to 10 cm diameter) or trypsinized and collected as a cell pellet prior to lysis. Cells grown in cell-culture flasks should always be trypsinized.
To lyse cells directly:
Determine the number of cells. Completely aspirate the cell-culture medium, and proceed immediately to step 2.
Note: Incomplete removal of cell-culture medium will inhibit lysis and dilute the lysate, affecting the conditions for binding of RNA to the RNeasy membrane. Both effects may reduce RNA yield.
To trypsinize and collect cells:
Determine the number of cells. Aspirate the medium, and wash the cells with PBS. Aspirate the PBS, and add 0.1–0.25% trypsin in PBS. After the cells detach from the dish or flask, add medium (containing serum to inactivate the trypsin), transfer the cells to an RNAse-free glass or polypropylene centrifuge tube (not supplied), and centrifuge at 300 x *g* for 5 min. Completely aspirate the supernatant, and proceed to step 2.
Note: Incomplete removal of cell-culture medium will inhibit lysis and dilute the lysate, affecting the conditions for binding of RNA to the RNeasy membrane. Both effects may reduce RNA yield.

2. Disrupt the cells by adding Buffer RLT.

   For pelleted cells, loosen the cell pellet thoroughly by flicking the tube. Add the appropriate volume of Buffer RLT (see Table 5). Vortex or pipet to mix, and proceed to step 3.

   Note: Incomplete loosening of the cell pellet may lead to inefficient lysis and reduced RNA yields.

Table 5. Volumes of Buffer RLT for lysing pelleted cells

Number of pelleted cells	Volume of Buffer RLT (µl)
<5 x 10^6	350
5 x 10^6 – 1 x 10^7	600

For direct lysis of cells grown in a monolayer, add the appropriate volume of Buffer RLT (see Table 6) to the cell-culture dish. Collect the lysate with a rubber policeman. Pipet the lysate into a microcentrifuge tube (not supplied). Vortex or pipet to mix, and ensure that no cell clumps are visible before proceeding to step 3.

Table 6. Volumes of Buffer RLT for direct cell lysis

Dish diameter (cm)	Volume of Buffer RLT (µl)*
<6	350
6–10	600

* Regardless of the cell number, use the buffer volumes indicated to completely cover the surface of the dish.

3. Homogenize the lysate according to step 3a, 3b, or 3c.
   See “Disrupting and homogenizing starting material”, page 22, for more details on homogenization. If processing ≤1 x 10^5 cells, homogenize by vortexing for 1 min. After homogenization, proceed to step 4.

Note: Incomplete homogenization leads to significantly reduced RNA yields and can cause clogging of the RNeasy spin column. Homogenization with a rotor–stator or QIAshredder homogenizer generally results in higher RNA yields than with a syringe and needle.

3a. Pipet the lysate directly into a QIAshredder spin column placed in a 2 ml collection tube, and centrifuge for 2 min at full speed. Proceed to step 4.

3b. Homogenize the lysate for 30 s using a rotor–stator homogenizer. Proceed to step 4.

3c. Pass the lysate at least 5 times through a blunt 20-gauge needle (0.9 mm diameter) fitted to an RNAse-free syringe. Proceed to step 4.

4. Add 1 volume of 70% ethanol to the homogenized lysate, and mix well by pipetting.
   Important: Do not centrifuge.
   Note: The volume of lysate may be less than 350 µl or 600 µl due to loss during homogenization.
   Note: When purifying RNA from certain cell lines, precipitates may be visible after addition of ethanol. This does not affect the procedure.
5. Transfer up to 700 µl of the sample, including any precipitate that may have formed, to an RNeasy spin column placed in a 2 ml collection tube (supplied). Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10,000 rpm). Discard the flow-through.*
   Reuse the collection tube in step 6.
   If the sample volume exceeds 700 µl, centrifuge successive aliquots in the same RNeasy spin column. Discard the flow-through after each centrifugation.*
   Optional: If performing optional on-column DNase digestion (see "Eliminating genomic DNA contamination", page 26), follow Appendix D (page 82), steps 1–4, after performing this step.

* Flow-through contains Buffer RLT or Buffer RW1 and is therefore not compatible with bleach. See page 6 for safety information.

6. Add 700 µl Buffer RW1 to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10,000 rpm) to wash the spin column membrane. Discard the flow-through.*

   Reuse the collection tube in step 7.

   Note: After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Be sure to empty the collection tube completely.

   Skip this step if performing optional on-column DNase digestion (page 82).
7. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x *g* (≥10,000 rpm) to wash the spin column membrane. Discard the flow-through.

   Reuse the collection tube in step 8.

   Note: Buffer RPE is supplied as a concentrate. Ensure that ethanol is added to Buffer RPE before use (see "Things to do before starting").
8. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently, and centrifuge for 2 min at ≥8000 x *g* (≥10,000 rpm) to wash the spin column membrane.

   The long centrifugation dries the spin column membrane, ensuring that no ethanol is carried over during RNA elution. Residual ethanol may interfere with downstream reactions.

   Note: After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Otherwise, carryover of ethanol will occur.
9. **Optional**: Place the RNeasy spin column in a new 2 ml collection tube (supplied), and discard the old collection tube with the flow-through. Close the lid gently, and centrifuge at full speed for 1 min.

   Perform this step to eliminate any possible carryover of Buffer RPE, or if residual flow-through remains on the outside of the RNeasy spin column after step 8.

* Flow-through contains Buffer RLT or Buffer RW1 and is therefore not compatible with bleach. See page 6 for safety information.

10. Place the RNeasy spin column in a new 1.5 ml collection tube (supplied). Add 30–50 µl RNAse-free water directly to the spin column membrane. Close the lid gently, and centrifuge for 1 min at ≥8000 x *g* (≥10,000 rpm) to elute the RNA.
11. If the expected RNA yield is >30 µg, repeat step 10 using another 30–50 µl RNAse-free water, or using the eluate from step 10 (if high RNA concentration is required). Reuse the collection tube from step 10.
    If using the eluate from step 10, the RNA yield will be 15–30% less than that obtained using a second volume of RNAse-free water, but the final RNA concentration will be higher.

Protocol: Reverse Transcription with Elimination of Genomic DNA for Quantitative, Real-Time PCR

Important points before starting

- The protocol is optimized for use with 10 pg to 1 µg of RNA. If using >1 µg RNA, scale up the reaction linearly to the appropriate volume.
- Set up all reactions on ice to minimize the risk of RNA degradation.
- **RNase inhibitor and dNTPs are already included in the kit components.** Do not add additional RNase inhibitor and dNTPs.
- RT Primer Mix (supplied) or gene-specific primers (not supplied) should be used. RT Primer Mix is optimized to provide high cDNA yields for all regions of RNA transcripts. If using gene-specific primers, we recommend using a final concentration of 0.7 µM or testing a range of final concentrations from 0.5 µM to 1 µM.
- **For convenience, premix RT Primer Mix and 5x Quantiscript RT Buffer in a 1:4 ratio if RT Primer Mix will be used routinely for reverse transcription.** This premix is stable when stored at –20°C.
- Separate denaturation and annealing steps are not necessary before starting the reverse-transcription reaction.
- If using a reaction volume of 200 µl or greater for reverse transcription, make sure the reaction tube is efficiently heated (e.g., if using a heating block, carefully fill each well with a drop of water so that heat can be efficiently transferred from the block to the tube).
- **After reverse transcription, the reaction must be inactivated by incubation at 95°C for 3 minutes.**
- If working with RNA for the first time, read Appendix A, page 18.
- For details on performing real-time PCR after reverse transcription, see Appendix C, page 23. For details on appropriate controls, see Appendix D, page 24.
- **Users of the FastLane® Cell cDNA Kit:** If you have purchased the QuantiTect Reverse Transcription Kit in order to perform additional reverse-transcription reactions with the FastLane Cell cDNA Kit, follow the protocol in the *FastLane Cell cDNA Handbook*. Do not follow the protocol in the *QuantiTect Reverse Transcription Handbook*.

Things to do before starting

- Dissolve any precipitates in gDNA Wipeout Buffer by vortexing. If necessary, briefly incubate the buffer at 37°C until the precipitates dissolve.

Procedure

1. **Thaw template RNA on ice. Thaw gDNA Wipeout Buffer, Quantiscript Reverse Transcriptase, Quantiscript RT Buffer, RT Primer Mix, and RNase-free water at room temperature (15–25°C).**

   Mix each solution by flicking the tubes. Centrifuge briefly to collect residual liquid from the sides of the tubes, and then store on ice.

2. **Prepare the genomic DNA elimination reaction on ice according to Table 1.**

   Mix and then store on ice.

   Note: If setting up more than one reaction, prepare a volume of master mix 10% greater than that required for the total number of reactions to be performed. Then distribute the appropriate volume of master mix into individual tubes followed by each RNA sample. Keep the tubes on ice.

   Note: The protocol is for use with 10 pg to 1 µg RNA. If using >1 µg RNA, scale up the reaction linearly. For example, if using 2 µg RNA, double the volumes of all reaction components for a final 28 µl reaction volume.

Table 1. Genomic DNA elimination reaction components

Component	Volume/reaction	Final concentration
gDNA Wipeout Buffer, 7x	2 µl	1x
Template RNA	Variable (up to 1 µg*)	
RNase-free water	Variable	
Total volume	**14 µl**	–

* This amount corresponds to the entire amount of RNA present, including any rRNA, mRNA, viral RNA, and carrier RNA present, and regardless of the primers used or cDNA analyzed.

3. **Incubate for 2 min at 42°C. Then place immediately on ice.**

   Note: Do not incubate at 42°C for longer than 10 min.

4. **Prepare the reverse-transcription master mix on ice according to Table 2.**

   Mix and then store on ice. The reverse-transcription master mix contains all components required for first-strand cDNA synthesis except template RNA.

   Note: If setting up more than one reaction, prepare a volume of master mix 10% greater than that required for the total number of reactions to be performed.

   Note: The protocol is for use with 10 pg to 1 µg RNA. If using >1 µg RNA, scale up the reaction linearly. For example, if using 2 µg RNA, double the volumes of all reaction components for a final 40 µl reaction volume.

Table 2. Reverse-transcription reaction components

Component	Volume/reaction	Final concentration
Reverse-transcription master mix		
Quantiscript Reverse Transcriptase*	1 µl	
Quantiscript RT Buffer, 5x†‡	4 µl	1x
RT Primer Mix‡	1 µl	
Template RNA		
Entire genomic DNA elimination reaction (step 3)	14 µl (add at step 5)	
Total volume	**20 µl**	–

* Also contains RNase inhibitor.

† Includes Mg^{2+} and dNTPs.

‡ For convenience, premix RT Primer Mix and 5x Quantiscript RT Buffer in a 1:4 ratio if RT Primer Mix will be used routinely for reverse transcription. This premix is stable when stored at –20°C. Use 5 µl of the premix per 20 µl reaction.

5. **Add template RNA from step 3 (14 µl) to each tube containing reverse-transcription master mix.**

   Mix and then store on ice.

6. **Incubate for 15 min at 42°C.**

   In some rare cases (e.g., if the RT-PCR product is longer than 200 bp or if analyzing RNAs with a very high degree of secondary structure), increasing the incubation time up to 30 min may increase cDNA yields.

7. **Incubate for 3 min at 95°C to inactivate Quantiscript Reverse Transcriptase.**

8. **Add an aliquot of each finished reverse-transcription reaction to real-time PCR mix (see Appendix C, page 23).**

   Store reverse-transcription reactions on ice and proceed directly with real-time PCR, or for long-term storage, store reverse-transcription reactions at –20°C. For real-time PCR, we recommend using a Rotor-Gene Kit, QuantiFast Kit, or QuantiTect Kit (see page 10).

Protocol I: Adherent or Suspension Cells

A. Preparation of Nuclear Extract from Cells

The following protocol is based on samples of approximately 8.8 x 10^6 cells, which correspond to HeLa cells grown to confluence in a 100 mm tissue culture plate. Each sample is one reaction. Prepare PBS/Phosphatase Inhibitors, Hypotonic Buffer and Complete Lysis Buffer as described above on page 5 in the section Buffer Preparation. Adjust the volumes according to the buffer preparation chart if using plates of different sizes. Place buffers and any tubes needed on ice before beginning assay.

Quick Chart for Preparing Buffers for Nuclear Extraction from Cells

Reagents to Prepare	Components	60 mm plate or 3.2 x 10^6 cells	100 mm plate or 8.8 x 10^6 cells	150 mm plate or 2 x 10^7 cells
PBS/Phosphatase Inhibitors	10X PBS	0.4 ml	0.8 ml	1.6 ml
	Distilled water	3.4 ml	6.8 ml	13.6 ml
	Phosphatase Inhibitors	0.2 ml	0.4 ml	0.8 ml
	TOTAL REQUIRED	**4.0 ml**	**8.0 ml**	**16.0 ml**
1X Hypotonic Buffer	10X Hypotonic Buffer	25.0 µl	50.0 µl	100.0 µl
	Distilled water	225.0 µl	450.0 µl	0.9 ml
	TOTAL REQUIRED	**250.0 µl**	**500.0 µl**	**1.0 ml**
Complete Lysis Buffer	10 mM DTT	2.5 µl	5 µl	10.0 µl
	Lysis Buffer AM1	22.25 µl	44.5 µl	89.0 µl
	Protease Inhibitor Cocktail	0.25 µl	0.5 µl	1.0 µl
	TOTAL REQUIRED	**25.0 µl**	**50.0 µl**	**100.0 µl**
*(Optional) Detergent	**TOTAL REQUIRED**	**1.25 µl**	**2.5 µl**	**5 µl**

*The addition of Detergent to the nuclear pellet may help with solubility of proteins, specifically those tightly associated with membranes or chromatin.

Step 1: Cell Collection

1. Aspirate media out of dish. Wash with 5 ml ice-cold PBS/Phosphatase Inhibitors. Aspirate solution out and add 3 ml ice-cold PBS/Phosphatase Inhibitors. If working with suspension cells, pellet cells and wash with 3 ml ice-cold PBS/Phosphatase Inhibitors and proceed to step 3.
2. Remove cells from dish by gently scraping adherent cells with cell lifter or, alternatively, pellet suspension cells by centrifugation. Transfer cells to a pre-chilled 15 ml conical tube.

   Note: Use of trypsin to detach cells can activate some signal transduction pathways.

3. Centrifuge cell suspension for 5 minutes at 200 x *g* in a centrifuge pre-cooled at 4°C.
4. Discard supernatant. Keep cell pellet on ice.

Step 2: Cytoplasmic Fraction Collection

Note: Make sure to verify cell lysis efficiency and release of the nuclei by comparing the appearance of cells before and after cell lysis (steps 1 and 2) using phase-contrast microscopy. Intact cells should appear as a dark central nucleus surrounded by a halo of less dense cytoplasm.

1. Gently resuspend cells in 500 µl 1X Hypotonic Buffer by pipetting up and down several times. Transfer to a pre-chilled microcentrifuge tube. Allow cells to swell by incubating for 15 minutes on ice.
2. Add 25 µl Detergent (for tissue, add 5 µl Detergent per every 100 µl Hypotonic Buffer) and vortex 10 seconds at the highest setting.
3. Check a small sample under the microscope to verify that cells have been efficiently lysed and that nuclei have been released. If the cells are not adequately lysed at this step, use an ice-cold Dounce homogenizer with a small-clearance (B) pestle to lyse the cells (refer to the Troubleshooting Guide in Section A of the Appendix, page 14, for further information).

   Note: Do not proceed with the centrifugation step until the cells are sufficiently lysed.
4. Centrifuge suspension for 30 seconds at 14,000 x *g* in a microcentrifuge pre-cooled at 4°C.
5. Transfer supernatant (cytoplasmic fraction) into a pre-chilled microcentrifuge tube. Store the supernatant at –80°C until ready to use. Use the pellet for nuclear fraction collection. OPTIONAL: Save a fraction of the supernatant (cytoplasmic fraction) for determination of fractionation efficiency by Western blot.

Step 3: Nuclear Fraction Collection

1. Resuspend nuclear pellet in 50 µl Complete Lysis Buffer by pipetting up and down. OPTIONAL: Add 2.5 µl Detergent to help solubilize membrane-associated nuclear proteins. A very viscous pellet may form and not completely resuspend. Alternatively, a Dounce homogenizer can be used (refer to the Troubleshooting Guide in Section A of the Appendix, page 14, for further information). Vortex 10 seconds at the highest setting.
2. Incubate suspension for 30 minutes on ice on a rocking platform set at 150 rpm.
3. Vortex 30 seconds at the highest setting. Centrifuge for 10 minutes at 14,000 x *g* in a microcentrifuge pre-cooled at 4°C. Transfer supernatant (nuclear fraction) into a pre-chilled microcentrifuge tube. OPTIONAL: Save a fraction of the supernatant (nuclear fraction) for determination of fractionation efficiency by Western blot.
4. Aliquot and store at –80°C. Avoid freeze/thaw cycles.

   Note: The presence of certain detergents or PMSF may interfere with the Bradford assay, thus perform a 1:50 or 1:250 dilution of your samples. Use the Complete Lysis Buffer as the blank with a similar dilution as the samples. Read the absorbance at 595 nm with a plate reader spectrophotometer (refer to the Appendix, Section A on page 13 for an example dilution series page 13). As an alternative, try Active Motif's ProStain™ Protein Quantification Kit (Catalog No 15001), which offers greater sensitivity and resistance to many contaminating agents.