Genetic and molecular studies in Hyperinsulinemic Hypoglycemia and congenital Polycystic Kidney Disease (HIPKD)

Anne Lore Kesselheim

UNIVERSITY COLLEGE LONDON

A thesis submitted to the University College London for
the degree of

Doctor of Philosophy

June 2018

Declaration

I, Anne Kesselheim confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgement

I would like to thank my supervisors Detlef Böckenhauer and Robert Kleta for offering me the opportunity and their trust to work on this project. I very much appreciate their guidance and support through the work presented in this thesis over the last three years. I am grateful to the J F Moorhead Trust for funding this work. I also thank all collaborators who were involved in this project. Thank you also to Michael Ludwig who encouraged me to move to London and has introduced me to Detlef and Robert.

I sincerely thank Ben Davies for giving me the chance to work in his lab at the Wellcome Trust Centre for Human Genetics at the University of Oxford where I could spend nine months to work on my CRISPR-Cas9 experiments under his supervision. His support was of great value for me and this project and I have learned a lot during this time. I also would like to express my special gratitude to the Davies' group who made me feel welcome from day one on and treated me as if I was part of their team. I in particular would like to thank Phalguni Rath for answering my countless questions around the CRISPR experiments, discussing results and the long hours spent in front of the FACS machine; Daniel Biggs for sharing his expertise in cell culture and his support whenever I needed it; Polinka Hernandez-Pliego for helping me genotyping hundreds of HepG2 clones.

My thanks also go to all my colleagues at the Centre for Nephrology for their constant support, advice, patience and listening. My special thanks to Vaksha Patel for her professional and personal support throughout this PhD and helping me to keep up my stamina; Horia Stanescu and Chris Cheshire for conducting the extensive bioinformatic analyses and sharing the data which have substantial impact on this project.

Thank you to my friends in London and Germany, in particular Steffi and Felice, for always being there for me especially when I was going through tough times. I am grateful to Grégory for his enduring patience, understanding and unconditional support. I cannot thank my family enough for their constant encouragement and understanding when I have not always been present in happy and sad situations.

I express my greatest thanks to my mother Ulrike and my sister Steffi for their love, endless support and unwavering belief in me. Thank you for always being there for me.

Finally, I dedicate this thesis to my father who left far too early and my beloved grandmother Lotti.

Abstract

Background

Hyperinsulinemic hypoglycemia (HI) and congenital polycystic kidney disease (PKD) are rare, genetically heterogeneous disorders. The co-occurrence (HIPKD) in 17 children from 11 unrelated families suggested a shared cause.

Methods

We ascertained the clinical phenotype and performed genetic studies. The effect of the identified shared mutation was assessed *in vitro*.

Results

All patients exhibited HI and enlarged polycystic kidneys. Whole genome linkage analysis in 5 informative families identified a single significant (LOD 6.5) locus on chromosome 16p13.2. A promoter mutation (c.-167G>T) in *PMM2* was found in all patients, either homozygous or *in trans* with *PMM2* coding mutations. Yet, typical systemic features of congenital disorder of glycosylation type 1a were absent and the diagnostic test of transferrin isoelectric focusing was normal.

The promoter mutation showed decreased transcriptional activity in patient kidney cells and impaired binding of the transcription factor ZNF143. *In silico* analysis suggests an important role for ZNF143 for the formation of a chromatin loop including *PMM2* that could affect tissue-specific transcription. In order to investigate this further in a chromatin conformation study a HIPKD cell model homozygous for the promoter mutation was generated with CRISPR-Cas9.

Conclusions

We report a rare disease characterized by the combination of hyperinsulinemic hypoglycemia and polycystic kidney disease. Our findings extend the spectrum of genetic causes for both disorders, provide insights into gene regulation and implicate glycosylation in the disease etiology. The identified promoter mutation appears critical for tissue-specific regulation of *PMM2* transcription, leading to an organ-specific phenotype and explaining *PMM2* pleiotropy.

Impact Statement

The work undertaken during this project expands the spectrum of genetic causes for HI and PKD, relevant for clinical research and diagnosis. This study also provides critical insights into gene regulation and underlines the importance of studying the impact of non-coding variants in human disease. Furthermore, it indicates that the higher organisation of the genome essentially contributes to correct gene regulation and that extensive research on this topic is of great value for both, basic and medical research.

The initial findings during this PhD have highlighted the dysfunctional gene regulation of the *PMM2* promoter variant. This substantially has contributed to postulate a hypothesis for the potential disease mechanism in HIPKD and eventually has helped to draw the attention onto gene regulation and 3D chromatin conformation. In this thesis, the generation of a HIPKD cell model with the CRISPR-Cas9 technology will take the project to its next level and will enable us to study the 3D conformation in health and disease on the example of HIPKD. Hence, this work will provide new important details of the spatial organisation of the human genome in general and will help to understand interactions within the genome and how these are affected when a mutation lies in a gene regulatory element.

In fact, about 30 million people in Europe are affected by rare diseases (same number in the US). 80% of rare disease worldwide are of genetic cause and often are life-threatening devastating disorders which frequently affect children.

The major issue with rare diseases is the lack of treatment for most of them. Research on rare diseases significantly improves the understanding of biology and molecular mechanisms within the human body in health and disease, which is crucial for the identification of drug targets with the major goal of developing new treatments. These may not only benefit patients with the rare disease that led to the identification of the drug, but also for more common diseases. Moreover, extensive research on rare diseases can improve diagnosis and thereby avoid unnecessary treatments and improve and personalise disease management.

Studies as conducted during this PhD with the aim to decipher a rare disease mechanism significantly contributes to the general understanding of biological processes important for human health and in particular the involvement of the non-coding genome and chromatin folding in human disease. More than a million enhancer elements are present within the human genome. Hence, it is not surprising that non-coding mutations which alter architectural modules can cause human disease. However, we are just at the beginning to understand how SNPs and structural variants impact genome folding. The combination of Hi-C data and gene editing tools such as CRISPR-Cas9 may allow for innovative new therapeutic inventions. These could be based on cell-related therapy and the insertion or elimination of architectural modules. Thus, studies which focus on the 3D genome itself and the impact of non-coding variants on genome folding are needed to significantly improve diagnostic and potentially therapeutic outcomes valid for research, medicine and industry.

Table of contents

CHAPTER 2: MATERIALS AND METHODS 68

List of Figures

List of Tables

Abbreviations

3D	3-dimensional
ABAT	4-Aminobutyrate Aminotransferase
AC	Adenylate cyclase
ADPKD	Autosomal dominant polycystic kidney disease
AP	Action potential
ARPKD	Autosomal recessive polycystic kidney disease
ASA	Allele-specific assay
ASN	Asparagine
ATP	Adenosine Triphosphate
BBS	Balanced salt solution
bp	Base pairs
BSA	Bovine serum albumin
BSS	Balanced salt solution
cAMP	Cyclic adenosin monophosphat, 3,5-cyclic AMP
CARHSP1	Calcium Regulated Heat Stable Protein 1
CD	Collecting duct
CDG-1a	Congenital disorder of glycosylation Type 1 A
CDH1	Cadherin 1
cDNA	Complementary DNA
CHF	Congenital hepatic fibrosis
ChIP	Chromatin immunoprecipitation
CKD	Chronic kidney disease
CNS	central nervous system
CRISPR	Clustered regularly interspaced palindromic repeats
CRISPR-HDR	CRISPR experiment with HDR donor
CRISPRed	CRISPR edited
crRNA	CRISPR RNA
CTCF	CCCTC-binding factor
DAG	Diacylglycerol
DCT	Distal convoluted tubule
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic acid
Dol-P	Dolichol-P linker
Dol-P-Man	Dolichol-P-Mannose

dPCR	Digital PCR
DSB	DNA double strand-breaks
EMSA	Electrophoretic mobility shift assay
Endo H	Endoglycosidase H
Epac	Exchange-protein directly activated by cAMP
ER	Endoplasmic reticulum
ESKD	End-stage kidney disease
FACS	Fluorescence-activated cell sorting
FAM	6-carboxyfluorescein
FBS	Foetal bovine serum
FCS	Fetal calf serum
fg	Femtogram
G1 phase	First growth phase in cell cycle
G2 phase	Second growth phase in cell cycle
GABA	Gamma-Aminobutyric acid
GCK	Glomerulocystic kidney disease
GCK	Glucokinase
GDH	Glutamate dehydrogenase
gDNA	Genomic DNA
gDNA	Genomic DNA
GEA	Gene expression assay
GEA	Gene expression assay
GFP	Green fluorescent protein
GFR	glomerular filtration rate
GlcNAc	N-acetylglucosamine
GLP-1	Glucagon-like peptide 1
GLP-1R	Glucagon-like peptide 1 receptor
GLUT2	Glucose transporter type 2
GLUT4	Glucose transporter type 4
GMO	Genetically modified organism
GPCR	G-protein coupled receptor
gRNA	Guide RNA
GSIS	Glucose-induced insulin secretion
GSIS	Glucose stimulated insulin secretion
HDR	Homology-directed repair
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HI	Hyperinsulinemic hypoglycemia/Hyperinsulinism
HIPKD	Hyperinsulinemic hypoglycemia and polycystic kidney disease
HNF4A	Hepatocyte Nuclear Factor 4 Alpha
IEF	Isoelectric focusing
IEQ	One islet equivalent
IP_3	Inositol trisphosphate

IP_3R	IP_3 receptor
iPSC	Inducible pluripotent stem cells
IR	Insulin receptor
IRS	Insulin receptor substrates
K_{ATP}	ATP-sensitive potassium channel
kb	Kilobase
kDa	Kilodalton
K_{ir} 6.2	inward rectifying potassium channel
KRBB	Krebs-Ringer bicarbonate buffer
LDL	Low-density lipoprotein
LiCl	Lithium chloride
LoH	Loop of Henle
M phase	Mitotic phase
Man-1-P	Mannose 1-phosphate
Man-6-P	Mannose 6-phosphate
MCT-1	Monocarboxylate transporter 1
MEF	Murine embryonic fibroblast
MGB	Minor-groove-binder
mM	Millimolar
MRI	Magnetic resonance imaging
MUT	Mutant
mV	Millivolt
NFQ	3' nonfluorescent quencher
ng	nanogram
NHEJ	Non-homologous end joining
nM	Nanomolar
NSB	No sample blank
nt	nucleotide
OCD	Oriented cell division
ON	Over night
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PC-1	Polycystin 1
PC-2	Polycystin 2
PC1 CTT	Polycystin 1 C-terminal tail
PCR	Polymerase chain reaction
PCT	Proximal convoluted tubule
PKA	Protein kinase A
PKC	Protein kinase C
PKC α	Protein kinase C α
PKD	Polycystic kidney disease
PLC	Phospholipase C
pM	Picomolar
PMA	Phorbol 12-myristate 13-acetate
PMM2	Phosphomannomutase 2

PNGase F	Peptide N-Glycosidase F
PP	Pancreatic polypeptide
PPIA	Peptidylprolyl isomerase A
PST	Proximal straight tubule
QC1	Quality control 1
QC2	Quality control 2
qRT-PCR	quantitative reverse transcription PCR
RD	Rare diseases
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RNPs	Ribonucleoproteins
ROI	Region of interest
RPF	renal plasma flow
rpm	Revolutions per minute
RRP	Readily releasable pool
RT	Room temperature
RyR	Ryanodine receptor
S phase	DNA replication (Synthesis phase)
SBS	Specific binding site
SEM	Standard error of the mean
SGLT2	Sodium-glucose co-transporter 2
sgRNA	Single guide RNA
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SNP	Single nucleotide polymorphism
SOC	Store-operated calcium channels
ssODNs	single-stranded oligodeoxynucleotides (homologous HDR DNA repair template)
SUR1	Sulfonylurea receptor 1
TADs	Topology associating domains
TALEN	Transcription activator-like effector nucleases
TCA cycle	Tricarboxylic acid cycle
TIDE	Tracking indels by Decomposition
TMEM186	Transmembrane Protein 186
TRPM2	Transient receptor potential cation channel 2
TRPP2	Transient receptor potential ion channel 2
trRNA	Trans-activating crRNA
UCSC	University of California, Santa Cruz
V	Volt
VDCC	Voltage-dependent calcium channels
wt/WT	Wildtype
ZNF	Zinc finger proteins
ZNF143	Zinc Finger Protein 143

Chapter 1: General Introduction

1.1 The nephron

The nephron represents the functional unit of the human kidney converting blood into urine in a process of filtration, reabsorption, secretion and excretion, which is “cleaning” the blood from waste products (metabolic waste, toxins) and maintaining homeostasis of blood volume and composition, essential for the human body. About 500,000-1,000,000 nephrons are present in each kidney. They are composed of two major compartments, the renal corpuscle and the renal tubule [1,2].

The renal corpuscle functions like a blood filter with the glomerulus forming a network of capillaries inside, surrounded by the Bowman’s capsule, a bowl-like structure formed of epithelial cells, collecting the glomerular filtrate. It is permeable for nearly all components of the blood apart from proteins, red and white blood cells and thrombocytes; these are restrained by specialised cells, the podocytes, which function as a barrier [3]. Around 300 litres of blood per day are passing through the glomerular capillaries, which derive from the afferent renal arteriole and around 180 litres of blood are filtered. The unfiltered blood leaves the glomerulus through the efferent arteriole [4]. The glomerular filtrate or primary urine contains all essential solutes ($Ca^{2+,}$ Na^+, K^+, Cl^-, Mg^{2+}, HCO_3^-, glucose, amino acids, vitamins and H_20) and also waste products and enters the lumen of the renal tubular system for further processing [5].

The tubular system is lined with epithelial cells, which function as a barrier between the blood and the tubular lumen where they facilitate the reabsorption of the essential solutes back into the bloodstream by active, energy-consumptive and passive transport processes through a well-orchestrated system of channels, pumps and transporters. The renal tubular epithelial cells are polar cells with an apical (lumen) and basal (blood) side. The proximal tubule epithelial cells are equipped with an apical brush border formed of microvilli, increasing the luminal cell surface dramatically for reabsorption processes.

The renal tubules are sectioned into different segments with unique functions defined by the altering specificity and morphology of the lining epithelial cells. The first segment is composed of the proximal convoluted tubule (PCT) and proximal straight tubule (PST) followed by the loop of Henle (LoH), distal convoluted tubule (DCT) and the collecting duct (CD). Apart from the LoH and the CD, which extends into the renal medulla, all parts of the nephron are embedded within the renal cortex, the outer zone of the kidney [6].

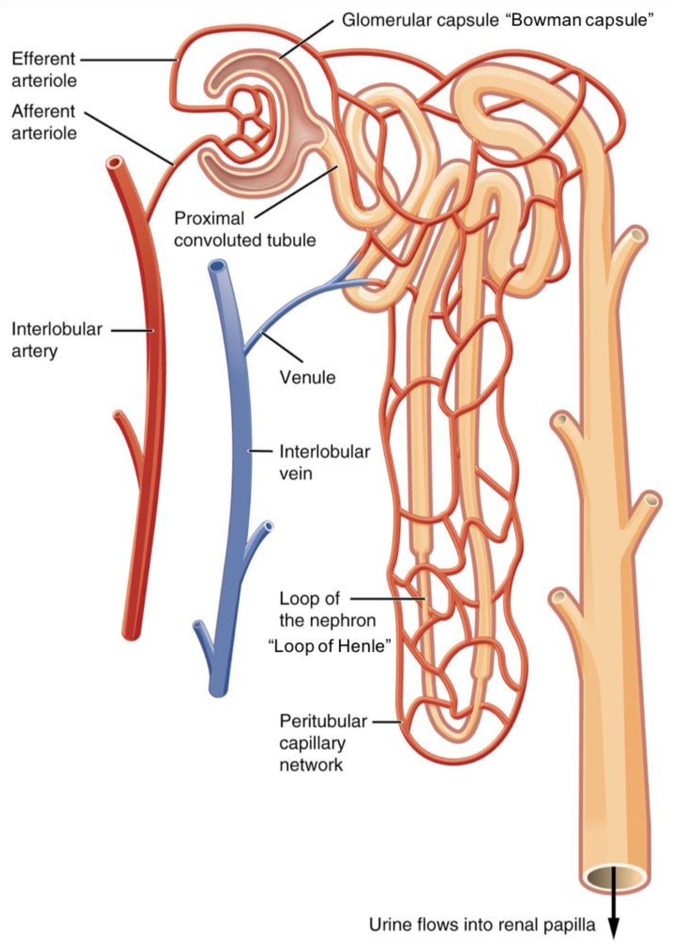


Figure 1: Anatomical structure of the nephron.
Blood enters the glomerulus through the efferent arteriole. The glomerular filtrate (primary urine) gets collected in the Bowman capsule and passes the renal tubular system for reabsorption of essential solutes, which into the circulation via the peritubular capillary network. Starting at the proximal convoluted tubule the filtrate passes further down through the loop of Henle and the distal convoluted tubule. The final filtrate (urine) finally arrives at the collecting duct for excretion. Figure taken from https://biologydictionary.net.

About 99% of the glomerular filtrate is reabsorbed via the tubular epithelium into the circulatory system through the peritubular arterioles located alongside the tubular system and are connected with the efferent arteriole. The DCT delivers the final filtrate, the urine, to the collecting ducts, which end in the renal collecting system from where the ureters drain the final urine to the urinary bladder for excretion [7]. Disturbances in these essential processes can result in general and specific metabolic imbalances and the study of these rare renal disorders help to explain the overall renal physiology. In particular, so-called renal tubulopathies have helped to understand the molecular physiology. But there are also more common conditions that illuminate the consequences of kidney dysfunction, for example chronic kidney disease (CKD), nephrolithiasis (kidney stones) and PKD.

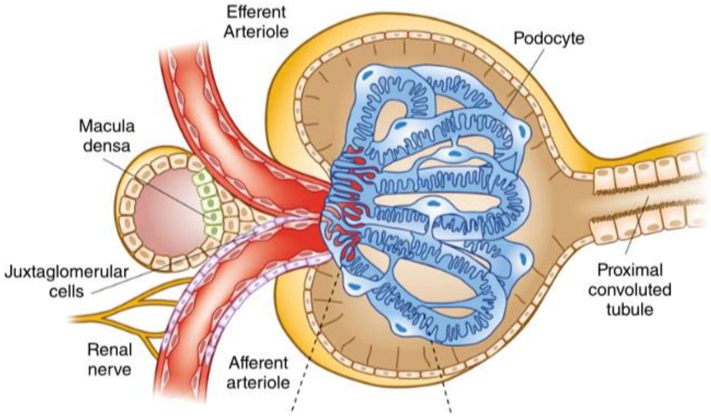


Figure 2: Anatomy of the renal corpuscle, cross section of the Bowman capsule.
A tuft of capillaries (glomerulus) surrounded by a single layer of epithelial cells of the bowman capsule. Afferent arteriole: Blood flow entering the capillaries of the glomerulus for filtration. Macula densa (MD, distal convoluted tubule) and juxtaglomerular cells (JC) regulate blood pressure. MD: Sodium detection, triggers contraction of afferent arteriole when sodium is high. JC: Renin secretion when sodium is low: Efferent arteriole: Unfiltered blood leaving the glomerulus. Proximal convoluted tubule: Tube lined with polar epithelial cells. Glomerular filtrate enters the renal tubular system for reabsorption and disposal of urine. Figure taken from http://cjasn.asnjournals.org/.

1.2 Rare diseases and how they teach us

Rare diseases (RDs) or also commonly called orphan diseases are typically Mendelian disorders. They are often severe, disabling, degenerative and life-threatening disorders of multisystemic nature, affecting different organ systems. There is no global agreement on when a disease is classified “rare”. In Europe for example, a disease is considered rare when less than one in 2,000 individuals are affected, whereas in the US this is the case when a disease is concerning fewer than 200,000 people. Worldwide the number of individuals affected by a RD was estimated between 350 and 400 million in 2014 [8-10]. Approximately 5,000–8,000 such highly heterogeneous RDs are known and about 80% of them identified being of genetic origin.

At least 150 of known RDs are rare kidney diseases, with 60–80 affected in 100,000 individuals in Europe and the US. Although such rare renal diseases are usually not lethal as good organ-replacement therapy is available, patients often suffer from multisystem features and have a poor quality of life, as the kidneys contribute to numerous homeostatic processes within the body. Renal replacement therapy, such as dialysis or transplant becomes necessary, when kidney function has deteriorated to such a degree that it is unable to maintain basic homeostatic function, a condition called end-stage kidney disease (ESKD). ESKD is due to genetically-caused renal conditions in at least 10% of all cases (adults and children) [11,12].

RDs cover a broad group and are organised in subgroups according to pathology, these for example include: Metabolic, autoimmune or neurodegenerative diseases, rare cancers, dysmorphic conditions and congenital malformations [9]. While individual RD are rare, patients affected by RD are frequent making research into RD an important endeavour. Moreover, research into describing their pathomechanisms significantly helps to better understand normal human physiology and pathology. The genetic causes of RDs are related to monogenic mutations, thus providing a lot of potential relevance in research, as the study of the phenotypic contribution of a single gene is easier to conduct compared to understanding the multiple genetic influences seen in complex disorders. Knowing the disease-causing gene and its protein product helps to decipher the principles of genetics, biochemical pathways and physiological processes, equally relevant also for common diseases and overall helping to identify potential targets for medical treatment and therapy for both, rare and common diseases [13]. The development of statins is an example for how studying a RD (homozygous familial hypercholesterolaemia) has helped to develop a universal drug to lower LDL (low-density lipoprotein) cholesterol within hypercholesterolemic conditions [14].

The small number of patients affected by an individual RD makes the development of effective drugs and treatments challenging for researchers, highlighting the importance of broad international collaborations and the formation of biobanks to gather clinical and genetic data as well as biospecimens. Such biobanks are of scientific and clinical relevance, contributing to the understanding of biology, the causes of disease or the identification of new disease genes, mutations and biomarkers that can help develop improved diagnostic and therapeutic approaches [9,15]. Moreover, RDs provide a lot of potential for innovative therapy approaches, such as gene therapy or enzyme-replacement therapy. In 2011 for example, 20% of all innovative products released to the market in Europe were specific for the treatment of RDs [16].

1.3 Cystic kidney diseases

Cystic kidney diseases are highly heterogeneous and can have genetic, developmental or acquired causes. They are often part of a rare syndrome appearing together with additional features in other tissues. Interestingly, many mutations causing renal cyst formation are often found in genes associated with the cilia centrosome complex and ciliary dysfunction is the shared pathomechanism (ciliary hypothesis) of such diseases. Therefore, cilia-related cystic kidney diseases belong to the group of ciliopathies. However, renal cysts also develop through mutations in other genes encoding for proteins which are not directly involved in cilia structure or function [17]. The most common ciliopathies are outlined below. Table 1 provides an overview of ciliopathies, their cyst-causing proteins and where these operate within the body.

Table 1: Ciliopathies and their causative genes.
Encoded proteins and their function. Table modified after Loftus and Ong [17].

Disease	**Gene**	**Protein**	**Function**
ARPKD	*PKHD1*	Fibrocystin/Polyductin	Exact function unknown. Expressed in apical membrane, primary cilia, and centrosomes. Interacts with polycystin-2
ADPKD	*PKD1*	Polycystin-1	Mechanosensitive or ligand-activated receptor; functionally complexed with polycystin-2. Expressed in basolateral plasma membrane, cilia, centrosomes
	PKD2	Polycystin-2	A non-selective Ca^{2+} channel. Functionally complexed with polycystin-1. Expressed in ER, basolateral plasma membrane, cilia, centrosomes
Nephronophthisis	*NPHP1-13*	Nephrocystin 2 (Inversin) Nephrocystin 3-6 GLIS2 protein Nephrocystin 8-9 CCCAP Meckelin IFT139 IFT144	Involved in primary cilia function and the cell cycle. Some localize to cell–cell junctions or focal adhesions
Medullary cystic kidney disease	*MCKD1*		Not known
	MCKD2 (UMOD)	Uromodulin	Possible role in maintaining integrity of thick ascending loop of Henle. Localization to cilia and centrosomes
HNF1β mutations		Transcription factor 2 (TCF2)	Regulates the transcription of several key cystic genes
Von Hippel-Lindau disease	*VHL*	VHL protein	Tumour suppressor gene acting through HIF and non-HIF-dependent pathways. Localization to cilia and role in cilia length control
Tuberous sclerosis complex	*TSC1*	Hamartin	Complexes with tuberin to inhibit mammalian target of mTOR
	TSC2	Tuberin	Complexes with hamartin. Localizes to primary cilia and regulates cilia length

1.3.1 Autosomal dominant polycystic kidney disease (ADPKD)

Inherited ADPKD is the most abundant monogenic kidney disorder and cause for renal failure (ESKD) with 1 in 400 to 1 in 1,000 cases in children and adults [18]. ADPKD belongs to the family of ciliopathies.

The detailed description of ADPKD shall provide a representative example of the complexity of PKD, its genetic causes and pathogenesis. Further examples will be outlined in brief.

1.3.1.1 Phenotype

Increasing numbers of fluid-filled renal cysts are clinically notable whilst the condition progresses with thousands of cysts of different sizes causing a massive enlargement and deformation of the kidneys at the advanced stage of disease [19]. ADPKD is a systemic disorder expressing abnormalities in different tissues, for example cysts developing in liver and pancreas, but can also affect the connective tissue (e.g. aortic dissection and intracranial aneurysms) and hypertension is a common symptom [20]. Studies on cellular and biochemical level have revealed that alterations of basement membrane and epithelial cells (renal and non-renal) are predominant features in ADPKD and the disease is therefore considered a disorder of epithelial cell growth [21,22].

1.3.1.2 Genetic cause and the encoded proteins

ADPKD is caused by mutations in either the *PKD1* or the *PKD2* gene, with mutations in *PKD1* (MIM 601313) being the most common with about 85% of all cases reported compared to 15% in *PKD2* (MIM 173910). To date, are 1,273 and 202 pathogenic variants known for *PKD1* and *PKD2*, respectively, which are categorised as definitely pathogenic, highly likely pathogenic or likely pathogenic in the ADPKD mutation database (http://pkdb.mayo.edu/). ADPKD mutations appear with high allelic heterogeneity, mainly leading to missense alterations (31% of all *PKD1* mutations) [23]. Overall, the expression of the disease is milder when *PKD2* is the cause and also the disease onset of *PKD2*-related ADPKD happens later in life and progression is slower until reaching the ESKD state [23,24]. A combination of mutations (trans-heterozygosity) in both genes are known and develop a more severe form of the disease [21].

PKD1 encodes for polycystin 1 (PC-1) and is located on the p-arm (short arm) on chromosome 16 (1994) whereas *PKD2* lies on chromosome 4 (long arm, q-arm) and translates into the protein polycystin 2 (PC-2) [25].

Polycystin 1 (PC-1)

PC-1 is a multifunctional protein involved in cell-cell/matrix adhesion, ciliary function and cell signalling processes [26,27]. It is found in various tissues (e.g. heart, liver) including the tubules in the kidney, where it is located in the cilium, the plasma membrane (lateral domain), cell-adhesion complexes (tight junctions, adherens junctions, desmosomes) and urinary exosomes. PC-1 expression is highest in the fetal epithelium but still detectable in the matured tissue [28].

Polycystin 2 (PC-2)

PC-2 belongs to the transient receptor potential ion channels protein family and is also called TRPP2. It is membrane-associated and found on intracellular compartments (endoplasmic reticulum (ER), Golgi apparatus), primary cilia, both apical and basolateral plasma membrane of polar epithelial cells and like PC-1 also in urinary exosomes. It acts as non-selective ion channel permeable for calcium, potassium and sodium but is predominantly present at intracellular compartments for the release of intracellular calcium stores [29]. Interactions between PC-1 and PC-2 in an ion channel signalling complex have been observed suggesting that PC-1 stimulates PC-2 to conduct its ion channel capacities [30]. Mutations in both genes have led to the disturbance of intracellular calcium regulation promoting cell proliferation [31]. Figure 3 depict the localisation of PC-1 and PC-2 in the cell.

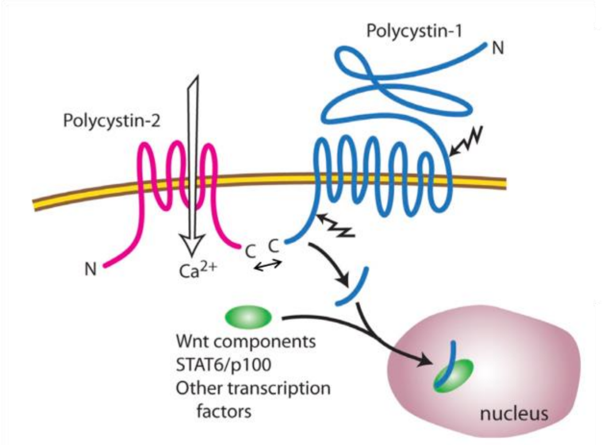


Figure 3: Topology of membrane proteins polycystin-1 (PC-1) and polycystin-2 (PC-2).
Both proteins are found on the plasma membranes of the cellular adhesion complex (PC-1), apical and basolateral membranes, membranes of ER and Golgi (PC-2) and the shaft and basal body of the primary cilium of epithelial cells (PC-1 and PC-2). PC-1 (blue) is a 450 kDa protein with 11 transmembrane domains and a large extracellular N-terminal domain. The short C-terminus tail is on the cytosolic side of the membrane PC-1 has different cleavage sides (toothed arrows). PC-2 (magenta) is a smaller transmembrane protein with 110 kDa spanning the membrane with 6 transmembrane domains. Both, N-terminal and C-terminal domains face the cytosol. PC-2 functions as a calcium channel (TRPP2). With their C-terminal tails PC-1 and PC-2 interact (double-headed arrow) with each other exerting distinct regulatory functions. Figure modified after Chapin 2010.

Function of polycystins

Mechanosensation and cilia

Apart from the intercalated cells, all renal tubular epithelial cells within the nephron and collecting ducts have a primary cilium on their apical side, which functions as a sensor. The cilium is formed of nine nonmotile microtubules. The external part has a microtubule core (axoneme) and with the basal body, derived from the intracellular mother centriole it is anchored with the cell. It reaches the cell lumen to exert its mechanosensory function mediating extracellular signals inside the cell (signal transduction). The cilium responds to either mechanical- or fluid flow-based stimulation with bending and this inducing calcium influx through PC-2 (TRPP2 calcium channel) [32,33]. PC-1 located on the cilium is thought to sense such environmental stimuli with its large extracellular domains and after conformation change activates PC-2 for calcium influx. This was confirmed in Pkd1 null mouse cells and human ADPKD kidney epithelial cells which failed to respond to mechanic stimuli [34]. In healthy cells the resulting calcium influx then triggers calcium release from the intracellular stores and the increasing calcium levels are transduced into a downstream Ca^{2+}-signal to control transcriptional activity in the nucleus [35]. Hence, mutations in any of the polycystins can change their response to fluid flow leading to dysregulation of target genes and their downstream cellular processes (tissue homeostasis, cell proliferation, apoptosis etc.).

Involvement in signalling pathways

G-protein signalling: PC-1 has G-protein coupled receptor capacities (GPCR) as some domains in its C-terminal tail (PC1 CTT) share sequence homology with GPCR and *in vitro* activate $G\alpha_{i/0}$proteins [36]. G-protein signalling pathways are involved in the regulation of cellular processes like fluid secretion, cell proliferation, -polarity and -differentiation present in cyst formation [37]. Downstream signalling effects of PC1 CTT are found within the JNK/AP-1 (c-Jun N-terminal kinase and activator protein-1) and calcineurin/NFAT (nuclear factor of activated T-cells) pathways. PC1 CTT has the ability to activate the JNK/AP-1 pathway through protein kinase C α (PKC α), small G-proteins (Cdc42 and Rac1) and $G\alpha$ and $G\beta\gamma$ subunits. Cell cycle regulation, cell proliferation and differentiation as well as apoptosis are processes controlled by this pathway and many of those are subject of PKD pathogenesis [38]. In the calcineurin/NFAT pathway PC1 CTT activates phospholipase C (PLC) mediated via $G\alpha_q$.

$G\alpha_q$-coupled PC1 CTT-dependent activation of PLC mediates an increase of intracellular calcium and leads to calcineurin (serin-threonin phosphatase) activation. Calcineurin eventually dephosphorylates NFAT, inducing its translocation into the nucleus for regulation of its target genes with documented involvement in cell development and adaptation in many cell types. NFAT and PC-1 are co-expressed in mice renal tubular epithelial cells during development and adulthood suggesting a shared role in a signalling pathway [39,40].

Growth regulation/Cell-cycle regulation: PC-1 and PC-2 are directly involved in cell cycle regulation. The related signal transduction pathways are described in brief. **JAK-STAT pathway:** PC-1 activates the Janus kinase (JAK) and the transcription factors STAT1 and STAT3, together causing the upregulation of p21 (kinase), which slows down or arrests the cell cycle at G_0/G_1 phase. **Id2:** The transcription factor Id2 (helix-loop-helix protein, HLH) regulates cell proliferation and differentiation by downregulating p21. Phosphorylated PC-2 directly interacts with Id2 and regulates its nuclear localisation, thus contributes to the regulation of such cellular processes. PC-1 regulates the phosphorylation of PC-2. Id2 is lacking a nuclear localisation signal needs to complex with a second HLH protein (E47). PC-2 is able to bind this complex and keeps it outside the nucleus, hence p21 transcription promotes cell cycle arrest. In the case of *PKD1* or *PKD2* mutations, PC-2 interaction with Id2 is restricted, explaining increased cell proliferation in ADPKD [41]. **STAT6/P100:** PC-1 cleavage is regulated by fluid flow. Under normal flow conditions in the renal tubules non-cleaved PC-1 and P100 (co-activator) are co-localised on cilia of epithelial cells. Together PC-1 and P100 translocate the transcription factor STAT6 from the nucleus to the cilia preventing STAT6-regulated transcription. Under "no flow" conditions (cysts) PC-1 gets cleaved at its C-tail (PC1 CTT$_{14kDa}$) and together with P100 translocates STAT6 to the nucleus enhancing transcription of STAT6 target genes promoting cell growth. P100 overexpression and high levels of nuclear STAT6 have been confirmed in ADPKD patients [42]. **mTOR**: mTOR is a kinase and the key protein in the mTOR pathway stimulating protein synthesis and cell proliferation. PC1 CTT interacts with TSC2 (tuberin) a regulator of mTOR. Mutations in TSC2 have been described to cause renal cysts (Shillingford et al. 2006). Active membrane-bound PC1 CTT has the ability to inhibit phosphorylation of TSC2 which retains TSC2 bound to the membrane repressing mTOR signalling. In the case of mutated PKD1 the mTOR pathway is active constantly causing uncontrolled cell growth [43]. **Wnt pathway:** Wnt signalling is involved in the regulation of cellular processes, such as embryonic development, cell fate, cell proliferation, cell migration and planar cell polarity (cell shape). The latter is subject to the non-canonical (β-catenin independent) Wnt pathway, the others are regulated during the canonical Wnt pathway (β-catenin dependent). The major Wnt ligands are glycoproteins. **Canonical Wnt pathway:** β-catenin is a subunit of the cadherin complex involved in cell-cell adhesion but also regulates gene transcription. Binding of Wnt ligands to β-catenin promotes its translocation into the nucleus activating transcription of its target genes. Cleaved PC1 CTT binds and co-translocates together with β-catenin into the nucleus. Bound in this complex β-catenin-regulated transcription is inhibited. In ADPKD patients and *PKD1* null cells Wnt signalling is upregulated and downstream effects enhanced [44]. **Non-canonical Wnt pathway:** The non-canonical Wnt pathway regulates planar cell polarity important for tubule formation in the kidney. The cell division of renal tubular cells has a pre-set parallel direction (oriented cell division, OCD), supporting lengthening rather than widening of the tubules. In PKD a loss of OCD is suggested promoting the expansion of the tubule diameter, which could mark a precursor-state of cystogenesis [45,46]. Another interesting model is that urine flow-induced signalling functions as a switch between the canonical and non-canonical pathway. Urine flow rates increase the expression of inversin, a protein involved in tubular left-right axis development. Upregulated inversin reduces β-catenin levels, which likely terminates the canonical Wnt pathway to initiate the β-catenin-independent Wnt signalling for polarisation of the renal epithelial cells needed for correct tubule development. In this way the two Wnt pathways could manage shape and function of the nephron. In mice, lacking functional inversin cyst formation was observed, as a result of unhindered canonical Wnt signalling [47]. In ADPKD patients dysfunctional *PKD1* potentially prevents the switch between the two Wnt pathways due to its inability to repress β-catenin-driven transcription resulting in enhanced cell growth. Inversin also interacts with nephrocystin proteins and is the causative gene for nephronophthisis type II, a disease expressing renal cysts as a major feature. Figure 4 summarises the signalling pathways PC-1 and PC-2 take part in.

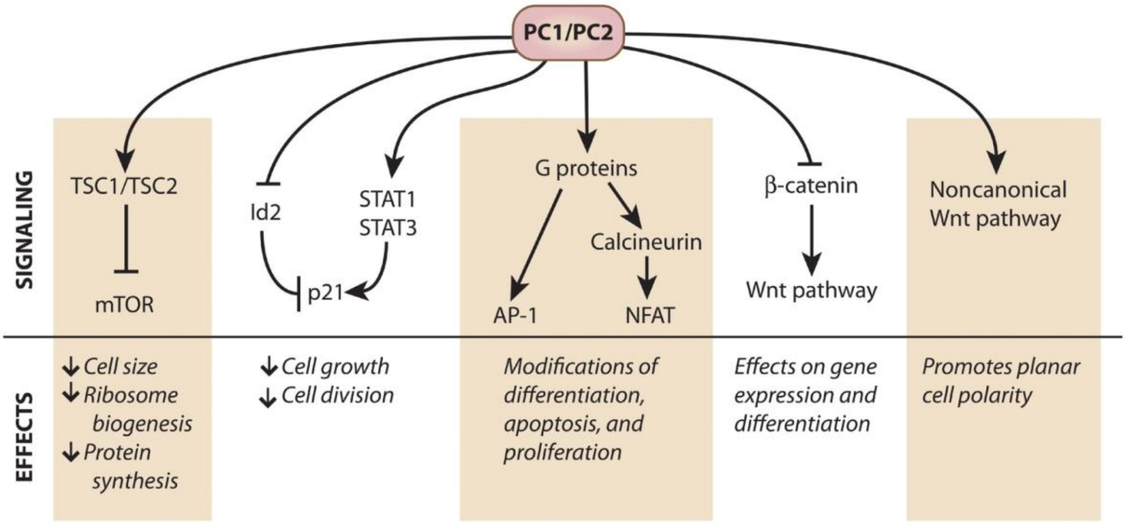


Figure 4: Signalling pathways with polycystin 1 and polycystin 2 involvement.
Figure taken from Chapin 2010.

Cystogenesis

Cyst formation is associated with major cellular processes becoming dysfunctional, such as enhanced cell proliferation and apoptosis of cyst-forming epithelial cells, loss of epithelial polarity and de-differentiation (regenerative process) and cell/matrix interactions. Phenotypic changes are also observed as the cyst-forming epithelial cells instead of exerting their absorptive function starting to become secretory.

In ADPKD, cysts develop from epithelial cells and can arise in any tubular section of the nephron due to cell proliferation and inflow of glomerular filtrate. Whilst initially they stay attached to their parental tubule, when reaching a diameter of approximately two millimetres they detach allowing only cAMP-mediated transepithelial chloride secretion into the cystic lumen through the basolateral Na-K-Cl cotransporter and the apical cystic fibrosis transmembrane conductance regulator, short CFTR [48,49].

Growth factors such as the epidermal growth factor (EGF), the EGF receptor (EGFR) and the transforming growth factor alpha (TGF alpha) are overexpressed in the cyst-forming epithelium promoting cell proliferation and hence cyst formation. In healthy tissue, cAMP mediates the inhibition of the MAPK/ERK pathway (cell-cycle regulation) restricting cell growth. In cyst-forming cells, it does the opposite. It highly accumulates due to reduced intracellular calcium concentrations, enhancing the proliferation of cyst-forming epithelial cells by activating the MAPK/ERK pathway [50]. The knock-out of apoptosis-preventing genes and the overexpression of apoptosis-promoting genes both also support cell proliferation and cyst formation as mouse models have shown [51]. Figure 5 depicts the process of cytogenesis on all levels.

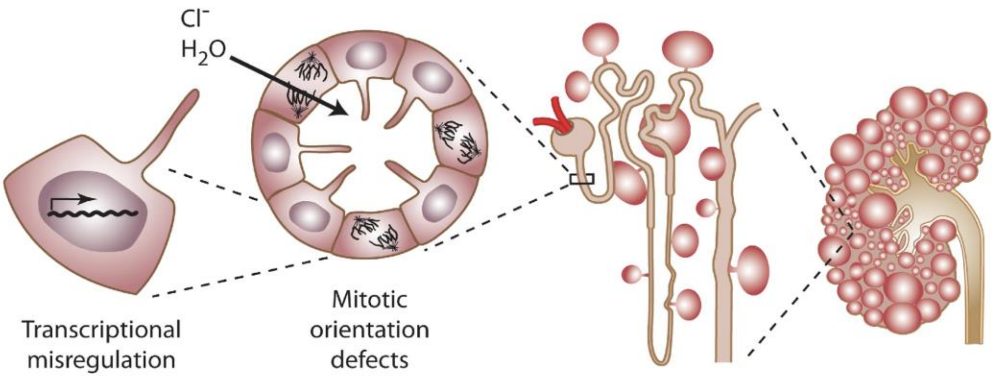


Figure 5: Cystogenesis on cellular level, the nephron and entire kidney.
Gene defect on PKD1 or PKD2 alter transcription changing cell proliferation and ion secretion resulting in fluid-filled cysts of different sizes. Cysts protrude from the tubules in different nephrons and all together changing the parenchyma of the whole organ. Figure taken from Chapin 2010.

On average a human adult kidney hosts about 1 million nephrons, however in ADPKD patients, cysts arise from only approximately 1,000 nephrons. If a germline mutation was the sole cause for cyst formation, consequently cysts derived from all nephrons would be expected. However, on the basis of analysis of *PKD1*- and *PKD2*-related cysts, the pathogenesis of cyst formation in ADPKD is thought to be a two-hit process (two-hit model Figure 6) [52,53]. In first place, a heterozygous germline mutation occurs on one allele of the ADPKD genes, which is present in all cells but only induces cyst formation in a minor number of renal tubular cells. Hence, an additional mutation of somatic origin – the second hit – on the other “normal” allele is needed to increase the functional loss or dysfunction of the polycystins resulting in extended cystogenesis. The second hit affects the wild type (wt) allele of one of the ADPKD genes and causes the loss of heterozygosity in those cells [54]. Despite being widely accepted the two-hit model is also criticised due to the observation, that in up to 60% of *PKD1*- or *PKD2*-related cysts a second somatic mutation was absent, but this might also be explained by technical limitations of the methods used back in the 1990s [23,55].

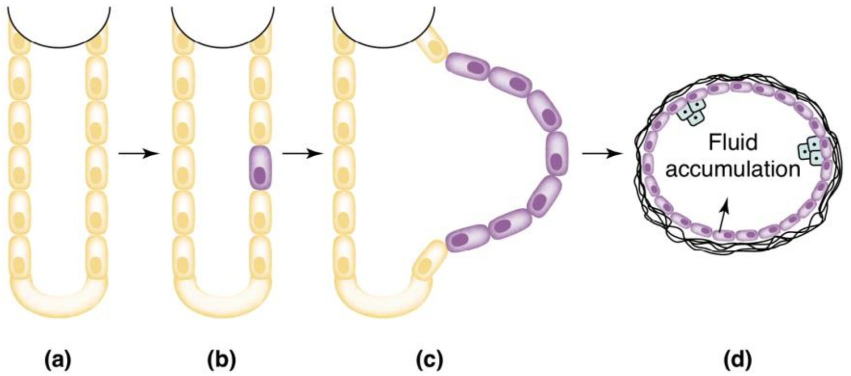


Figure 6: Two-hit model of cyst formation in ADPKD.
(a) Renal tubular cells all having the same PKD mutation in heterozygous state (first hit). (b) Somatic mutation on the wt allele randomly occurring in one of the cells (second hit). (c, d) Clonal proliferation and tubular dilatation of the now compound heterozygous cell (two hits) and cyst development with ion secretion into the cyst lumen. Figure taken from Pei 2001 [30].

Leonhard et al. in 2015, observed cyst formation in an ADPKD mouse model and raised the idea of a snowball effect of cyst formation proceeding from one initial cysts, which itself originates from a somatic second hit mutation in one cell, prompting the development of further cysts due to possible enhanced PKD-signalling [56]. Another more recent viewpoint of cystogenesis is described by the threshold and dosage-sensitive model, initially proposed by Lantinga-van Leuwen et al. in 2004 whilst working with mice with hypomorphic *PKD1* or *PKD2* 57. They propose that cystogenesis is not only provoked by a combination of a germline and somatic mutation but also might be dependent on gene dosage. Meaning, when reduced expression of *PKD1* and *PKD2* reaches a certain level this then drives cyst formation (haploinsufficiency) and that reduced expression can be induced not only by genetic causes but also due to environmental or stochastic (random fluctuations in the expression of individual genes) factors [58]. To date and despite some variations in available data, the clinical threshold level of remaining PC-1 and PC-2 function necessary to maintain cell function is estimated at around 40% [59].

1.3.2 Autosomal recessive polycystic kidney disease (ARPKD)

Autosomal recessive polycystic kidney disease is predominantly seen in neonates and children with in incident of 1 in 20,000 live births [18].

Phenotype

ARPKD typically represents a combination of renal cysts in both kidneys and congenital hepatic fibrosis (CHF) or Caroli disease [60]. Renal cysts are already notable in the unborn child. Fluid-filled cysts form mainly in the collecting duct and stay attached to the parental tubule. As a consequence, the collecting ducts in affected individuals elongate and this causes poor corticomedullary differentiation resulting from the extension of the collecting ducts from medulla to cortex. The cysts in ARPKD are usually quite small, below 3 mm, but with increasing numbers the kidneys enlarge [61].

Histologically the main feature of CHF in ARPKD patients is a ductal plate malformation with portal fibrosis, bile duct proliferation and hypoplasia (underdevelopment) of portal vein branches with hypertension as a consequence. Liver involvement in form of the Caroli phenotype shows cysts within the intrahepatic bile ducts [62].

Genetic cause and encoded protein

ARPKD belongs to the group of ciliopathies and is caused by mutations in only a single gene, *PKHD1*, which is located on chromosome 6p12.3-12.2 [63]. To date, nearly 750 different mutations have been described (www.humgen.rwth-aachen.de) with severe and milder onset depending on the two mutations on the *PKHD1* alleles Whereas two truncating mutations are lethal usually during the perinatal period, one or two missense mutations present a milder form of the disease [64]. *PKHD1* encodes for the large receptor-like 450 kDa transmembrane protein fibrocystin (polyductin), with a single transmembrane span, a large extracellular N-terminus and its C-terminal domain facing the cytoplasm [65].

Fibrocystin, its function and cystogenesis

Fibrocystin is found within the kidney in the cortical and medullary collecting ducts, the liver in the biliary ducts and pancreas in the pancreatic ducts. In renal epithelial cells it is predominantly a ciliary protein and localises to the cilia membrane, the basal body, but also the apical plasma membrane as well as the mitotic spindle [65,66]. Like other cystogenic proteins (e.g. *PKD1* and *PKD2*) this suggests fibrocystin is involved in cellular processes, such as cell proliferation and apoptosis. Mutations are likely to negatively influence the regulation of such processes and this may initiate cystogenesis due to uncontrolled cell growth as seen in other ciliopathies. However, the exact function of fibrocystin is still unknown and its role in the human remains speculative [17,67].

Besides its ciliary targeting signal, fibrocystin has a nuclear localisation signal (NLS) and a polycystin-2 binding domain. This implies a possible translocation into the nucleus as well as an interaction with polycystin-2 (PC-2). Both may indicate involvement of fibrocystin in nuclear signal transduction, independently or in synergy with PC-2, eventually regulating structural cellular processes [68]. Evidence of nuclear translocation has been described in a study by Hiesberger et al. [69]. They were able to show proteolytic cleavage within the C-terminal domain of fibrocystin in dependence of intracellular calcium levels and protein kinase C (PKC) activity. The cleaved product carries the NLS and conclusively should be capable to translocate into the nucleus, suggesting a role in signal transduction alike other proteins known from their dysfunction in ciliopathies (e.g. PC1 CTT).

1.3.3 Glomerulocystic kidney disease

Glomerulocystic kidney disease (GCK) is a rare form of PKD. Glomerular cysts are defined as dilatation of the Bowman space more than two to three times the normal size. When at least 5% of the kidney is interspersed with glomerular cysts, the phenotype is recognised as a renal condition [70]. Glomerular cysts are difficult to recognise. To distinguish them from tubular renal cysts, the appearance of the glomerular tuft - as long as not degraded - within the cysts helps for indication [71]. GCK commonly occurs as a feature in a genetically caused syndrome and the affected gene often encodes for a protein associated with the primary cilium or the centrosome - the very common pathogenesis in PKD as discussed above. Dysregulation of transcription in signalling pathways controlling renal development seems also to be the mechanism for cystogenesis, like seen in typical ciliopathies when GCK is of hereditary nature [72]. However, the causes and origins to develop GCK are quite heterogeneous and of hereditary and non-hereditary origin and are categorised in five subtypes after Lennerz et al. [71]. Table 2 lists the five subtypes of GCK.

Table 2: Categorisation and key features of GCK conditions Lennerz et al. 2010 [79].

Type	**Name**	**Common feature**	**Subcategories**	**Key features**
I	**GCK in PKD**	PKD genes	ADPKD (*PKD1*, *PKD2*)	Inheritance pattern: Age, shape, size, bilateral, liver and/or other organ cysts
			ARPKD (*PKHD1*)*	
II	**GCKD = hereditary GCK**	GCKD genes	*UMOD*	Inheritance pattern and GCKD specific features
			HNF1β	
			NOS	
III	**Syndromic GCK**	Recognised syndrome, **absence of dysplasia**	Too many to list, please see Table 2 in Lennerz et al. 2010.	Malformation syndromes with typical manifestations ±**obstruction**
			Localised GCK	Unilateral or segmental
IV	**Obstructive GCK (non-hereditary)**	Urinary tract abnormalities	With dysplasia	Smooth muscle cell collarettes, cartilage, IM
			Without dysplasia	Urinary tract abnormalities, uni-or bilateral
V	**Sporadic GCK**	No familial pattern, No dysplasia, No syndrome, No obstruction	Ischemic GCK	Vascular changes, unilateral, age
			drug-induced pattern	Bilateral, history of drug exposure

Abbreviations: ADGCKD: Autosomal dominant glomerulocystic kidney disease; ADPKD: Autosomal dominant polycystic kidney disease; ARPKD: Autosomal recessive polycystic kidney disease; FHGCKD: Familial hypoplastic glomerulocystic kidney disease; GCK: Glomerulocystic kidney; GCKD: Glomerulocystic kidney disease, synonymous with heritable/familial GCK; GCKD (NOS): Glomerulocystic kidney disease, not otherwise specified; IM: immature mesenchyme; PKD: Polycystic kidney disease; PKD1-2: ADPKD genes; PKHD1: ARPKD gene; TCF2: FHGCKD gene encoding hepatocyte nuclear factor–1 β (HNF1β); UMOD; Uromodulin, ADGCKD gene

*indirectly due to mutations in HNF1β inhibiting *Pkhd1* expression in mice [73].

GCK (Type III, syndromic GCK) is presented in at least 43 malformation syndromes. The most common of them is tuberous sclerosis, but also Zellweger's syndrome, nephrotic syndrome and von Hippel Lindau disease are well known [71].

1.4 The Islets of Langerhans of the human pancreas

The pancreas is one of the key organs of the digestive tract with an exocrine and endocrine function in the human body where it plays a major role in nutrient digestion and glucose homeostasis (energy homeostasis), respectively. The majority of its mass is dominated by the exocrine tissue which harbours the acinar, the exocrine, cells. They produce the pancreatic juice, a liquid which contains the digestive enzymes, such as amylase, pancreatic lipase and trypsinogen crucial for macro digestion, which they pass directly into the small intestine via the pancreatic duct. Within the exocrine pancreatic tissue, little cell clusters are embedded, formed of five different hormone-secreting cell types representing the endocrine part of the organ [74]. They are surrounded by a double-basement membrane and besides consisting of the endocrine cells also vascular, immune and often neuronal cells complete their cytoarchitecture [75]. A healthy human pancreas accommodates approximately 3.2 million islets with an average diameter of 110 µm, which in total occupy only 1-2% of the total mass of the pancreas [76,77]. However, they receive up to 20% of the blood supply [78].

These cell clusters were first described in 1869 by Paul Langerhans and therefore are called the islets of Langerhans but also commonly described as human pancreatic islets. Those islets are recognised as highly organised endocrine microorgans which are essential for the regulation of glucose homeostasis, sensitively responding to changes in blood glucose levels by secreting pancreatic hormones in an endocrine, paracrine and autocrine manner [74,79]. The total number of cells in an islet equivalent (IEQ, volume of 150-µm diameter sphere) is estimated being around 1,560. The most prominent and most important endocrine cell types for glucose homeostasis hosted by the pancreatic islets are the α-cells (glucagon) and β-cells (insulin) with about 20-30% and 60-70% per islet, respectively. The remainder is shared by the remaining three cell types, δ-cells (somatostatin), PP or γ-cells (pancreatic polypeptide) and the G or ε-cells (ghrelin) [76,80]. The pancreatic hormones secreted by these cells are associated with a distinct function contributing to glucose homeostasis, with insulin and glucagon at the forefront.

Pancreatic islets are quite variable in terms of their layout and relative population of the endocrine cells within one cell cluster, which differs from islet to islet, from individual to individual and from species to species. The cellular arrangement within human islets is, differently to rodent islets, of a heterogeneous nature with no distinct recognisable pattern. This implies an apparent absence of hierarchy between the different cell types [80]. Moreover, human islets contain numerous blood vessels and the endocrine cells are located alongside without a distinct order, in contrast to mice, suggesting that the direction of the blood flow does not majorly influence particular interactions [81]. β-Cells are directly adjoined by α-cells and/or δ-cells and this association is so strong that they often remain coupled to the α-cells even when in suspension [82]. These observations strongly suggest this unique cell arrangement in human islets is of functional nature and indicates paracrine cell signalling. Paracrine signalling is depended on a close proximity between cells as many signalling molecules otherwise get degraded quite quickly within the interstitium and the blood, therefore this process is independent from vascular perfusion – both these requirements are fulfilled by the human pancreatic islets cytoarchitecture [75,83]. Besides the secreted hormones, there are additional signalling molecules, mainly neurotransmitters (e.g. glutamate, ATP, acetylcholine and GABA) active within the islets, contributing to the paracrine and autocrine processes and adding on a complex picture of signalling pathways within these small microorgans.

Since the early 2000s, islets transplantation is successfully applied to treat type 1 diabetes mainly due to the insulin-producing potential and functional significance of the β –cells [84].

1.5 Peptide hormones secreted by the Islets of Langerhans

1.5.1 Pancreatic glucose homeostasis by insulin and glucagon - principle

Pancreatic islets are glucose sensors which react to changes in glucose concentration through a number of metabolic processes accomplished by the endocrine cells. The main drivers balancing blood glucose levels are insulin and glucagon, secreted by the β-cells and α-cells of the endocrine pancreas, respectively. Whilst glucagon increases the blood glucose levels, insulin does the opposite [85]. In this way, the optimal blood glucose levels are kept within a tight range of 4-6 mM (~90 mg/mL). During fasting periods, sleep for example or between meals, when glucose levels decrease glucagon gets released by the α-cells to guarantee an immediate glucose supply (endogenous glucose) from the degradation of glycogen in the liver (hepatic glycogenolysis) to prevent hypoglycemia [86]. The β-cells, by contrast, respond to increasing blood glucose levels from glycogen conversion and food intake with the secretion of insulin [87]. Released into the blood, insulin acts on different tissues. It binds to its specific receptors on adipose and muscle tissue allowing the uptake of exogenous glucose by these tissues from the blood and this in turn lowers the blood glucose levels, preventing the risk of hyperglycemia [88]. But it also stimulates the liver to produce more glycogen from glucose (glycogenesis) to maintain the glycogen stock needed for fasting periods [89]. Figure 7 represents the principle of insulin- and glucagon-regulated glucose homeostasis.

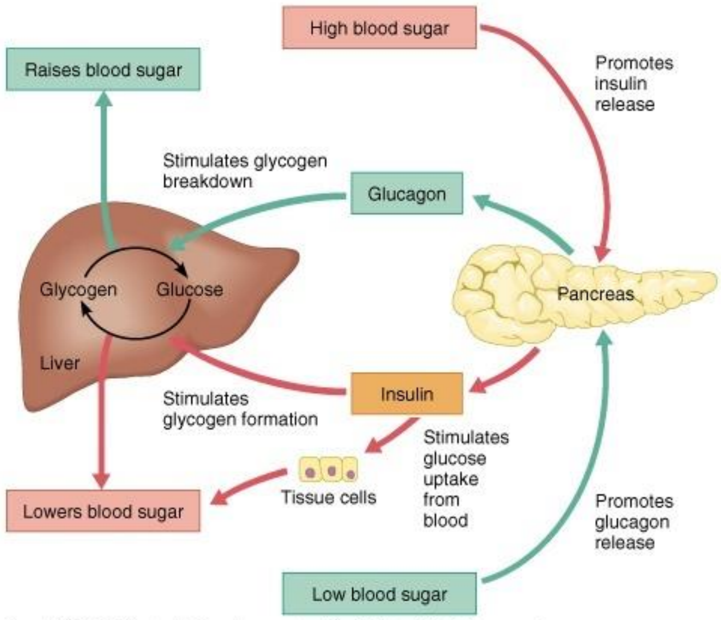


Figure 7: Principle of glucose homeostasis regulated by insulin and glucagon.
Red pathway: High blood glucose levels stimulate β-cells in the pancreas to release insulin. Insulin prompts tissue cells (adipocytes, muscle cells) to take up glucose from blood and stimulates glycogen conversion from glucose stores in the liver. This eventually lowers the blood glucose levels. Green pathway: Low blood glucose levels causing the α-cells within the pancreas to secrete glucagon, which stimulates glycogen degradation within the liver to produce and liberate glucose into the blood stream. This in consequence raises the blood glucose levels. Figure taken from https://physiology-11.wikispaces.com.

The additional hormones released by the δ-cells, γ-cells and ε-cells of the endocrine pancreatic cells are modulators of this complex process of the insulin and glucagon-controlled glucose homeostasis and their role will be discussed more detailed below.

1.5.2 Insulin and Insulin signalling

Insulin is exclusively secreted by the β-cells in pancreatic islets. Besides its crucial role in glucose homeostasis, Insulin also exerts anabolic functions within the body and is involved in carbohydrate metabolism, lipid metabolism and protein synthesis with its ability to stimulate glucose uptake by the liver, adipose tissue and skeletal muscle cells from the blood [90]. In these tissues, insulin is a key mediator for different metabolic processes such as, enhancing glycolysis, activating glycogen synthesis, stimulating lipid synthesis (adipose tissue only) and decreasing lipolysis. The latter resulting in the reduction of plasma fatty acid levels. Within the protein metabolism, insulin promotes protein synthesis and prevents protein degradation in muscle tissues [90-94].

Figure 8 shows a simplified scheme of insulin signalling in the human body. The gate for glucose uptake by insulin-sensitive cells/tissues is the glucose transporter type 4 (GLUT4), a transmembrane protein. By binding of insulin to its receptor (IR) on the cell surface of the targeted tissues, it activates a complex signal cascade needed for glucose uptake into the cytoplasm. On the extracellular side insulin binds to the α-subunit of the receptor with consequent activation of the β-subunits which exert tyrosine kinase activity. It autophosphorylates itself and subsequently other intracellular proteins, the so-called insulin receptor substrates (IRS) [95]. This is where the insulin signal begins to get transferred through a kinase-dependent signal cascade (PI3K/AKT pathway), to modulate the downstream anabolic processes of insulin stimulation (glycogen synthesis, protein synthesis, lipid synthesis), in particular the glucose uptake from blood. In order to induce glucose uptake, AKT2 initiates the migration of GLUT4-containing vesicles to the cell membrane by phosphorylating SNARE proteins [96]. Details around insulin secretion are outlined in section 1.6.

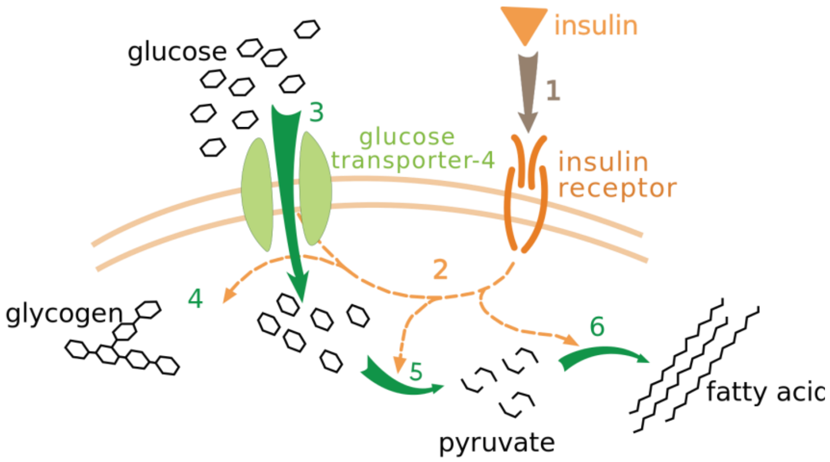


Figure 8: Overview of insulin signalling.
Secreted insulin binds to the extracellular domain of the insulin receptor on the surface of insulin-sensitive cells. Insulin binding initiates a kinase-dependent signal cascade mediating the translocation of GLUT4 for glucose uptake and initiates the downstream anabolic processes of insulin signal transduction (Glycogen synthesis, lipid synthesis, glycolysis). Figure taken from https://courses.lumenlearning.com.

1.5.3 Glucagon, glucagon secretion and signalling

Glucagon exerts its major role in the liver, driving glycogenolysis and gluconeogenesis and inhibiting glycogenesis and glycolysis and to prevent hypoglycaemia. It is the opponent of insulin and is secreted during hypoglycemic conditions to return blood glucose to normal levels. Apart from glycogen break down, glucagon also mediates the recruitment of glycerol from adipocytes and amino acids for gluconeogenesis in the liver [97,98]. It also increases fatty acid break down during β-oxidation (liver) and the formation of ketone bodies when carbohydrates are not available for energy production [99]. Glucagon also has important functions in the heart and the kidneys. It positively influences the cardiovascular performance and it increases the renal plasma flow (RPF), glomerular filtration rate (GFR) and electrolyte excretion in the kidneys as a consequence of its stimulatory effect on vasodilation (widening of blood vessels) [100,101].

Glucagon secretion works very similar to the mechanism of insulin secretion. The α-cells sense low blood glucose levels and become electrically active, encouraging them to fire action potentials (APs). Paracrine signalling between α-and β-cells is thought to be one of the factors enabling α-cells to sense low insulin levels after lowering blood glucose. However, the overall process is related to several factors (e.g. the central nervous system) and yet not well understood [102-104].

Action potential firing is promoted by the opening of voltage-gated sodium channels (Na_v channels) and subsequently also calcium channels (L-, N-, T-type). This prompts additional specific voltage-dependent calcium channels (P/Q-type) to open. The increasing intracellular calcium concentration in turn stimulates glucagon granule exocytosis allowing the release of glucagon. As blood glucose levels rise, and glucose enters the α-cell through the sodium-glucose co-transporter 2 (SGLT2) the intracellular glucose metabolism consequently raises ATP production, hence inhibiting membrane-embedded K_{ATP} channels (same channel as in β-cells). The intracellular potassium accumulation results in membrane depolarisation and this likely causes the inhibition of Na_v and P/Q-type calcium channels due to the, for them, insufficient voltage change, potentially lowering the amplitude of APs, thus the inhibition of glucagon release. Differently to β-cells, it is the P/Q-type calcium channel which predominantly triggers glucagon release in α-cells. The increase in circulating glucose levels then prompts the suppression of glucagon release. However, this process to date is not completely understood [105].

Glucagon signalling pathways are induced by binding of glucagon to its receptor to the surface of various cells in different tissues, e.g. liver, pancreas, heart, kidney, brain and smooth muscle. In the following, only the glucagon signal transduction in the liver shall be outlined in brief: The glucagon receptor is G-protein coupled and once glucagon is bound, the signal gets mediated via G-proteins ($G_s\alpha$, G_q) to activate the adenylate cyclase (AC, $G_s\alpha$). AC then boosts cAMP (cyclic adenosine mono phosphate) and protein kinase A (PKA, G_q) production, necessary to regulate the activity of the key enzymes within the hepatic glucose metabolism to eventually promote glycogenolysis and gluconeogenesis [106]. Glucose produced from glycogen stores enters the blood stream via the glucose transporter 2 (GLUT2) [107].

1.5.4 Somatostatin

Somatostatin is known for its extensive inhibitory function, operating in various organs and tissues and therefore is considered a key player in homeostatic processes. It acts as a neurotransmitter within the central nervous system (CNS) and also as a neurohormone inhibiting the release of growth hormone. Within the gastrointestinal tract and the exocrine and endocrine pancreas, it regulates processes such as, glandular secretion and smooth muscle contractility (e.g., contractility of stomach and intestine) in an autocrine, paracrine and neuronal manner and is either secreted by endocrine cells or neuroendocrine neurons. [108-111]. Somatostatin-producing and -secreting cells are so called δ-cells (D-cells), which are predominantly present in the central and peripheral nervous system, the gut and the endocrine pancreas. Within the pancreatic islets cluster, δ-cells are the third most prominent cell type exerting an inhibitory effect on insulin and glucagon secretion with their ability to release somatostatin [112].

Although not yet completely understood, somatostatin-induced glucagon inhibition seems to be glucose-dependent as experiments with mouse pancreatic islets have indirectly shown: Whilst somatostatin-deficient mouse islets did not indicate any decrease in glucagon secretion under high glucose condition, this was the opposite in control wt islets suggesting an inhibitory effect of glucose on the endocrine α-cells mediated by somatostatin [113]. In terms of β-cells, somatostatin secretion creates a negative feedback loop on insulin secretion, mediated by urocortin3 (Ucn3). Interestingly, this peptide hormone is co-secreted together with insulin by the pancreatic β-cells. Once released, it binds to the Crhr2α receptor (Corticotropin-releasing hormone receptor 2 alpha) on the surface of δ-cells promoting the release of somatostatin which then can exert its inhibitory function on the β-cells [114].

The mechanism, how somatostatin finally applies its inhibitory function on α-cell and β-cells follows the same principle: After somatostatin has bound to its receptor (SSTR2) on the human α- and β-cell surface it inhibits the action potential firing and prevents Ca^{2+}-

influx through VDCCs. These effects together eventually prevent the Ca^{2+}-dependent hormone exocytosis in α- and β-cells [115].

In addition, it was recently described, that ghrelin, the fifth hormone secreted by the endocrine pancreas, boosts somatostatin release from δ-cells, which leads to the inhibition of insulin release by binding of somatostatin to its receptor on the β-cells surface [116].

Figure 9 depicts a model of Watts et al., 2016 how α-, β- and δ-cells interact within the process of glucose homeostasis [117].

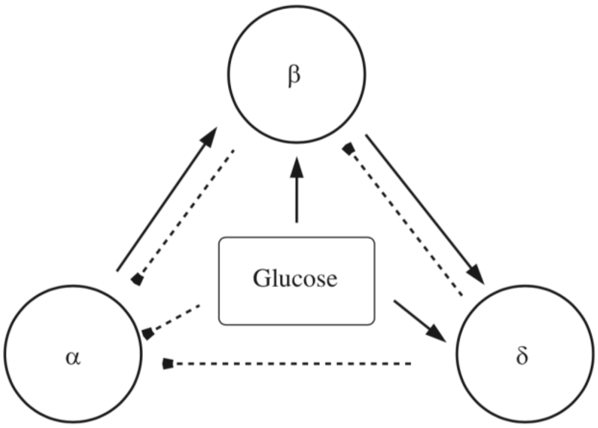


Figure 9: Overview of paracrine regulation between α-, β- and δ-cells in glucose homeostasis.
Black arrows indicate stimulation. Dashed lines indicate inhibition. High glucose levels stimulate β- and δ-cells, responding with insulin and somatostatin release, respectively but inhibits glucagon release from α-cells. Stimulated β- and δ-cells also suppress stimulation of α-cells and glucagon release. In contrast, glucagon-secretion from α-cells stimulates β-cells and insulin secretion. Figure taken from Watts et al. 2016 [127].

1.5.5 Ghrelin

Compared to the other hormones well known to be of pancreatic origin since decades, the presence of ghrelin as being one of them has been discovered rather recently. Ghrelin is an orexigenic (appetite stimulant) peptide secreted by the endocrine cells within the gastric mucosa [118] Shortly after its discovery, scientists provided the evidence for its presence and endocrine function within the pancreas whilst performing double- or triple-immunostainings with the other four islets hormones [119]. Earlier it was already reported that ghrelin levels were high during fasting periods, when the demand on insulin was low and conversely a decreasing concentration of ghrelin was noted when the demand on insulin was high. This helped to come to and verify the conclusion, that it acts as an inhibitor of glucose stimulated insulin secretion from β-cells in humans within the pancreatic islets [120]. As mentioned above (1.5.4), scientist could show, that ghrelin affects insulin secretion indirectly by promoting somatostatin secretion and this in turn has an inhibiting effect on insulin release. This provides a good example for the complex crosstalk between the different cells within the pancreatic islets [116].

1.5.6 Pancreatic Polypeptide

The PP or γ-cells respond to food intake with the release of the pancreatic polypeptide. Its role is not completely understood yet, but it is known to be involved in several regulatory processes, such as appetite regulation in both directions, anorixigenic and orexigenic [121,122] and the inhibition of pancreatic exocrine secretion [123].

It acts as a mediator between the gastrointestinal tract and the CNS, which are connected via the vagus nervus to regulate the energy homeostasis [124] and is part of the neurohormonal system, balancing energy accumulation (food intake) and energy expenditure [125].

The predominant stimulus of pancreatic polypeptide secretion comes from the parasympathetic vagus nervus (autonomic parasympathetic nervous system) which, opposed to the sympathetic nervous system (“fight-or-flight” response), is responsible for body functions when at rest, e. g. the stimulation of digestion (energy conservation) and the promotion of satiety [126]. The secretion is biphasic with a cephalic and gastrointestinal phase (peripheral), the first promoting the orexigenic effect on appetite whereas the peripheral phase does the opposite. On the cephalic primary level, external stimuli, such as taste or smell are recognised by the CNS which responds through vagal efferent neurons with the stimulation of pancreatic polypeptide release. On the gastrointestinal level (secondary phase), various stimuli induce the secretion of the pancreatic polypeptide, such as 1) protein and lipid consumption, 2) gastric distention due to food intake and following activation of gut epithelial mechanoreceptors, 3) bioactive peptides (e.g. cholecystokinin, CKK) modulating postganglionic efferent neurons and also 4) hypoglycemia [127]. However, the autonomic neuronal influence on human islets is somewhat unclear. Differently to mouse pancreatic islets, human pancreatic islets are sparsely innervated, and this rather suggests an indirect influence on pancreatic hormone secretion. This could be for example through changing the local blood flow or neurotransmitters being spilled over into the islets microcirculation [128].

1.6 The insulin secretion pathway

1.6.1 Glucose induced insulin secretion pathway

The glucose-induced insulin secretion pathway is dependent on changing membrane potentials. β –cells are sensing changes in plasma glucose and this initiates the insulin secretion pathway (Figure 10). Blood glucose enters the cell via the human glucose transporter GLUT2 and gets rapidly metabolised during glycolysis within the cytosol. Pyruvate, the product of the glucose breakdown, is then further metabolised in the citric acid cycle also known as Krebs cycle and provides the driving force for ATP synthesis. The increasing cytosolic ATP concentration causes the closure of ATP-sensitive potassium channels (K_{ATP}) located in the β-cell membrane. These K_{ATP} channels are membrane protein complexes composed of a K_{ir} 6.2 pore formed of four subunits and a regulatory subunit made of four SUR1 (Sulfunylurea receptor 1) subunits [129-132]. Kir 6.2 is encoded by the *KCNJ11* gene and SUR1 by the *ABCC8* gene, both genes are associated with congenital hyperinsulinism [133,134]. At a state of low blood glucose levels, the K_{ATP} channels remain open allowing potassium ions to leak out of the cell constantly and this is keeping the β-cells hyperpolarised and electrically inactive, with a membrane potential of about -65 mV (cell inside negative). The increasing glucose levels and the following rise of intracellular ATP concentration affect the activity of the K_{ATP} channels due to enhanced ATP-binding and this in consequence is leading to channel closure. This results in the depolarisation (a temporary accumulation of positive charge inside the cell) of the β-cell membrane and, in consequence, enables voltage-dependent calcium (Ca^{2+}) channels (VDCC) to open upon a certain threshold potential and Ca^{2+-}influx to happen. The increasing Ca^{2+}-concentration then finally triggers the release of insulin stored in insulin granules via exocytosis (Ca^{2+}-triggered exocytosis) [135,136]. Figure 10 depicts the glucose-induced insulin secretion pathway [137].

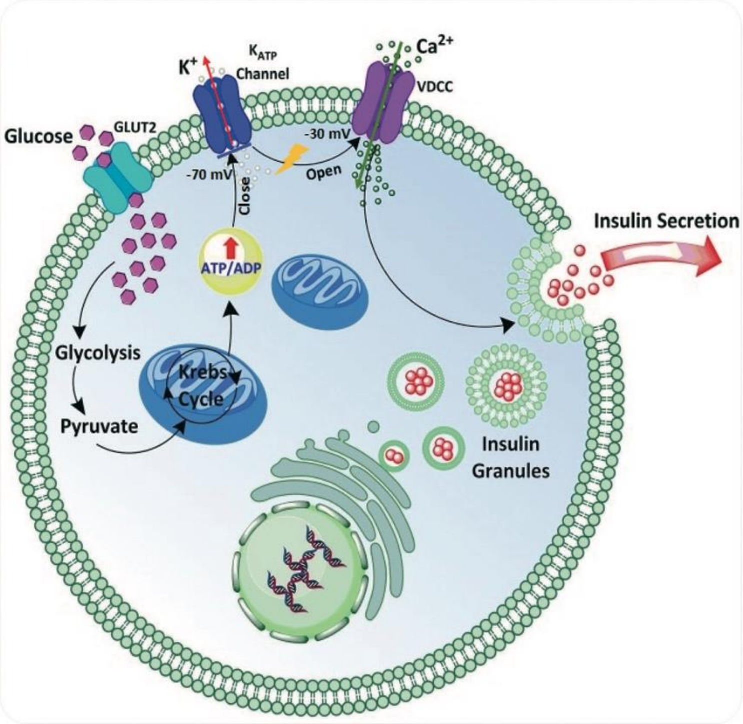


Figure 10: Glucose-induced insulin secretion pathway in human pancreatic β-cells.
Glucose molecules bind and enter the cytoplasm via the membrane embedded glucose transporter GLUT2. Glucose gets metabolised in glycolysis and Krebs cycle and drives ATP production. The raising intracellular ATP concentration promotes K_{ATP} channels to close preventing potassium efflux. This depolarises the cell membrane and action potential firing causing VDCC to open. The increasing calcium concentration triggers insulin granule fusion with the cell membrane and insulin release. Figure taken from Castiello 2016.

1.6.1.1 Biphasic character of insulin release

Insulin secretion in response to glucose is a biphasic process which can last for hours. Biphasic insulin release is glucose-dependent and happens when the extracellular glucose concentration increases dramatically from basal levels (3 mM) below the physiological range (4-6 mM) to suprastimulatory concentration of 16.7-22 mM (Grodsky 1972). During the first phase, insulin gets released rapidly and in high concentration within the first 5-10 minutes after the stimulus. It is thought that the amount of insulin released during the first phase originates form a so-called ‘readily releasable pool’ (RRP) stored in insulin granules which are already located close to the plasma membrane. In contrast, insulin released during the second phase is stored in granules which are located deeper within the cell and first need to be recruited to the cell membrane for fusion and exocytosis [138]. These granules are called the ‘Newcomer granules’. As the process of Insulin secretion lasts for hours, the majority of insulin released derives from the second phase. Only about 15% of the total insulin released, results from first phase secretion during the first hour of stimulation. When this first phase is over the insulin release rate drops but is being maintained over hours by the second phase with a lower release rate until normal glucose levels are restored [139]. Figure 11 depicts a biphasic insulin release profile [140].

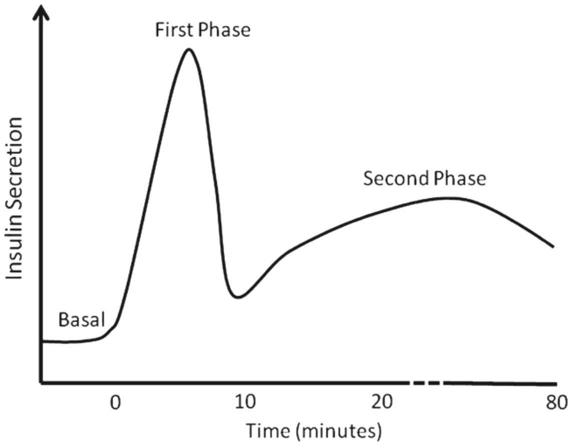


Figure 11: Glucose-dependent biphasic insulin release profile.
Basal insulin levels are low. With a strong glucose stimulus at time point zero (0) a rapid and strong insulin release happens during the first 10 minutes of stimulation (First phase). Insulin secretion rate then drop but stays above the basal level for an extended amount of time. Figure taken from Meloni 2013.

1.6.1.2 Exocytosis of insulin granules

The insulin granule exocytosis, particularly during the second phase release, is a complex process of intracellular vesicle trafficking and fusion of insulin granules to and with the cell membrane [141]. It is a SNARE-protein (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) mediated process [142]. More than 30 types of SNARE proteins are found in mammals representing a transmembrane protein family in particular located in cell organelles and vesicles, which form SNARE complexes to allow fusion of vesicles for exocytosis or intercellular trafficking [143]. SNARE proteins are located in the vesicle (v) membrane and the target (t) membrane meant to fuse with each other and are therefore subdivided into (v)-SNAREs and (t)-SNAREs [144]. The transport of the insulin granules to the cell periphery is also depended on the cytoskeleton, with involvement of the F-actin filament and microtubule network. Long-range movement of the insulin granules from the core of the β-cell to the periphery happens along microtubules and is kinesin-mediated. For the involvement of the F-actin filament, two models do exist:

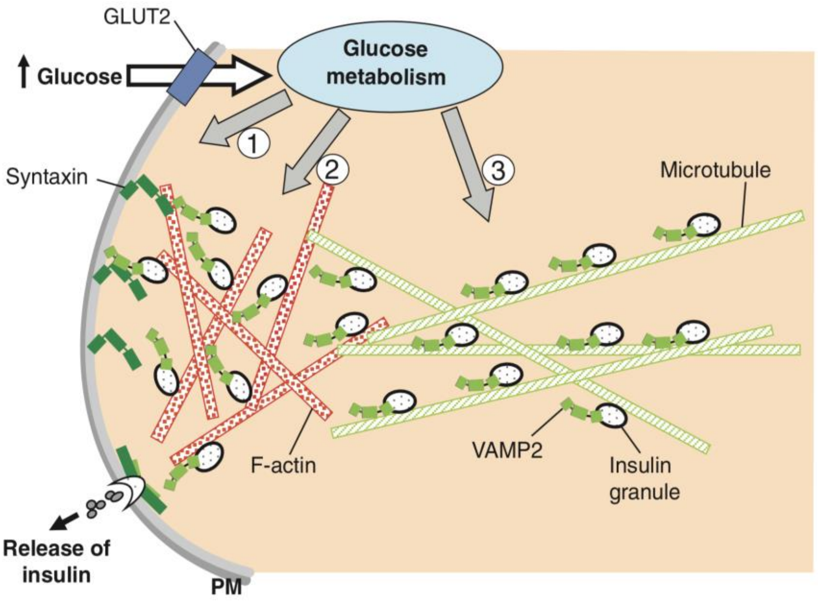


Figure 12: Model of insulin granule recruitment during the second phase of insulin secretion supposed by Wang et al. [154].
Glucose enters the cell via GLUT2 when blood glucose levels are high. The changing intracellular glucose metabolism drives the rearrangement of the sublaminar F-actin filament (red) to unblock the way for SNARE protein complex formation (Syntaxin, VAMP2) between vesicles and cell membrane and allows migration of insulin granules first from the cell core along the microtubule network (green) to the periphery and then via the f-actin tracks further to the plasma membrane for fusion and insulin release. Figure taken from Wang et al., 2009.

The first and older model, the ‘barrier model’ suggests, that the F-actin filament undergoes a reorganisation in β-cells dependent on the glucose metabolism. It is also thought, that a connection between microtubules and the F-actin filament is built at the cell margins forming a bridge for the insulin granules which then “walk” along the actin tracks to the cell membrane. In this scenario it is supposed, that under basal glucose conditions, the densely packed F-actin fibres block the way for the insulin granules to reach the plasma membrane and this in consequence prevents the formation of the SNARE protein complexes, crucial for vesicle-membrane fusion. In contrast, raising cytosolic glucose concentrations, when blood sugar levels are elevated, induce the reorganisation of the F-actin network and this allows insulin granules to access the plasma membrane, SNARE protein complexes to form and eventually enabling granule fusion resulting in the release of insulin as demonstrated in Figure 12 [145]. Another, more recent, model agrees with the barrier function of the actin filament during fasting (low glucose) periods to prevent suboptimal insulin release, but claims, that insulin granules are transported to the cell membrane directly across microtubules only, independently from the F-actin network being reorganised during periods of glucose stimulation [146].

1.6.2 GLP-1 induced insulin secretion pathway

The glucagon-like peptide 1 (GLP-1) belongs to the family of incretins (blood glucose decreasing metabolic hormones) and has the ability to enhance insulin secretion. It is produced and secreted by the L-cells located in the intestine, mainly the ileum and colon, upon nutrient intake [147,148]. The insulinotropic effect of GLP-1 is glucose dependent. L cells are enteroendocrine cells, which sense preferentially carbohydrates and fat rather than proteins and respond with GLP-1 release to lower raised blood glucose levels [149]. It is encoded by the same gene as glucagon and both share the same precursor peptide together with further proteins which either have their role in the intestine or the pancreas. The maturation of proglucagon to GLP-1 is depended on tissue-specific posttranslational processing [150]. GLP-1 has also glucagonostatic properties at normal physical conditions. However, the inhibiting mechanism on glucagon is not completely clarified but is likely an indirect effect due to GLP-1-mediated stimulation of somatostatin [151]. Apart from its effect on insulin and glucagon secretion, GLP-1 exerts many other functions on different organs. For example, it has an influence on gastrointestinal motility as well as a presumed protective effect on the heart and the neural system [152,153].

Depending on the concentration of GLP-1 exposed to β-cells or pancreatic islets the downstream GLP-1-induced signalling pathway to stimulate insulins secretion differs. In experimental settings with isolated β-cells or pancreatic islets the effect of both, nanomolar (nM) and picomolar (pM) levels of GLP-1 on insulin release have been investigated and were summarised in Shigeto et al. 2015 and 2017 [154,155]: The cAMP/PKA-dependent and cAMP/Epac-dependent pathways are induced by nanomolar concentrations of GLP-1 and are well established and commonly accepted (Figure 14). Rather recently, experiments with GLP-1 concentrations within the physiological picomolar range (1-10 pM) have been performed with isolated pancreatic islets or β-cells. The results provide the evidence, that picomolar concentrations of GLP-1 in an experimental set up have the potential to depolarise the plasma membrane, increase calcium influx and raise insulin secretion in an cAMP/PKA-independent manner [156]. This pathway is now considered being the major pathway for the stimulation of insulin release at physiological GLP-1 concentrations (Figure 13). At this point it should be noted, that the physiological GLP-1 levels do not exceed 30 pM even directly after food intake.

GLP-1 stimulation of β-cells can experimentally be mimicked by PMA (Phorbol 12-myristate 13-acetate), a phorbol ester which activates PKC [157]. It was used in this study.

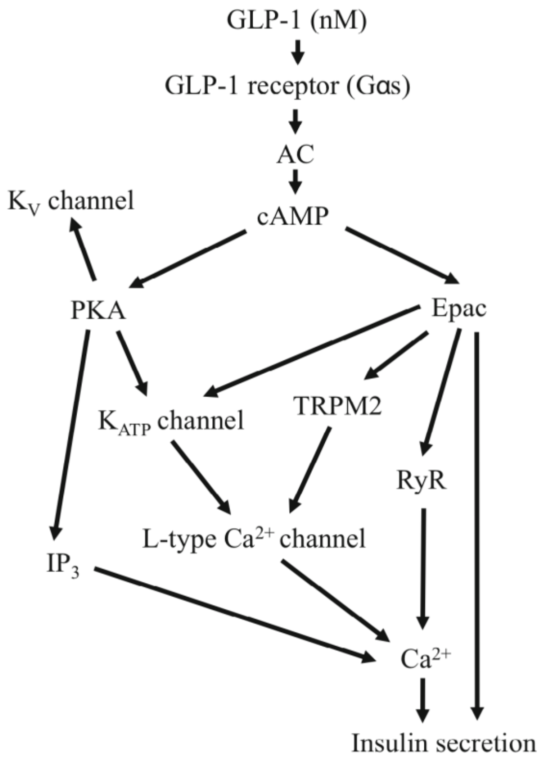


Figure 14: Model for nanomolar (nM) concentration-mediated GLP-1-induced insulin secretion pathway.
cAMP/PKA-dependent and cAMP/Epac-dependent pathway. **PKA-dependent route:** GLP-1 binds to its receptor on the β-cell surface and activates the AC. AC uses ATP as a substrate to transform it into cAMP (3'5'-cyclic AMP) and pyrophosphate. cAMP activates PKA. PKA initiates K_{ATP} channel closure (depolarisation and L-type Ca^{2+} channel opening), phosphorylates IP_3R-formed intracellular Ca^{2+} channels. IP_3 and Ca^{2+} together prompt the IP_3R-formed Ca^{2+} channel to open (first due to IP_3 binding which causes conformation change and allows Ca^{2+} to bind and this supports the opening of the central pore) and Ca^{2+} gets released from the store. Increasing Ca^{2+} concentration prompts insulin release. After depolarisation PKA also activates voltage-dependent potassium channels (K_V) to open for repolarisation of the membrane. **PKA-independent route:** High cAMP levels also recruit Epac to promote insulin secretion. Epac interacts with K_{ATP} channels, TRPM2 cation channels and RyR-Ca^{2+}-channels to initiate depolarisation (through cation influx) and intracellular Ca^{2+}-accumulation for following insulin release. Figure taken from Shigeto 2017.

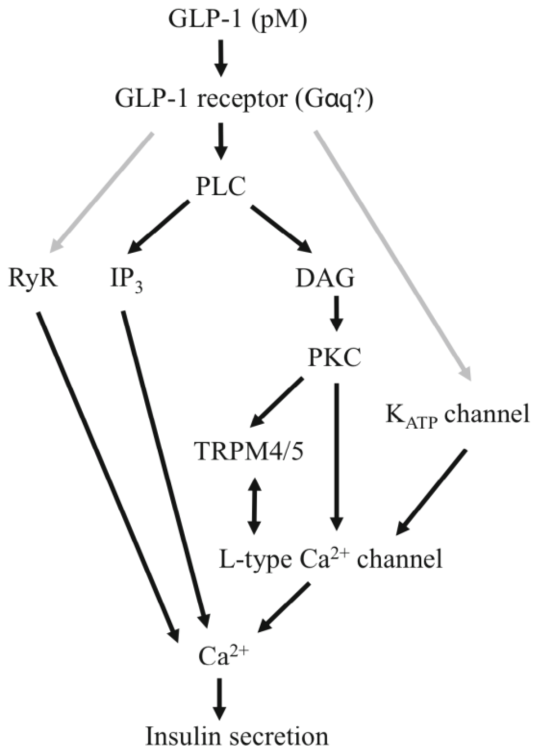


Figure 13: Model for picomolar (pM) concentration-mediated GLP-1-induced insulin secretion pathway.
PLC/PKC-dependent pathway. GLP-1 binds to the GLP-1 receptor on the β-cell surface and induces PLC activity. This results in IP_3 and DAG production. The raising IP_3 concentration induces Ca^{2+} mobilisation from intracellular calcium stores. The increasing DAG production activates PKC which activates L-type Ca^{2+} channels promoting Ca^{2+} influx. The raising Ca^{2+} concentration also stimulates Ca^{2+}-dependent TRPM4/5 channels for additional cation influx (Na^+). GLP-1 also promotes K_{ATP} channel closure resulting in increased intracellular K^+ concentration. The raising intracellular cation concentration (Ca^{2+}, K^+, Na^+) depolarises the cell membrane which results in calcium-dependent insulin release. GLP-1 also mediates RyR-based release of Ca^{2+} stores. Figure taken from Shigeto 2017.

1.7 Hyperinsulinemic hypoglycaemia (Hyperinsulinism)

Hyperinsulinism (HI) describes the excessive secretion of insulin by the β-cells of the pancreas. HI can originate due to either decreasing insulin sensitivity of glucose removing cells (liver and muscle cells) stimulating β-cells to secrete more insulin or the hypoglycaemia associated form (Harris 1924) often due to gene mutations affecting the insulin secretion pathway. The consequence of HI-related hypoglycaemia bears a high risk of a permanent brain damage due to an undersupply with glucose resulting in hypoglycaemic brain injury. A variety of clinical symptoms are seen in patients with HI with most severe presentation in newborns, having features such as apnea, seizures and unresponsiveness. Less severe presentation comes with rather non-specific symptoms for example, poor feeding, irritability and lethargy. HI can also be associated with cardiomyopathy, hepatomegaly and developmental syndromes [158,159,160].

1.7.1 Genetic cause, encoded proteins and molecular pathomechanism

The genetic cause for HI is quite heterogeneous with at least eight different mutations in key genes regulating insulin secretion being reported affecting adequate insulin release from β-cells: *ABCC8, KCNJ11, GLUD1, GCK, HADH, HNF4A, SLC16A1 and INSR*. The encoded proteins are mainly β-cell-associated or associated with insulin-responsive cells. The latter being the case for mutations in the insulin receptor gene, *INSR* (MIM 609968, Chr19p13.2). HI occurs with an incident of 1 in 50,000 live births in the general population and 1 in 2,500 in particular communities with a high rate of consanguinity [161]. Inheritance of HI mainly is recessive but also some dominate mutations have been reported. In total, around 55% of cases presenting HI result from mutations in the above listed genes, the remainder is of unknown genetic cause. However, amongst them, mutations in *KCNJ11* and *ABCC8* are the most common cause of inherited HI with 40%-45% of all cases. Whereas mutations in the remaining genes account for only 5%-10%. Successful treatment of HI depends on the causative mutation and the function or dysfunction of the encoded protein. HI patients with mutations in *GLUD1*, *HADH* and *HNF4A* usually respond well to diazoxide treatment. In contrast, those with mutations in *ABCC8* or *KCNJ11* they often need surgery. Diazoxide acts on potassium channels, causing them to open [162].

The molecular pathophysiology within HI known so far is outlined below in more detail. In addition, Figure 15 depicts an overview of molecular disease mechanisms in pancreatic β-cells affecting insulin secretion in HI.

1.7.1.1 K_{ATP} channel mutations (*KCNJ11*, *ABCC8*)

KCNJ11 (MIM 161820, Chr11p15.1) and *ABCC8* (MIM 600509, Chr11p15.1) encode for the subunits of the ATP-dependent potassium channel (K_{ATP}) $K_{ir}6.2$ and SUR1, respectively. Located in the β-cell plasma membrane the K_{ATP} channel functions like a mediator between glucose metabolism and electrical membrane potential crucial for insulin secretion. HI is associated with loss-of-function mutations in these genes. Under healthy circumstances, high glucose levels increase the intracellular ATP concentration and ATP-binding initiates channel closure with subsequent membrane depolarisation resulting in calcium influx and eventually the coupled insulin release. $K_{ir}6.2$ opening and closure is regulated by intracellular ATP and ADP concentrations via two different nucleotide binding sites. ATP exerts an inhibitory function to the channel and stimulates channel closure once bound. ADP has a stimulatory effect on the channel prompting it to open in a magnesium-dependent manner, allowing potassium efflux. Hence, when intracellular magnesium (Mg) is absent, only the inhibitory effect is accomplished. Mutations in *KCNJ11* prevent MgADP-binding to $K_{ir}6.2$ causing permanent membrane depolarisation and excessive insulin release independently from blood glucose concentrations. Mutations in the *ABCC8* gene occur with a much higher frequency compared to those in *KCNJ11* [163,164]. Since SUR1 and $K_{ir}6.2$ are co-expressed for correct channel formation, mutations in The SUR1 gene (*ABCC8*) triggering abnormal protein maturation, defective assembly of the SUR1 β-subunits forming the channel or also wrong protein trafficking to the cell surface can result in the absence of $K_{ir}6.2$, as well leading to continuous membrane depolarisation and insulin secretion [165].

1.7.1.2 Glutamate dehydrogenase mutations (GLUD1)

The glutamate dehydrogenase (GDH) is encoded by the *GLUD1* gene (MIM 606762, Chr10q23.2) and mutations are causative for a gain of protein function in protein-sensitive HI. GDH is leucin-sensitive and involved in amino acid-stimulated insulin secretion in the mitochondrion and stimulates α-ketoglutarate formation from glutamate in a GTP/ADP ratio-dependent manner. Low glucose levels (high ADP) and leucine activate GDH (leucine-sensitive state) resulting in the increase of intracellular ATP from glutamate metabolism, thus prompting K_{ATP} channel closure. High GTP levels reverse GDH-sensitivity to leucine (leucine-insensitive state) preventing an effect on K_{ATP} channels and related insulin release. *GLUD1* mutations can change the sensitivity of GDH to high GTP levels making it unresponsive, preventing its inhibition and this leads to excessive secretion of insulin [162,166].

1.7.1.3 Glucokinase mutations (*GCK*)

The glucokinase is encoded by the *GCK* gene (MIM 602485, Chr7p13) and is one of the key enzymes in glucose metabolism. It catalyses the phosphorylation of glucose to glucose-6-phosphate after glucose uptake through the membrane transporter GLUT2 in β-cells and contributes to K_{ATP} channel activating ATP levels. Mutations in GCK are enhancing the enzymes' affinity to glucose and consequently the glycolysis rate, resulting in exorbitant insulin release independently from its demand [167].

1.7.1.4 Solute carrier family 16, member 1 mutations (*SLC16A1*)

Mutations in the *SLC16A1* gene (MIM 610021, Chr1p13.2) are causative for exercise-induced HI due to the accumulation of lactate and pyruvate and the consequent increase of ATP levels – eventually resulting in excessive insulin release during elaborate exercise. *SLC16A1* encodes for the monocarboxylate transporter 1 (MCT-1), which sits in the mitochondrial membrane promoting lactate and pyruvate transport into the lumen. Although, its expression in β-cells is limited to reduced levels under normal conditions, mutations in *SLC16A1* promote gain-of-function and cause increased expression on the β-cell surface, enhancing its response to extracellular lactate and pyruvate levels [168].

1.7.1.5 *HNF4A* and *HADH* mutations

The exact molecular pathophysiology of HI mutations causing excessive insulin release due to mutations in the *HNF4A* (MIM 600281, Chr20q13.12) and *HADH* (MIM 609975, Chr4q25) genes are unknown to date. The transcription factor *HNF4A* Hepatocyte nuclear factor 4A) has multiple regulatory functions within the pancreatic islets on the transcriptional level, causing loss-of -function mutations [169]. *HADH* encodes the hydroxyacyl-coenzyme A dehydrogenase. *HADH* is a key enzyme in mitochondrial β-oxidation catalysing the penultimate reaction before Acetyl-CoA is generated to enter the Krebs cycle. However, the exact dysregulation responsible for elevated insulin secretion is yet not understood [170].

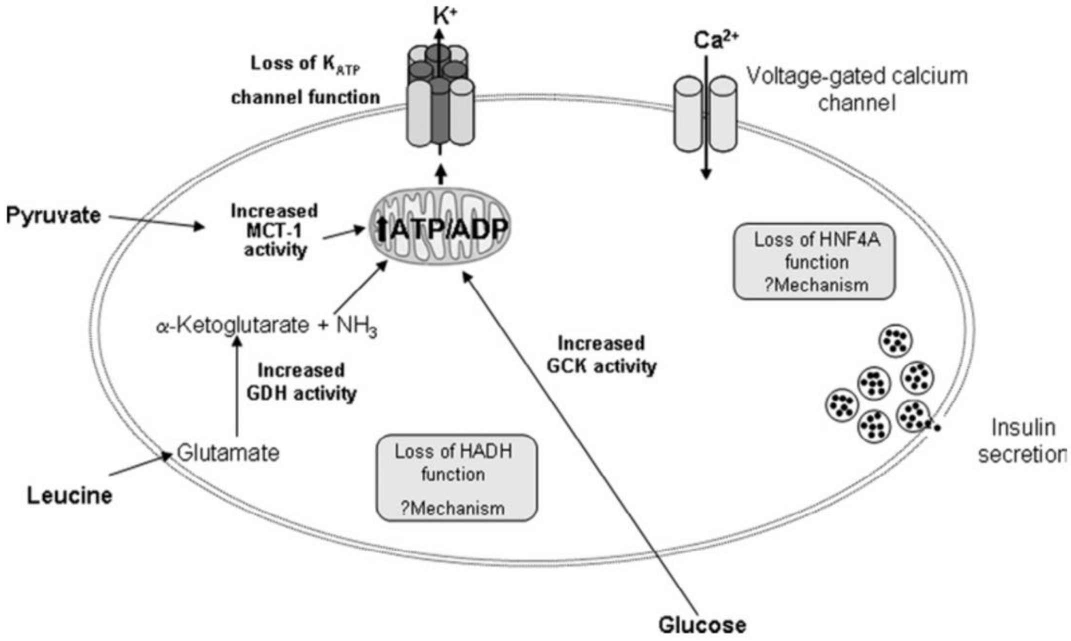


Figure 15: Molecular mechanisms in pancreatic β-cells underlying hyperinsulinemic hypoglycaemia.
Mutations in the *KCNJ11* and *SUR1* genes, encoding the K_{ATP} channel cause loss-function mutations. Mutations in *GCK*, *SLC16A1* and *GLUD1* cause gain-of-functions mutations in the encoded proteins glucokinase, MCT-1 and glutamate dehydrogenase, respectively. The molecular mechanisms of mutations in HNF4A and HADH (Hydroxyacyl-coenzyme A dehydrogenase) causing excess secretion of insulin are unknown. Figure taken from Flanagan 2011.

1.8 Congenital disorders of glycosylation

Congenital disorders of glycosylation (CDG) are inherited metabolic disorders based on mutations affecting the biosynthesis (Type I) or processing (Type II) of glycoconjugates (glycoproteins, glycolipids). The transfer of polysaccharides (glycans) onto proteins or lipids are either N-linked, via the amide group of asparagine or O-linked, via the hydroxyl group of a serine or threonine at the N-terminus. N-linked glycosylation (please refer to section 1.8.1.3 for details) is a much more complex process happening in the cytoplasm, the endoplasmic reticulum (ER) and the Golgi, compared to O-linked glycosylation, which starts at the Golgi.

Several hundred enzymes and transporters are involved in posttranslational protein glycosylation pathways, which are potential targets for a glycosylation defect and CDG. To date, only 28 CDGs caused by mutations in different genes are known, with in total 16 N-linked CDGs, 6 O-linked CDGs, 4 CDGs with both N- and O-linked defects and 2 with defects in lipid glycosylation. CDGs are highly heterogeneous ranging from mild to severe diseases with multisystem features to types with abnormalities only in a single organ. Major features within CDGs are of neurological nature due to significant brain damage, such as psychomotor retardation, ataxia, eye abnormalities, structural brain abnormalities and hyporeflexia to name a few [171,172].

The most powerful screening test for a N-linked glycosylation defect and the diagnosis of CDG is the assessment of serum transferrin levels in patients analysed in an isoelectric focussing assay. An example is represented in Figure 16.

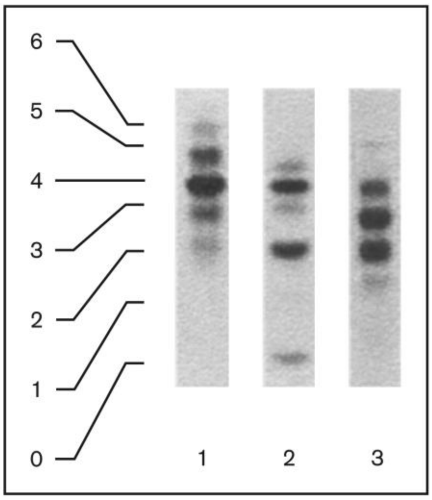


Figure 16: IEF of serum transferrin of CDG-1 and CDG-2. (1) wt control. (2) CDG1 and (3) CDG2. Figure taken from Jaeken and Carchon 2001.

Deficiencies in the phosphomannomutase 2 (*PMM2*) characterising the CDG subtype CDG-1a is the most common variant within CDGs and is described below in more detail.

1.8.1 Congenital disorder of glycosylation Type 1 A

The congenital disorder of glycosylation Type 1 A (CDG-1a) is by far the most frequent form of CDG, with more than 700 affected individuals known worldwide in 2010 [173].

It was first reported in 1980, as a well-described rare autosomal recessive disease due to coding mutations within the *PMM2* gene [174]. Commonly also named PMM2-CDG or Phosphomannomutase 2 deficiency, it is characterised by the defective post-translational glycosylation of N-linked Glycoproteins due to the defective synthesis of mannose-1-phosphate [175,176].

Most reported cases are due to compound heterozygous *PMM2* mutations, only a few present a homozygous genotype (N216I) [177], Y106F [178]. Homozygosity for complete loss-of-function mutations is lethal suggesting that a minimum of residual enzymatic PMM2 activity might be crucial for survival [179].

1.8.1.1 The phenotype within CDG-1a

Initially, after being identified CDG-1a was characterised as a combination of developmental delay, distinct fat pads, cerebellar hypoplasia and inverted nipples. A few years later, the application of the transferrin screening (isoelectric focussing analysis) helped to develop a broader view on the consequences of functional *PMM2* mutations. In contrast to what was thought in the beginning, it became evident, that within CDG-1a any organ system can be affected and it became obvious that CDG-1a presents a multisystem disorder with multiple features in a highly variable manner ranging from mild to very severe expression [180,181] . At the neonatal stage, infants commonly present with muscular hypotonia, inverted nipples and abnormal subcutaneous fat pads, but also additional dysmorphia like large protruding ears or a high forehead or feeding problems (anorexia, vomiting and diarrhea) can be the case – however any of these symptoms can also be missing [181]. Overall, predominant features in CDG-1a are of neurological and dysmorphic nature. Kidney (e.g. tubular microcysts, proximal tubulopathy), liver features (e.g. Hepatomegaly, liver fibrosis) and hyperinsulinism have also been reported [175].

The disease progress was categorised in four stages by Hagberg et al.: 1) Multisystemic stage (Babies), 2) Ataxia-mental retardation stage (Childhood), 3), Leg atrophy (Teenagers) and 4) Hypogonadal stage (Adulthood) [182]. This pattern still applies to most cases, given the patients survive the early years of life as the mortality is quite high with 20% of those patients presenting the “visceral” form of CDG-1a, a combination of neurological and non-neurological features often lethal due to organ failure [183]. Table 3 provides a summary of the different features present in CGD-Ia.

Table 3: Clinical features noted in CDG-1a patients. Taken from Grunewald [181].

Neurology	Axial hypotonia; peripheral neuropathy; hyporeflexia; developmental delay; seizures; stroke-like events; micro- and macrocephaly; myopathy
Gastroenterology/hepatology	Failure to thrive; vomiting; protein-losing enteropathy; liver dysfunction; hepatomegaly; chronic diarrhea
Neonatology	Hydrops; ascites; pleural effusions; multiorgan failure; failure to thrive; floppy baby
Haematology	Thrombocytopenia; coagulopathy; thrombosis; anaemia; leukocytosis
Endocrinology	Hyperinsulinaemic hypoglycaemia; hypothyroidism; hypergonadotropic hypogonadism; growth retardation
Clinical genetics	Mild dysmorphic facial features; inverted nipples
Orthopaedics	Osteopenia; joint contractures; kyphosis/scoliosis; short limbs; arthrogryposis
Ophthalmology	Abnormal eye movements; squint; cataract; retinitis pigmentosa; nystagmus; (transient) cortical blindness
Radiology	Small cerebellum, olivopontocerebellar hypoplasia; delayed myelinisation
Histology	Liver fibrosis; liver cirrhosis; lamellar inclusions in hepatocytes; renal microcysts
Dermatology	Abnormal fat distribution
Nephrology	Nephrotic syndrome; tubulopathy; cystic kidneys, enlarged kidney
Immunology	Recurrent infections; hypogammaglobulinaemia
Cardiology	Cardiomyopathy; pericardial effusions, cardiac ischemia, cardiac tamponade
Biochemistry	Hypoalbuminaemia; elevated transaminases; low cholesterol, triglycerides; decreased antithrombin III; decreased factor VIII, IX and XI; decreased protein C and S; elevated FSH, LH and prolactin; low free T4, increased ferritinaemia

Although hyperinsulinism and renal cysts have been observed in CDG-1a, this is only the case in a subset of patients and associated with the typical additional features. Moreover, nephromegaly and glomerular cysts, the renal phenotype in HIPKD, has not been described in CDG-1a.

Aside from CDG and CDG-1a, several congenital disorders are described which are caused by defective N-glycosylation, e.g. LAD II, Galactosemia type I-III or the Ehlers-Danlos syndrome [184], reflecting its importance within biological processes.

1.8.1.2 Genetic cause – The *PMM2* gene

The *PMM2* gene is relatively small, with eight exons, and is located on chromosome 16p13.2 as identified by Matthijs et al. in 1997 [185,186]. Two years before, in 1995, Schaftingen et al. [187] were able to report, that CDG-1a (MIM#212065) underlies a phosphomannomutase deficiency. They concluded from results of transferrin analyses of CDG-1a patients an early defect during the N-glycosylation pathway as the glycosylation pattern on transferrin blots were normal but overall reduced in patient samples, indicating decreased mannosylation of targeted proteins [187].

Matthjis et al. have found eleven more different missense mutations in *PMM2* and defined the gene as “the” causative gene for the congenital disorder of glycosylation type 1a [188,189]. To date, more than 100 pathogenic variants of CDG-1a have been reported, including missense, nonsense and frameshift mutations, splicing defects and exon-skipping mutations within *PMM2,* with the c.G425A (Arg141His) transition being the most common, *PMM2* variant with 75% [190-192].

1.8.1.3 Function of PMM2 – N-linked glycosylation

The encoded enzyme, phosphomannomutase 2 (PMM2) is an isomerase and a key enzyme within posttranslational N-linked glycosylation, which catalyses the isomerisation of mannose 6-phosphate (Man-6-P) to mannose 1-phosphate (Man-1-P), a precursor to GDP-mannose [185].

Posttranslational modifications (PTM) are enzyme-catalysed additions of functional groups (e.g., glycosylation, phosphorylation, methylation, acetylation) onto proteins after protein biosynthesis. Many proteins undergo PTMs to fully develop their structural and functional maturation [193]. More than half of all naturally occurring proteins are glycosylated [194] including various types of proteins such as structural and functional (enzymes, receptors, hormones) proteins, protective proteins (Mucin) or proteins of the immune system (Immunoglobulins), to fulfil their function within the cell, on its surface or beyond [195].

The involvement of PMM2 in N-glycosylation is placed very early within this posttranslational pathway. The enzyme catalyses the isomerisation of mannose 6-phosphate (Man-6-P) to mannose 1-phosphate (Man-1-P), a precursor of GDP-mannose (nucleotide sugar) necessary for the synthesis of the lipid-linked glycan Dolichol-P-Mannose (Dol-P-Man), the carrier and donor of oligosaccharides for transfer onto nascent proteins in this pathway [209]. The assembly of N-glycans starts on the cytosolic side of the rough ER (RER) membrane with the formation of Dol-P-Man: The dolichol-P linker (Dol-P) sits in the RER membrane and functions as an acceptor for mannose sugars attached to GDP-mannose (mannose donor substrate), which before were transferred from Man-1-P – after conversion through PMM2 action – onto GDP. Prior to receiving mannoses two molecules N-acetylglucosamine (GlcNAc) are attached to Dol-P through a pyrophosphate linkage (GlcNAc$_2$-PP-Dol) – the first assembly step. The GlcNAc$_2$-PP-Dol precursor glycan then is ready for elongation starting with the reception of one mannose sugar obtained from GDP-mannose and is further build up to a biantennary structure with, in total, five mannose sugars (Man$_5$-GlcNAc$_2$-PP-Dol) building the core glycan. Five mannosyltransferases (I-V) are catalysing these reactions. At this point, the core glycan gets “flipped” across the RER membrane facing now the luminal side for further modifications and elongations with glucose and mannose sugars to form the final glycan (Glc$_3$Man$_9$GlcNAc$_2$-PP-Dol). In the case of CDG-1a when PMM2 performance is dysfunctional, the formation of the core glycan with its tree build of five mannose sugars remains absent due to disturbed Man-6-P to Man-1-P conversion [196]. The early stage and PMM2 action of posttranslational N-glycosylation is depicted in Figure 17.

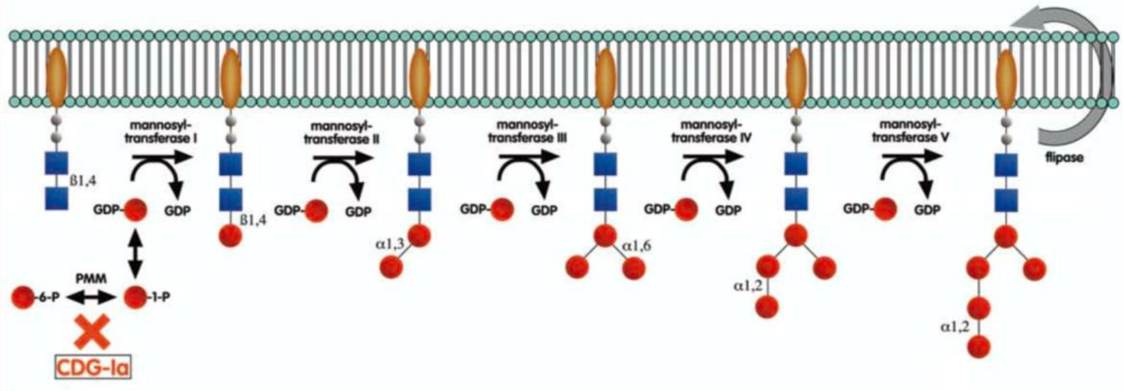


Figure 17: Early stage of posttranslational N-linked glycosylation.
Involvement of PMM2 and impairment in the case of CDG-1a are included. Black arrows indicate enzyme activity. In CDG-1a PMM2 function is impaired and affects the conversion of Man-6-P to Man-1-P, crucial for GDP-mannose formation, the mannose donor to build the core glycan dolichol-P-mannose (Dol-P-Man). Orange ellipse: Dolichol-linker. Grey circles: Phosphate. Blue squares: N-acetylglucosamine. Red circles: Mannose sugar. Figure taken from Marquardt 2003.

Once on the luminal side of the RER, the final glycan gets then transferred onto an Asn residue of the growing protein due to the cleavage of the phosphate bond between dolichol and the second GlcNAc (oligosaccharyltransferase-catalysed). After some additional trimming of the oligosaccharide branches, the glycoproteins are ready for vesicular transport to the cis-Golgi where further modifications take place. These include the partial or complete removal of mannose to finally allow the completion of the three major N-glycan residues (medial Golgi): High mannose glycans (2 GlcNAc + many mannose residues), complex glycans (2 GlcNAc + any number of other sugar residues) or hybrid glycans (2 GlcNAc + mannose and other sugar residues). Those mature N-linked glycoproteins are then finally transferred from the trans-Golgi to the plasma membrane, where they are either embedded or secreted [196,197]. Figure 18 shows an overview of the process after forming the core glycan.

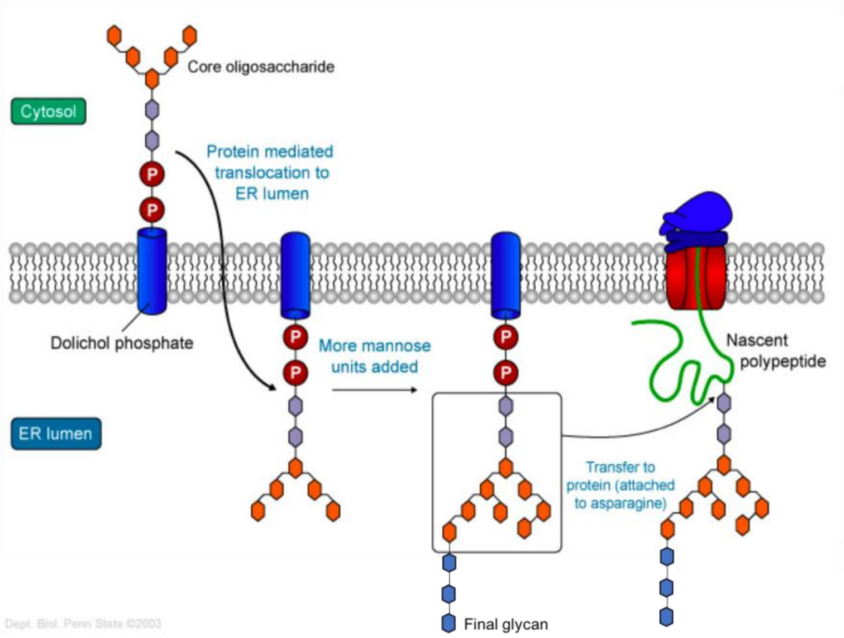


Figure 18: Posttranslational N-linked protein glycosylation.
The core oligosaccharide or glycan (Man_5-$GlcNAc_2$-PP-Dol) forms on the cytosolic side of the membrane of the rough ER (RER). It gets “flipped” across the membrane to the RER lumen where more mannose and three glucose units are added to form the final glycan, which eventually gets attached to an Asn-residue of the nascent protein. Grey hexagons: GlcNAc. Orange hexagons: Mannose molecules. Blue hexagons: Glucose molecules. Figure taken from https://wikispaces.psu.edu.

1.9 The CRISPR-Cas9 system – an overview

Since its discovery and first application in early 2013, the CRISPR-Cas9 system has revolutionised the scientific world as being the simplest and most cost-effective gene editing tool known so far, which allows for manipulation of nearly any DNA sequence in a diverse range of cell lines and organisms. CRISPR (clustered regularly interspaced palindromic repeats) sequences were discovered in some bacterial strains and identified as being part of the adaptive immune system in many bacterial and archaeal genomes against invading viruses and pathogens [198]. Those strains are able to incorporate short sequences of the foreign DNA in their genome and upon reoccurring infection by the same pathogen, the bacterial genome “memorises” the invading DNA and the CRISPR sequences get transcribed into crRNAs activating the CRISPR system to eventually silence the harmful DNA [199]. Different CRIPSR-Cas systems of bacterial and archaeal origin are known of which the CRISPR-Cas9 system from *Streptococcus pyogenes* is the most widely used and best understood [200].

The active CRISPR-Cas9 complex is formed of three major components: The Cas9 endonuclease which acts as a pair of genomic scissors, the trRNA:crRNA duplex (gRNA) which guides the Cas9 protein to an individual target site and the protospacer adjacent motif (PAM), the recognition site for Cas9 within the genome, crucial for its cutting activity. The working principle of the system relies on DNA double strand-breaks (DSB) created by Cas9 at precise genomic loci due to the target-specific nature of the gRNA. This gRNA is a short 20 nt long piece of RNA, which can be individually designed according to the targeted genomic region [199], thus making CRISPR-Cas9 such a versatile gene editing tool. Since its discovery it has been undergoing a rapid development and nowadays is an indispensable gene editing tool across many fields of biological research and industries, including the human health and agricultural sector [201].

1.9.1 DNA repair after CRISPR-Cas9 cleavage

In eukaryotic cells, DNA damage can occur at any locus, any time and can potentially result in genomic inaccuracies, which can have a significant impact on the function and behaviour of a cell. In order to prevent the occurrence of such harmful DNA alterations, eukaryotic cells have developed so called checkpoints (G1/S, G2/M) between the phases of the cell cycle to detect and repair DNA double strand breaks. [202]. Cas9-induced double strand breaks (DSBs) also activate these cellular DNA repair mechanisms within the nuclei of the targeted organism and can follow two possible repair pathways, called non-homologous end joining (NHEJ) and HDR, the latter following the principles of homologous recombination (Figure 19). NHEJ DNA repair takes place mainly during the G1 phase [202]. The simplest and most abundant mechanism of the two DNA repair mechanisms is the error-prone NHEJ repair pathway which randomly introduces *indel* mutations at the target site. This often results in the formation of DNA frameshifts and premature STOP codons causing gene knock-outs [203]. The HDR repair is of rare nature but highly precise and allows the introduction of specific nucleotide modifications ranging from a small single nucleotide substitution to the insertion of an entire gene. The supply of a DNA repair template homologous to the target site and carrying the desired mutation enables biologists to introduce the specific DNA alterations of choice into the genome in a scarless manner [204]. These DNA repair templates, also called HDR donor template or just single-stranded oligodeoxynucleotides (ssODN), are composed of two homology arms compatible to the CRISPR site, flanking the desired insertion. In the case of large insertions plasmid donors, which should at least have 1-2 kb of sequence homology for stable repair, are commonly used. For smaller changes, ssODNs are more sufficient [205]. Studies observing HDR rates at different stages in the cell cycle revealed, that HDR events are most efficient during the S-phase (39x higher than in M-phase, 24x higher than in G_1/G_0 phase). This is likely to be explained by the presence of the sister chromatid which can be used as a template for DNA repair. Studies have shown that the intracellular DNA repair mechanism, instead of recruiting the homologous chromosome, preferentially uses the sister chromatid for DSB repair. The efficiency of HDR is generally low (<10%) and the CRISPR activity generally high, the majority of DSBs will be repaired after the NHEJ pathway, as has been reported [206]. Hence, each CRISPR-Cas9 experiment will result in a genomically mixed population of cells with combinations of wt alleles, NHEJ-repaired alleles, the desired HDR-edited alleles and also heterozygous versions of these options. To date, many human cell lines and organism have been successfully modified with the CRISPR-Cas-9 system, either by introducing specific changes into the genome or via gene knock-outs [207].

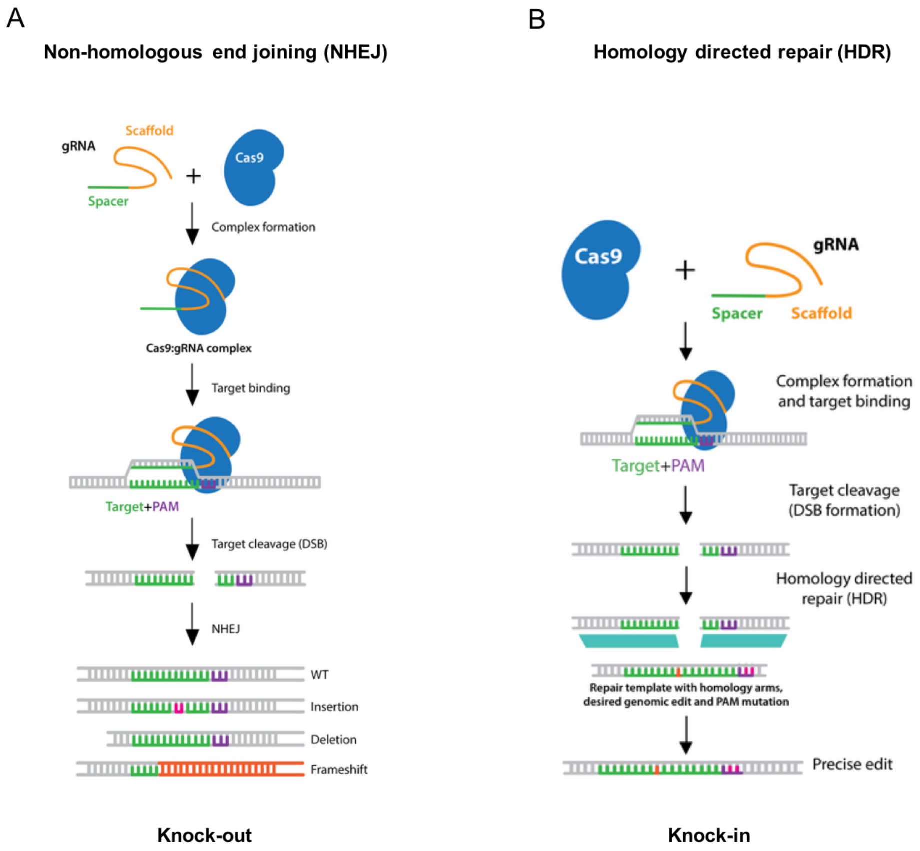


Figure 19: DNA repair mechanisms after CRISPR-Cas9 cleavage.
Target-specific crRNA (Spacer, green) and universal trRNA (Scaffold, yellow) build a RNA-duplex establishing the mature gRNA which forms a ribonucleoprotein complex with the Cas9 endonuclease. Led by the specific crRNA sequence, the mature gRNA guides the Cas9:gRNA complex to the specific target site on the genomic DNA. Cas9 protein recognises the PAM site (purple) and cuts the DNA three to four nucleotides upstream of the PAM. The CRISPR-induced DNA double strand break (DSB) activates the cellular DNA repair mechanism which can follow two routes. A NHEJ: Indels form as a result of randomly occurring insertions and deletions at the cut site to fix the cleavage often causing frameshifts and premature STOP codons, which lead to the disruption of the gene (knock-out). B HDR: In presence of a homologous DNA repair template which has a desired DNA change incorporated in its sequence, this attaches to the flanking regions of the target site with its left and right homology arms. It covers the cleavage allowing the DSB to be repaired accordingly and the desired DNA modification gets introduced simultaneously. The PAM site on the template usually gets mutated to prevent the CRISPR-Cas9 complex to bind and cleave it. Figure taken from blog.addgene.org.

Figure 20 demonstrates Cas9 and donor template DNA binding after creating the DSB at the CRISPR target site. This model is based on studies by Richardson et al. [208] who aim to explain the biophysics of the interaction between Cas9, sgRNA and the DSB to improve homology-directed genome editing.

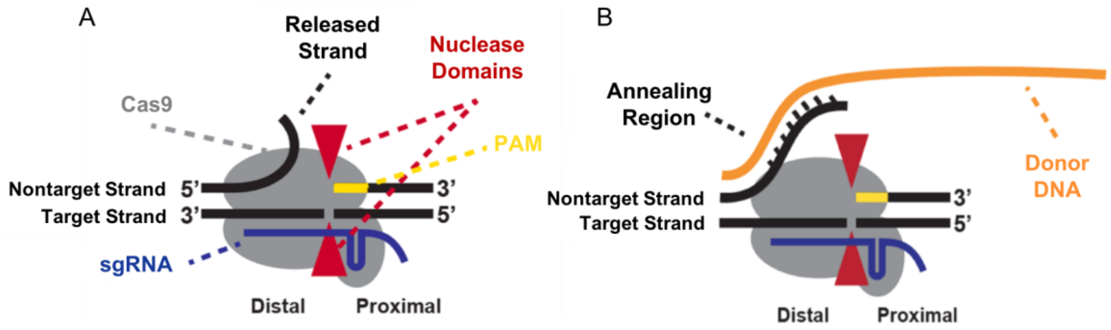


Figure 20: Cas9 and donor DNA interaction at the CRISPR site.
Target strand refers to the strand targeted by the complementary sgRNA. A: Binding of Cas9 to the cleaved DNA. The Cas9 (grey) and sgRNA (blue) complex binds tightly to three of the four ends of the DSB. The 3'-end of the nontarget strand distal of the cut is free. Red triangles indicate the nuclease domains of Cas9 performing the cut. PAM site (yellow) is proximal to the cut site on the nontarget strand. B: Interaction of the DNA donor template with its complimentary (non-target) strand and the Cas9:sgRNA complex. The donor DNA first binds to the 5'-end of the nontarget strand released from the Cas9:sgRNA complex which gets elongated whilst the three remaining ends of the DSB are still in tight complexation with the Cas9:sgRNA duplex. Figure taken from Richardson 2016 and http://blog.addgene.org/.

According to their findings after measuring Cas9 dissociation rates from the DNA target site the following process can be explained:

After Cas9 has produced the DSB, it releases the 3'-end of the leading strand, upstream of the cut first but remains attached to the 5'-end of the leading strand downstream of the cut and also to both ends of the reverse strand. This allows the single-stranded donor DNA to anneal complementary to the freely available 3'-end of the leading strand, whilst the tight complexation between Cas9:sgRNA duplex and the other three ends is maintained. This prevents the re-annealing of leading and reverse strand. A bit later Cas9 detaches and releases the 5'-end of the leading strand. The DNA now starts opening on this side and allows the HDR donor to anneal there as well. The HDR donor now bridges the gap on the site of the leading strand completely, which then gets repaired (DNA-Polymerase) and elongated (5'->3') according to the sequence of the DNA donor template, harbouring the desired change (e.g. SNP). When the elongated leading strand meets its 5'-end, the ends get ligated (ligase). The reverse strand then gets repaired according to the edited leading strand and the genomic change is introduced successfully.

1.10 Aims

This study focuses on HIPKD, a novel rare kidney disease caused by a promoter mutation in the *PMM2* gene which clinically is characterised by cystic kidneys, hyperinsulinism and in some cases also presents with liver malformations. As the disease is triggered by a newly recognised promoter mutation in the *PMM2* gene, dysfunctional gene regulation and expression are implicated as the molecular causes for HIPKD.

Overall, this work shall contribute to disclose the molecular mechanisms within HIPKD, to better understand why the phenotype settles in three distinct organs in this individual case. In particular, this PhD has focused on the functional consequences of the promoter mutation, both with respect to *PMM2* gene regulation as well as assessing the effect of altered glycosylation on insulin secretion to confirm and extend the results obtained previously [209]. In addition, it was of interest to generate a HIPKD cell model with the promoter mutation integrated in homozygous state using CRISPR-Cas9.

Aim 1: Investigate how the c.-167G>T promoter mutation affects *PMM2* gene expression.

Differences in gene and allelic expression levels on the basis of the c.-167G>T promoter variant were assessed in primary renal cells of one of the compound heterozygous patients using quantitative digital PCR. The genes (*TMEM186, CARHSP1, ABAT*) surrounding *PMM2* and two additional control genes (*PPIA, CDH1)* were also included in the gene expression assay to investigate whether their expression is also affected in HIPKD.

Aim 2: Investigate the effect potentially impaired N-glycosylation on insulin secretion.

Enhanced insulin secretion in HIPKD is assumed to be linked to defective protein N-glycosylation. To prove this, human pancreatic islets were used to mimic a glycosylation defect as assumed in HIPKD patients and were deployed in two sets of insulin secretion assays to enable subsequent measurements of the insulin levels.

Aim 3: Development of a HIPKD cell model to enable further investigation of the 3D chromatin conformation in HIPKD and to prove the hypothesis of altered chromatin loop formation.

As the promoter is poorly conserved through evolution, a human cell model for HIPKD was needed. To this end, HepG2 cells were genome edited using the CRISPR-Cas9 technology. As part of this, a homologous DNA repair template was provided to allow the cells to follow the homology directed repair pathway for integration of the HIPKD variant in homozygous state.

Chapter 2: Materials and Methods

2.1 Cell culture methods

All cell culture methods were performed under sterile conditions working under a Class II laminar flow hood.

2.1.1 Primary renal cells for digital PCR assays

2.1.1.1 Obtaining and culturing primary renal cells from compound heterozygous patient

Primary renal cell cultures were obtained from the renal cortex of a compound heterozygous patient including the c.-167G>T *PMM2* promoter mutation (c.-167G>T/c.422G>A) and a control kidney, kindly provided by Dr Jill Norman (UCL Centre for Nephrology). Cortical tissue was minced into explants <1 mm^3. Explants were plated into pre-wetted flasks (Corning) with multiple explants in each flask and cultured in a minimal volume of 4.5 g/L D-Glucose DMEM GlutaMax™-I (Gibco) supplemented with 10% fetal calf serum (FCS) (Sera Laboratories International Ltd), 1% antibiotic-antimycotic (Sigma) at 37°C, 5% CO_2. Explants were left undisturbed for at least one week to maximise attachment and the initiation of cellular outgrowth. Thereafter the medium was changed every 4-5 days taking care not to dislodge the explants from the substrate. At confluence, cellular outgrowth and explants were trypsinised and cell suspension plated in new flasks. Cells were cultured under standard conditions (37°C, 5% CO_2) with the medium changed, every 4-5 days. Cells were passaged at confluence and split 1:3.

Primary renal cells were further maintained in 4.5 g/L D-Glucose DMEM GlutaMax™-I (Gibco) with 10% fetal bovine serum (Gibco), 2.5 µg/mL Amphotericin B (PAA laboratories) and antibiotics (100 U/ml penicillin and 100 µg/ml streptomycin, Gibco) at 37°C, 5% CO_2 and grown until about 80% confluence with a regular medium change every three to four days.

2.1.2 Human pancreatic islets for deglycosylation studies

2.1.2.1 Human pancreatic islets: Samples and culture

Two batches of human pancreatic islets, which were deemed unsuitable for transplantation, were obtained from the Oxford consortium for islet transplantation (OXCIT, University of Oxford). The following information was provided for batch 1: Islets originated from a 54-year-old female donor (BMI: 31). Islet purity and viability determined was 50% and 65%, respectively. Islets were supplied three days after isolation in a suspension (~50mL) containing a total number of 13,000 islets, an equivalent of 21,000 IEQs. IEQ is considered a unit for pancreatic islets and refers to an islet volume with a diameter sphere of 150 µm. Batch 2 was obtained from a 59-year-old female donor with a BMI of 27. Purity and viability accounted for 76% and 70%, respectively. In total, 20,000 IEQ were provided one day after isolation in a 50 mL suspension. Human pancreatic islets were maintained in 5.6 mM D-Glucose CMRL 1066 transplant media without phenol red supplemented with 10% purified human albumin, 20 mM HEPES, 1% penicillin/streptomycin and 1 mM glutamine (OXSOP 5.2.7). Islets culture medium was prepared at the OXCIT and kindly provided for usage throughout the experiments. If islets were meant to be kept in culture, they were seeded (100 µL/100 IEQ seeding density) in non-coated T75 flask (Corning) for maintenance at 37°C, 5% CO_2.

2.1.2.2 Krebs-Ringer bicarbonate buffer (KRBB buffer)

Insulin secretion studies according to Assay 1 (2.3.1) were performed in KRBB (modified after [210]). Like other salt solutions, KRBB buffer maintains the physiological pH and osmotic balance and provides cells with water and inorganic ions. In one litre of ultrapure water, HEPES (25 mM), NaCl (115 mM), $NaHCO_3$ (24 mM), KCl (5 mM), $MgCl_2$ x 6 H_2O (1 mM), BSA (0,1%), and $CaCl_2$ x H_2O (2.5 mM) were combined under continuous stirring. The pH of 7.3-7.5 was adjusted with 1 M NaOH and 1 M HCl and the solution then filter sterilised, aliquoted and stored at 2 - 8°C. Prior to the experiment, KRBB buffer was supplemented with 2.8 mM and 28 mM glucose for low and high glucose-stimulated insulin secretion, respectively.

2.1.2.3 Balanced salt solution (BBS) after Gey and Gey 1936

For insulin secretion studies (2.3.2) with Batch 2 of human pancreatic islets a physiological BBS after Gey and Gey [211] was prepared. The following ingredients were added to 500 mL of ultrapure water and dissolved one by one under continuous stirring conditions: Sodium chloride (128.56 mM), potassium chloride (4.96 mM), sodium bicarbonate (27 mM), magnesium chloride hexahydrate (1 mM), disodium hydrogen phosphate dehydrate (0.84 mM), potassium dihydrogenphosphate (0.22 mM) and magnesium sulfate heptahydrate (0.28 mM). After the salts had dissolved the solution was supplemented with calcium chloride dihydrate (2 mM) and BSA (0.25 g). It was important to add the calcium chloride dihydrate bit by bit as it is a difficult to dissolve compound. The complete dissociation of all ingredients and for adjustment of a physiological pH of 7.3-7.5 the salt solution was gas-flushed with CO_2 before and after filter-sterilising (0.2 cellulose acetate filter) the solution. 40 mL aliquots of the balanced salt solution were aliquoted in 50 mL tubes and stored at 4°C. The different solutions needed for the assay were prepared prior to the experiment on the basis of the balanced salt solution and either supplemented with 2 mM glucose or 20 mM glucose. For deglycosylation, 4000U/mL Peptide N-Glycosidase F (PNGase F) and 4000U/mL Endoglycosidase H (Endo H) were added to a part of the 2 mM glucose solution and half of the 20 mM glucose solution was supplemented with PMA.

2.1.2.4 Radioimmunoprecipitation assay buffer (RIPA buffer)

Human pancreatic islets were lysed in radioimmunoprecipitation assay buffer (RIPA lysis buffer) composed of Tris-HCl pH 8 (50 mM), NaCl (150 mM), NP-40 (1%), Sodium deoxycholate (0.5%), SDS (0,1%), Protease inhibitor and combined in ultrapure water for intracellular insulin measurement.

2.1.3 HEK293T cells and HepG2 cells for genome editing with CRISPR

2.1.3.1 HEK293T cells

Culture of HEK293T cells

HEK293T cells (Davies lab) were routinely cultured in T25 flasks in MEF medium composed of DMEM high glucose (Sigma) supplemented with 10% FCS (Gibco), 1x L-Glutamine (Sigma) and antibiotics (100 U/ml penicillin and 100 µg/ml streptomycin, Sigma) and maintained at 37°C and 5% CO^2 fumigation. Cells were subcultured every two days with a subcultivation ratio of 1:3 cells per T25 flask. After carefully aspirating the medium, cells were washed ones with 1x PBS and incubated with 1 mL 0.05% Trypsin-EDTA (Sigma) per T25 flask for 2 minutes for detachment. To stop the reaction

4 mL of MEF medium were added and the cell suspension was pelleted for 5 minutes at 1,000 rpm in 15 mL conical tubes. Cells were suspended in MEF medium, composed of DMEM supplemented with and seeded after counting with a hemocytometer. For cryopreservation, cell culture medium was supplemented with 10% dimethylsulfoxide (DMSO, Sigma) and cells were stored in Nunc® CryoTubes at -80°C. When taken from cryostocks, HEK293T cells were thawed quickly in a water bath at 37°C for 90 seconds and immediately transferred into culture flask containing MEF cell culture medium and kept at 37°C, 5% CO^2.

Semi-reverse transfection of HEK293T cells

The assessment of two CRISPR-Cas9 expression constructs (Table 12), each carrying a target-specific gRNA and the Cas9 gene, was first tested in HEK293T cells, as these cells are known to be “easy to transfect”. The constructs were delivered to the cells in a semi-reverse transfection whilst cells are in suspension using the FuGENE® HD transfection reagent (Promega).

Beside the two CRISPR constructs to test, three controls were incorporated. The GFP-expressing plasmid pCAGFP [212], a CRISPR construct which was used successfully in a previous, independent experiment and non-transfected to compare as a negative control. Transfections were carried out in 24-well plates (Corning). After detaching the cells, 80,000 cells per well in a total volume of 500 µL MEF medium without antibiotics were transfected with 550 ng of plasmid DNA and DNA:FuGENE® HD ratio of 1:5. The transfection mix was prepared in serum-free Opti-MEM™ medium (Gibco) in sterile 1.5 mL reaction tubes and plasmid DNA and Opti-MEM™ together were made up to a final volume of 25 µL. To prewarm the FuGENE® HD reagent, 2.5 µL were added to the Opti-MEM™ and incubated for 5 minutes prior of adding the plasmid DNA. To promote FuGENE® HD and plasmid DNA complexation the content was gently mixed by flicking the tube and then incubated for 15 minutes at RT. Meanwhile, the cells were prepared and seeded in the 24 wells and the transfection mix then added dropwise to the cells. Transfected HEK293T cells were incubated 48 hours at 37°C and 5% CO_2 before lysis (s. 2.4.1.2.).

2.1.3.2 HepG2 cells

Culture of HepG2 cells

HepG2 cells (ATCC® HB-8065™) were routinely cultured in T25 flasks in α-MEM medium with ribonucleosides and desoxyribonucleosides (Gibco) supplemented with 10-

20% FCS, antibiotics (100 U/ml penicillin and 100 µg/ml streptomycin, Sigma), 1x non-essential amino acids (Sigma), 1x GlutaMAX™ Supplement (Gibco), 1x L-Glutamine (Sigma), 25 mM D-Glucose (Sigma), 50 µg/mL Linoleic acid bovine serum albumin (Sigma), 0.4 µg/ml Hydrocortisone (Sigma), 0.04 µg/mL Thyroid Releasing Hormone (Sigma), 10 µg/mL Insulin (Sigma) and 0.002 µg/mL Sodium Selenite (Sigma) and maintained at 37°C with 5% CO^2. HepG2 cells were passaged every 3-4 days with a seeding density of 750,000 cells per T25 flask. For detachment, cells were washed twice with 1x PBS (Gibco) and treated with 1.5 mL 0.25% Trypsin-EDTA (Gibco) per T25 flask for 4-7 minutes. The reaction was terminated by adding the double amount of cell culture medium containing serum. In order to obtain a single cell suspension and to avoid clumping, cells were passed through a 21G needle before seeding. Cell culture medium was changed 24 hours after subculturing. For cryopreservation, cell culture medium was supplemented with 5% DMSO and cells were stored in Nunc® CryoTubes at -80°C. When taken from cryostocks, vials with cells were thawed quickly in a water bath at 37°C for about 90 seconds, transferred into a 15 mL conical tube containing 9 mL of HepG2 culture medium and spun for 5 minutes at 1,000 rpm. Immediately after, the cell pellet was resuspended in HepG2 cell culture medium and cells were transferred into a T25 culture flask containing a total of 7 mL HepG2 cell culture medium and kept at 37°C, 5% CO_2. HepG2 cells were tested negative for mycoplasma contamination.

Semi-reverse transfection and co-transfection of HepG2 cells

In order to create the desired SNP (PMM2 c.-167G>T) in HepG2 cells, the CRISPR-Cas9 expression construct of choice and a specifically designed ssODN were delivered to the cells in a semi-reverse co-transfection.

Semi-reverse transfections and co-transfections were done using the FuGENE® HD transfection reagent and the protocol was optimised for HepG2 cells. At first, different amounts of the CRISPR plasmid DNA (550 ng, 800 ng and 1 µg) only were transfected to assess the right proportion of plasmid DNA to add. Later on, different amounts of 10 µM ssODN working stock (800 ng, 400 ng, 320 ng and 200 ng) were delivered to the cells together with the CRISPR plasmid in semi-reverse co-transfection experiments. For both, semi-reverse transfections and co-transfections, experiments were performed in 24 well plates. Cells were trypsinised and separated as described before and 250,000 cells were seeded per transfection in a total volume of 500 µL HepG2 culture medium without antibiotics. The transfection mix was added immediately after plating the cells.

The optimal DNA:FuGENE® HD ratio to use was 1:6. A total amount of 1 µg plasmid DNA were used per transfection and combined with 6 µL of FuGENE® HD. The transfection mix was prepared in serum-free Opti-MEM™ medium (Gibco) in sterile 1.5 mL reaction tubes. Plasmid DNA and Opti-MEM™ together were made up to a final volume of 25 µL. Before adding the plasmid DNA to the mix, 6 µL of the FuGENE® HD reagent were added to the Opti-MEM™ and incubated at RT for 5 minutes to prewarm the reagent. After adding the right amount of plasmid DNA, the transfection mix was gently mixed by flicking the tube and then incubated for additional 15 minutes at RT. In case of a co-transfection, 400 ng (1 µL) of the 10 µM ssODN were added to the mix additionally, prior the 15-minute incubation. In the meantime, the cells were seeded onto the wells and the transfection mix then added dropwise to the cells. Transfected HepG2 cells were incubated 48 to 72 hours at 37°C and 5% CO_2. During each experiment, at least one sample of cells got transfected with the GFP-expressing plasmid pCAGFP[213213213213213213213213213213213] to serve as a positive control and one or two wells of non-transfected cells were used as a negative control (MOCK). Both controls were also used to set up the FACS cell sorter later on.

2.1.4 Monitoring transfection efficiency of GFP-expressing cells

GFP expression of delivered plasmids to HEK293T and HepG2 cells and CRISPR constructs was monitored 24–72 hours after semi-reverse transfections and co-transfections using the ZOE™ Fluorescent Cell Imager (Bio-Rad). Pictures were taken from the brightfield and the green fluorescent channel, to be able to compare cell density and GFP-signal.

2.1.5 Electroporation of HepG2 cells (CRISPR-RNP approach)

When the CRISPR/Cas9 system was delivered to HepG2 cells as ribonucleoproteins (RNPs) together with the ssODN, electroporation was used. RNPs and HDR donor were transferred into the cells with the Neon® Transfection System (ThermoFisher) using the NEON® 10 µL electroporation tips provided by the Kit. In total, four samples were incorporated in the experiment: The sgRNA/Cas9 complex and the ssODN were delivered to the cells in duplicate next to a control electroporation with the GFP plasmid pCAGFP [213] only and a MOCK sample of non-electroporated cells.

Three to four days after passaging, HepG2 cells were harvested as described previously. For each electroporation ~1.4*10^6 HepG2 cells were used and plated afterwards onto 24 wells in 500 µL – 1 mL of HepG2 culture medium without antibiotics. The 24-well plate was prepared beforehand and put in the cell culture incubator (37°C, 5% CO^2) to prewarm the medium and adjust the pH. For each electroporation, a transfection mix was prepared in 1.5 mL tubes consisting of ~1.1 µg sgRNA (0.6 µL), 20 µM Cas9 protein (1 µL) and 4.4 µL Buffer R and the mix was then incubated for 15 minutes at RT to allow RNP complexes to form (RNP mix). Meanwhile, the ssODN, HepG2 cells and electroporator were prepared. First, 3 µL of the ssODN (100 µM stock) were diluted in 7 µL Buffer R. The NEON® pipette station was put under the laminar flow hood and the NEON® tube (contains electrode) was inserted. 3 mL of Buffer E were then added to the NEON® tube and the electroporation programme set to the following parameters: 1500 V, 30 ms, 1 pulse. Next, HepG2 cells were counted and the adequate number of cells per experiment needed was collected at 1,000 rpm for 5 minutes and the pellet dissolved in the right amount of Buffer R (7 µL per electroporation plus 1x excess) provided by the Neon® kit. This was done just before electroporation, to keep cell stress as low as possible. Suspended cells had to be kept on ice prior to usage. When the 15-minute incubation time was up, 1 µL of the ssODN dilution was added to each RNP mix and 7 µL cell suspension were added to this mix. With the electroporation pipette and the NEON® 10 µL electroporation tip, the cell/RNP/ssODN mix was mixed carefully (2-times up and down) and the whole volume then taken. Here, it was crucial to avoid bubble formation inside the NEON® tip. The tip was placed inside the NEON® tube containing Buffer E and the programme was started. Immediately after electroshocking, the cells were seeded in one well of the previously prepared 24-well plate and the plate put in the incubator at 37°C, 5% CO^2. After 3 days of incubation, CRISPRed HepG2 cells were harvested as described previously and three-quarters put in cryopreservation at -80°C as a backup. The remainder was used to extract gDNA to assess the cutting efficiency and overall success of the experiment (s. 2.4.6). In addition, 3x 5,000 cells were seeded in ultra-low density on 10 cm dishes and incubated at 37°C, 5% CO^2 for about two weeks, to obtain single cell colonies (s. 2.4.7) for genotyping (s. 2.4.8).

2.2 Quantitative digital PCR methods

Using quantitative digital PCR (dPCR), a series of gene expression assays (GEA) and an allele-specific assay (ASA) were performed. For both sets of experiments, cDNA obtained from primary renal cell cultures from a compound heterozygous patient and a healthy person (control) was used as a template in dPCR. For further details regarding primary renal cell culture and origin, RNA isolation and cDNA synthesis please go to sections 2.2.1, 2.2.2 and 2.1.1.

Quantitative differences in the overall *PMM2* gene expression of the HIPKD patients were analysed by GEA. Besides *PMM2*, the expression levels of three additional genes, *TMEM186*, *CARHSP1* and *ABAT*, the neighbour genes of *PMM2,* were analysed to explore whether these are affected by the *PMM2* promoter mutation. In addition, the expression of two housekeeping genes, *CDH1* and *PPIA* was measured as well.

In addition to the c.-167G>T SNP on one allele within the *PMM2* promoter, the compound heterozygous patient also carries a SNP within the coding region of *PMM2* (c.422G>A; Ref.-Seq.: NM_000303.2) located on the opposite allele controlled by the wt *PMM2* promoter. The nature of these two mutant *PMM2* alleles (Allele 1 and Allele 2) in the compound heterozygous patient was used to design an allele specific assay (ASA) for dPCR. This has allowed the investigation of potential alterations in expression levels of the two *PMM2* alleles due to the c.-167G>T promoter mutation on Allele 2. Here, also cDNA served as the input material. Quantification was implemented using two specific differently labelled probes, both spanning nt position c.422 (Ref.-Seq.: NM_000303.2) on both alleles. To distinguish between the two *PMM2* alleles, the SNP c.422G>A was used and incorporated in the probe design (Table 5) for the respective allele (Allele 1). Figure 21 shows how the two ASA TaqMan® probes were designed to bind at the *PMM2* target locus c.422G>A.

Figure 21: Scheme depicting TaqMan® probe-binding for allele specific discrimination (ASA) in a compound heterozygous patient.
The red asterisk indicates the PMM2 promoter SNP G>T at position c.-167 on Allele 2 (MUT promoter) of the PMM2 gene. The blue bar indicates probe binding overspanning the wt version of position c.422G (Allele 2 with MUT promoter). The green bar specifies probe binding to allele 1 (WT promoter) carrying the G>A SNP at position c.422. The blue and green asterisks mark probe-labelling with the two fluorophores FAM™ (green) and VIC® (blue). MUT = mutant. WT = wildtype.

Absolute quantification of differential gene expression in primary renal cells was implemented and analysed using the chip-based QuantStudio™ 3D Digital PCR system (ThermoFisher) and the QuantStudio™ 3D Analysis Suite Cloud Software (ThermoFisher). All specific TaqMan® Genotyping assays were purchased from Thermo Fisher Scientific, Paisley, UK. The TaqMan® probes are covalently labelled with a fluorophor (FAM™or VIC®) at the 5'-end for detection. One allele-specific assay and six gene expression assays were included in this study (Table 7).

2.2.1 RNA isolation from primary renal cells

In order to perform quantitative digital PCR gene expression experiments, RNA was isolated from primary renal cell cultures, derived from a patients' and a control (wt) kidney at different passages and days (Table 4) using TRIzol® (ThermoFisher) according to the manufacturer's protocol for adherent cells. RNA was dissolved in ultrapure DEPC-treated Water (Sigma). The concentration of isolated RNA (Table 4) was determined using the NanoDrop™ ND1000 spectrophotometer (ThermoFisher). RNA was stored at -80°C.

2.2.2 cDNA synthesis from RNA samples

Each of the RNA isolates was transcribed into cDNA by a one-step reverse transcription (RT) reaction using the iScript™ Reverse Transcription Supermix for RT-qPCR Kit (Bio-Rad). A total RNA input of 1,000 ng per RNA sample was transcribed into cDNA resulting in a total cDNA amount equivalent to 50 ng/µL of RNA per RNA sample. As suggested by the manufacturers' protocol a pooled no reverse transcriptase control for patient and control and a no template control (for ASA) were used as negative controls. cDNA samples were stored at -20°C until usage in digital PCR experiments.

Table 4: Patient (grey) and control (white) primary renal cell samples used for RNA isolation

Sample Kidney	Passage	RNA concentration	Assay used
Patient 1 (P1)	Passage 4	151.9 ng/µL	ASA/GEA
Patient 2 (P2)	Passage 3	171.4 ng/µL	ASA/GEA
Patient 3 (P3)	Passage 3	200.6 ng/µL	ASA/GEA
Patient 4 (P4)	Passage 1	423.6 ng/µL	ASA/GEA
Control 1 (C1)	Passage 4	319.7 ng/µL	GEA
Control 2 (C2)	Passage 4	278.5 ng/µL	GEA
Control 3 (C3)	Passage 4	318.8 ng/µL	GEA

2.2.3 TaqMan® assays

2.2.3.1 Allele specific assay

Allelic discrimination was implemented using a target-specific Custom TaqMan® SNP Genotyping Assay (ThermoFisher) designed through the Thermo Fisher TaqMan® Assay Design Tool. The assay contains a primer/probe mix for amplification and detection of the targets. For allelic discrimination two differently fluorophore-labelled probes were present in the assay. ASA probe and primer sequences and probe labelling are listed in Table 5 and Table 6.

Table 5: TaqMan® assay used for allele specific assay in dPCR setting

Assay ID	Target gene	Sequence 5'-3'	Fluorophore at 5'-end (Dye)
AHGJ9ED	*PMM2*	CCAAGAAGAAC**G**CATTGA-3'-NFQ-MGB	FAM™ (green)
	PMM2	CCAAGAAGAAC**A**CATTGA-3'-NFQ-MGB	VIC® (blue)

The following primer set was part of the allele specific assay for amplification in dPCR.

Table 6: Primer contained in TaqMan® assay AHGJ9ED used for allele specific assay

Target	**Direction**	**Sequence 5'-3'**
PMM2	Forward	GTGTCCCCTATTGGAAGAAGCT
PMM2	Reverse	CTTTTGTCTTATATTTTCTTTTTTATCGAGTTCGTAGA

2.2.3.2 Gene expression assay

TaqMan® gene expression assays used for expression studies (Table 7) were ordered predesigned. More detailed information regarding approximate probe-binding motif within the gene of interest can be ascertained through the assay IDs.

Table 7: TaqMan® assays used for quantitative dPCR gene expression assays

Target gene	**Assay ID**	**Fluorophore at 5'-end (Dye)**
PMM2	Hs00756707_m1	FAM™
TMEM186	Hs01049388_m1	FAM™
CARHSP1	Hs00183933_m1	FAM™
ABAT	Hs00609436_m1	FAM™
PPIA	Hs04194521_s1	VIC®
CDH1	Hs01013955_m1	FAM™

2.2.4 Control sample for allele specific assay

As a control, two linearized plasmid constructs [pcDNA™3.1(-) PMM2_wt.dna; pcDNA™3.1 (-) PMM2_mut.dna] were used in a 1:1 ratio, mimicking the two *PMM2* alleles of a patient being heterozygous for the c.422G>A SNP but both being under control of a wt promoter (done by Dr Daniela Iancu, UCL Centre for Nephrology).

2.2.5 Digital PCR sample preparation

For the final reaction, cDNA samples and master mix were prepared separately for each assay. All cDNA/plasmid samples were made to a volume of 6 µL using nuclease free water (Sigma). The master mix contained the Quantstudio® 3D Digital PCR Master Mix, 20x TaqMan® assay and nuclease-free water and were made up to the number of chips to be loaded per sample. Then, cDNA sample and master mix were combined and a total volume of 14.5 µL was loaded on each chip according to the manufacturer's recommendations (ThermoFisher).

2.2.6 PCR program

All dPCR reactions were performed using the GeneAmp PCR system 9700 (ThermoFisher) and samples were amplified with the following conditions: Initial denaturation at 96°C, 10 mins followed by 39 cycles of 60°C, 2 mins and 98°C for 30 sec and final 2 mins at 60°C for extension.

2.3 Insulin secretion assays with human pancreatic islets

The effect of deglycosylation on the function of human pancreatic islets was examined by, firstly, conducting two different deglycosylation assays (Assay 1 and Assay 2) using two batches of human pancreatic islets to then measure insulin secretion in a human insulin ELISA assay. Both assays had two main stages, a deglycosylation period evoked due to the addition of glycosidases and a glucose-stimulated insulin secretion series (GSIS). The main differences between the assays were 1) the order in which the two treatment periods (Deglycosylation and GSIS) took place, 2) for the second assay (Assay 2) the PKC activator PMA (phorbol-12 myristate-13 acetate) to enhance insulin secretion was used in addition and 3) the process itself. Whereas during the first assay (Assay 1), the islets were exposed to the different stimuli in a flowing sequence (transfer from well to well) and each sample had passed all the different treatment stages, in the second assay islets were split in three parts and kept separately, one group for each condition of the GSIS part. Figure 22 and Figure 23 illustrate the principle and performance of both assays. During the first assay (2.3.1) deglycosylation took place after a series of glucose-stimulated insulin secretion (2.8 mM glucose and 28 mM glucose) with Peptide N-Glycosidase F (PNGase F, NEB). In the second assay (2.3.2), deglycosylation was evoked by using two glycosidases, PNGase F and the more specific Endoglycosidase H (Endo H) and was conducted before glucose stimulation testing three conditions (2 mM glucose, 20 mM glucose, 20 mM glucose plus PMA).

2.3.1 Assay 1: Glucose-stimulated insulin secretion (GSIS) and PNGase F treatment

The glucose-stimulated insulin secretion and subsequent PNGase F treatment series was performed using human pancreatic islets Batch 1. After arrival, the islet suspension was placed in the incubator for islets settlement. Then islets were suspended in 5 mL medium and 1,500 IEQ (volume calculated) were seeded in a 10 cm non-coated petri dish containing 15 mL of islets culture medium. Subsequently, islets were incubated ON at 37°C, 5% CO_2 prior to start the experiment. The remaining islets were kept in culture as described in section 2.1.2.1. The experiment was done in 24-well plates (NUNC) with Millicell® cell culture inserts (Merck Millipore) for islets transfer. Before islets seeding, the 24-well plates were prepared as shown in Figure 22. KRBB buffer and previously prepared KRBB glucose solutions were filled in 24-wells (1.3 mL per well) and those then equilibrated for 60 mins at 37°C and 5% CO_2 to adjust the pH and temperature of the bicarbonate-containing KRBB buffer. In total, four samples (3x experimental condition, 1x control) were assayed. The ON culture was harvested in a 15 mL tube. After islets settlement on the bottom of the tube, the majority of the medium was removed and only 1.5 mL were left. In this way an islets concentration of 100 IEQ/100 µL was achieved to precisely place the desired number of IEQs (300 IEQ/300 µL per insert) in Millicell® cell culture inserts. Prior to the assay, seeded islets were washed and equilibrated in the KRBB 2.8 mM glucose solution for one hour at 37°C and 5% CO_2. For GSIS, all four samples were incubated in 1) low (2.8 mM), 2) high (28 mM) and again 3) low glucose (2.8 mM). For the last condition to test (4), three islets samples were exposed to 2,000 U PNGase F in KRBB buffer without the addition of glucose (0 mM glucose). The fourth sample was used as a control. Before adding it to the wells, PNGase F was mixed with 100 µL of its dedicated enzyme buffer (10x GlycoBuffer 7 “G7”, 500 mM Na_3PO_4) and subsequently added to the three wells containing KRBB buffer only (1:10 dilution). The control sample was supplemented with the G7 enzyme buffer to keep conditions consistent. Each treatment lasted for 60 mins under known culture conditions (37°C, 5% CO_2) and when the time was up, islets were transferred into the next well. For measurement of secreted insulin, medium samples were taken before (T_0, 300 µL) and after the 60-minute incubation (T_{60}, 1 mL).

Finally, islets were harvested in 1 mL radioimmunoprecipitation assay buffer (RIPA lysis buffer, 50 mM Tris-HCl pH 8; 150 mM NaCl; 1% NP-40, 0.5% sodium deoxycholate; 0,1% SDS; Protease inhibitor; 18.2 MΩ H2O) for intracellular insulin measurement. Cell lysates and medium samples were stored at -80°C and -20°C, respectively for analysis with ELISA later on (2.3.3).

Figure 22: Experimental set up for human pancreatic islets Batch 1 (Assay 1).
Plate preparation and experimental process for glucose-stimulated insulin secretion and PNGase F treatment assay. Low Glu: KRBB buffer plus 2.8 mM glucose (1.3 mL/well). High Glu: KRBB plus 28 mM glucose (1.3 mL/well). PNGase F (2000 U in 10x G7 enzyme buffer, 500 mM Na_3PO_4) in KRBB without glucose (1.3 mL/well). G7: 10x G7 enzyme buffer only in KRBB without glucose (1.3 mL/well). RIPA: Grey arrows indicate the transfer of the islets throughout the experiment. Blue arrow: Total amount of time of islets stimulation.

2.3.2 Assay 2: Deglycosylation assay with PNGase F and Endo H for insulin secretion studies

Insulin secretion studies following deglycosylation with PNGase F and, in addition, Endo H were done using human pancreatic islets Batch 3. Differently to the previous set up, the deglycosylation of human pancreatic islets took place prior to glucose-stimulated insulin secretion during this experiment. In total, six independent islets samples were treated with deglycosylating enzymes in a first experimental stage. In parallel, another six independent non-deglycosylated islets samples were used as a control group for direct comparison. During the second stage of the assay both sets of islets, deglycosylated and non-deglycosylated, were split in three groups for glucose stimulation, testing islets response to the following three conditions: 1) 2 mM glucose for basal islets stimulation, 2) Exposure to 20 mM glucose for enhanced stimulation and 3) Exposure to 20 mM glucose plus PMA to boost enhancement of insulin secretion additionally. For each condition, two independent samples of deglycosylated islets and two independent samples from the control group were available, allowing subsequent insulin measurements in duplicate. The assay was performed in 24-well plates which was prepared as shown in Figure 23 and 1 mL of the previously prepared glucose solution added per well. Before seeding the islets, the plate was equilibrated for one hour in the cell culture incubator (37°C and 5% CO_2).

After collection, islets were maintained in a non-coated T75 flask in their dedicated culture medium ON (2.1.2.1). Prior to the experiment, 24-well Millicell® cell culture inserts (Merck) were placed in 24 well plates. Human pancreatic islets were then seeded onto the polycarbonate membrane of the insert at a density of 150 islets/well. To deglycosylate cell surface proteins the cells were incubated at 37°C for 120 minutes, in balanced salt solution (2.1.2.3) prepared after Gey and Gey, supplemented with 2mM glucose in the presence or absence of 4000U/mL PNGase F and 4000U/mL Endo H. Immediately following deglycosylation the cells were washed and transferred into the next wells for glucose-stimulated insulin secretion. Islets were exposed to three different conditions as mentioned above. Islets were incubated for 60 minutes (37°C, 5% CO_2) in balanced salt solution supplemented with 2mM glucose, or 20mM glucose, or 20mM glucose plus the protein kinase C activator PMA (500 nM). After incubation, the medium of each well was collected and stored at -20°C for the measurement of secreted insulin with ELISA. In order to determine the intracellular insulin content, all islets samples used in this assay were harvested in RIPA cell lysis buffer (2.1.2.4) after the experiment was finished and carefully sonicated (Hielscher) on ice (3x 12 pulses, 60% amplitude) to ensure the breakdown of the cells. Islets lysates were frozen at -80°C for intracellular insulin measurements.

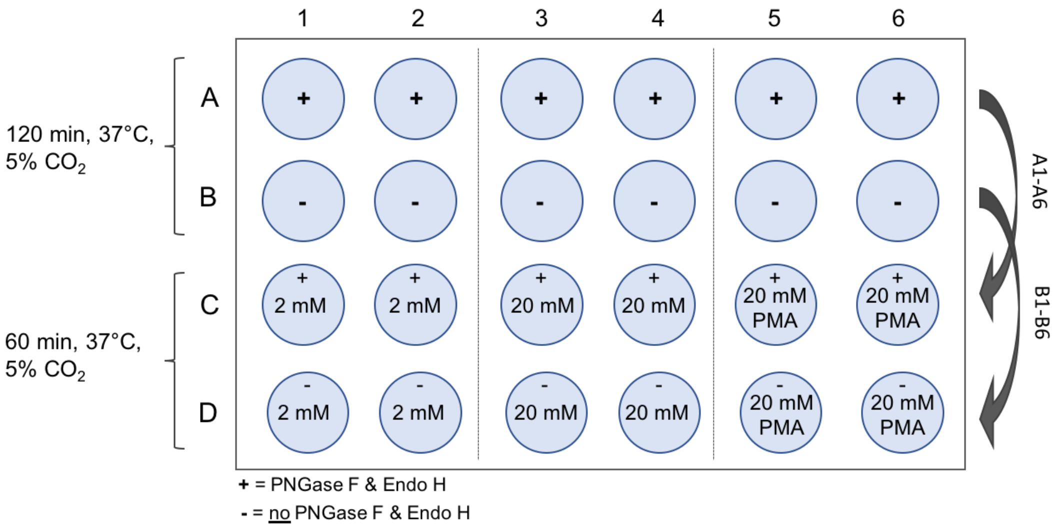


Figure 23: Experimental set up for human pancreatic islets Batch 2 (Assay 2).
Plate preparation and experimental process for deglycosylation studies with PNGase F, Endo H and subsequent glucose-stimulated insulin secretion with PMA. All solutions were prepared in balanced salt solution (BSS) and 1 mL were added per 24-well. +: with PNGase F and Endo H. -: Without PNGase F and Endo H. A: Deglycosylation. 150 IEQ/well were seeded in BSS containing 2 mM glucose with PNGase F and Endo H. B: Control (non-deglycosylated). 150 IEQ/well were seeded in BSS containing 2 mM glucose without PNGase F and Endo H. Rows A and B were incubated for 120 minutes at 37°C and 5% CO_2. After incubation, islets were transferred into rows C and D. C and D 1-2: Glucose-stimulated insulin secretion condition 1 (Basal stimulation). BSS containing 2 mM glucose with (+) or without (-) PNGase F and Endo H. C and D 1-2: Glucose-stimulated insulin secretion condition 1 (Basal stimulation). C and D 3-4: Glucose-stimulated insulin secretion condition 2 (Enhanced stimulation). BSS containing 20 mM glucose with (+) or without (-) PNGase F and Endo H. C and D 5-6: Glucose-stimulated insulin secretion condition 3 (PMA-enhanced stimulation). BSS containing 20 mM glucose and 500 nM PMA with (+) or without (-) PNGase F and Endo H. Rows C and D were incubated for 60 minutes at 37°C and 5% CO_2.

2.3.3 ELISA assay for Human insulin measurement

The medium samples and lysates collected from both glucose-stimulated insulin secretion and deglycosylation assays with human pancreatic islets Batches 1 and 2 (2.3.1, 2.3.2) were analysed in an enzyme-linked immunosorbent assay (ELISA) using the Human Insulin ELISA Kit (Millipore). The sample volume always accounted for 20 µL. Blanks (NSB, RIPA buffer) and standards (2, 5, 20, 50, 100, 200 µU/mL) were measured in duplicates (20 µL) and used for correction of sample absorbance values after reading. To guarantee accuracy of the measurement two low and high insulin quality controls, QC1 (concentration range: 7.3-15 µU/mL) and QC2 (concentration range: 19-39 µU/mL), were measured in parallel. Each sample (secreted and intracellular) and also the quality controls were measured once. The assay was performed according to the manufacturer's instructions and read at 450 nm (Beckman Coulter, DTX 880). When outside of the range of the standard curve, samples were diluted accordingly prior to measurement. Differences in insulin secretion between treatment groups were assessed by one-way or two-way ANOVA and post hoc Bonferroni or student’s t-test, respectively.

2.4 CRISPR-Cas9 genome editing methods

To generate the desired HepG2 cell line, a CRISPR/Cas9 vector-based strategy was chosen for scarless gene editing. In order to be able to track and select transfected cells, the Cas9 expression vector pSpCas9(BB)-2A-GFP (Addgene plasmid # 48138, PX458) [203], kindly provided by Feng Zhang, was used. The CRISPR vector contains the hSpCas9 gene (insert) from *S. pyogenes* and an EGFP at the C-terminal site of the insert. Due to digestion of the vector with *BbsI*, the target-specific gRNA oligonucleotides could simply be ligated into the backbone by adding matching 5'-overhangs. *BbsI* is a Type II restriction enzyme which cuts a few nucleotides upstream its recognition sites and therefore got replaced by the inserted double-stranded gRNA oligo. Figure 24 depicts a schematic overview of the cloning strategy when working with the Cas9 expression vector pSpCas9(BB)-2A-GFP (PX458).

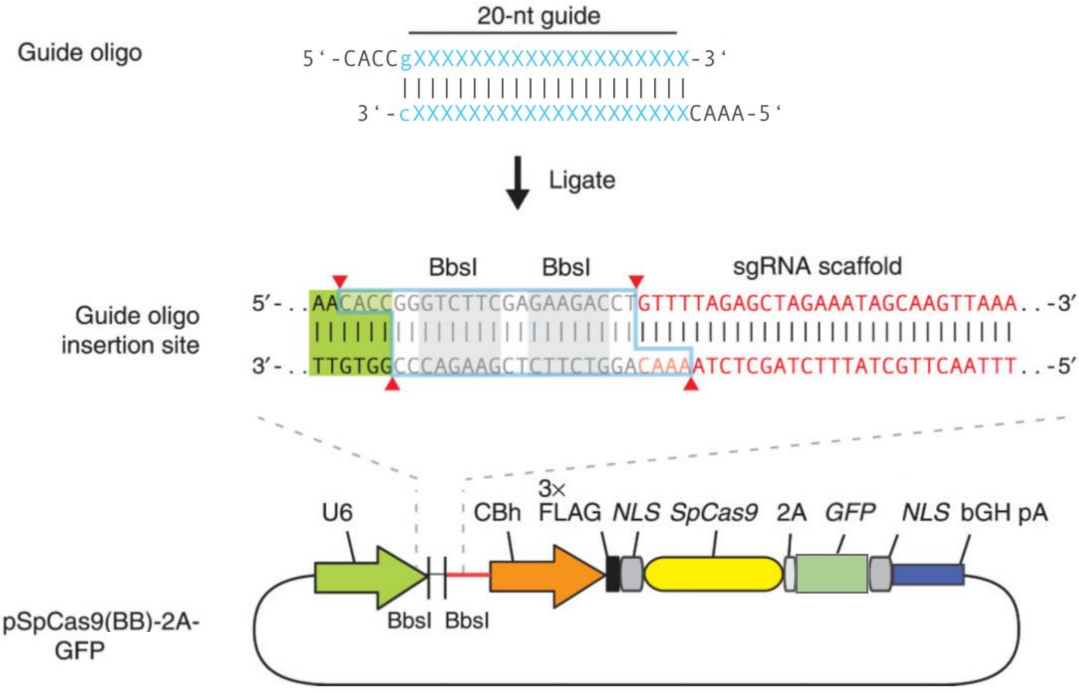


Figure 24: Schematic overview showing the insertion of the target-specific gRNA oligo using the Cas9 expression vector pSpCas9(BB)-2A-GFP (pX458).
Top: Double-stranded guide oligo composed of the individual target-specific 20-nt guide sequence (blue Xs), an additional guanine (blue g), if not present naturally and a matching 5'-overhang complementary to *BbsI* cut site in pX458 for ligation. Middle: Guide oligo insertion site and universal sgRNA scaffold sequence. Bottom: pSpCas9(BB)-2A-GFP with all its features from left to right: U6 promoter, *BbsI* guide oligo ligation site, Cbh promoter, 3x FLAG tag, NLS nuclear localisation signal, *S. pyogenes* Cas9 gene, viral 2A self-cleaving peptide, green fluorescent protein GFP, NLS, bovine growth hormone polyadenylation signal (bGH pA). Taken from Ran et al., 2013

2.4.1 Molecular biology methods for genome editing with CRISPR-Cas9

2.4.1.1 DNA sequences and primer design

Human *PMM2* promoter sequence was obtained from the UCSC genome browser (https://genome.ucsc.edu/) and corresponds to the human genome assembly GRCh38/hg38. Primers were chosen directly from the DNA sequence and analysed using the Thermo Fisher Tm calculator (https://www.thermofisher.com/uk/en/home.html) for appropriate annealing conditions.

2.4.1.2 Genomic DNA extraction from HepG2 and HEK293 cells

Genomic DNA (gDNA) was obtained from pooled or clonal single cell populations in order to sequence verify and genotype the *PMM2* CRISPR site to evaluate either the overall outcome of the CRISPR gene editing (cell pool) or to identify positive clones (single cell colonies) post-transfection. When the cells reached a decent confluency, lysis buffer (0.1% SDS, 400 µg/mL Proteinase K, 40 µg/mL RNAse A) was either added directly to 96 well plates or cells were harvested beforehand and lysed in reaction tubes. In both cases, cell lysis took place ON at 37°C. For precipitation, 1/10 volume of an 8 M LiCl solution and one volume isopropanol were added and lysates were mixed thoroughly. Samples or 96 well plates were spun for 10 to 15 minutes at 13,000 rpm, the supernatant removed, gDNA pellets then washed 1x with 500 mL 70% ethanol and spun again under previous conditions. Pellets were placed at 37°C for 5 to 10 minutes to evaporate remaining ethanol. gDNA was resolved in 10 to 50 µL ultrapure water, depending on the size of the pellet, quantified using NanoDrop™ and stored at -20°C.

2.4.1.3 Hot-Start PCR

For amplification of the *PMM2* CRISPR site in HepG2 cells (and also HEK293T cells) and genotyping of picked single cell HepG2 colonies, a Hot-Start PCR protocol was chosen, using the IMMOLASE™ DNA Polymerase Kit (Bioline, Cat. No. BIO-21046) to increase DNA yield and purity. IMMOLASE™ is a heat-activated thermostable DNA polymerase which is inactive at room temperature. It provides extremely high yields of DNA and significantly reduces non-specific products (primer-dimers, mis-primed products) compared to standard polymerases. All reactions were supplemented with Q-solution (Qiagen), a substance which additionally improves the PCR reaction by modifying the melting behaviour of the DNA. This is in particular beneficial when working with GC-rich templates or templates with a high degree of secondary structure. The formulation of Q-Solution is unknown.

All reactions were carried out in dedicated thin-walled PCR microcentrifuge tubes in a thermocycler (VWR Quattrocycler) for DNA amplification. The cycle conditions (Denaturation, primer annealing, elongation) were adjusted specifically to primer set used and amplicon length. PCR reactions were performed in a total reaction volume of 20 µL containing 1 µL of gDNA (10 – 300 ng) or 2 µL of gDNA when extracted from single cell colonies, 2 µL of 10X ImmoBuffer, 0.6 µL of $MgCl_2$ (50 mM), 0.4 µL of forward and reverse primers (10 µM), 0.4 µL of dNTPs (10 µM), 0.2 µL of IMMOLASE DNA Polymerase, 4 µL of Q-Solution (5x) and 11 µL of sterile ultrapure water. The amplification of gDNA was carried out using the following conditions: Hot start incubation at 95°C for 10 mins to activate the IMMOLASE DNA Polymerase, followed by 35 cycles of denaturation (94°C, 30 s), primer annealing (temperature primer depended, 30 s) and elongation (72°C, 30 s). A final elongation at 72°C for 2 min completed the DNA amplification. PCR products (5 µL) were analysed on 2% agarose gel. Table 8 lists primer sequences used in Hot-Start PCR.

Table 8: PCR primer used to amplify genomic CRISPR site within *PMM2* in HepG2 cells

Name	**Target**	**Direction**	**Sequence 5'-3'**	**Amplicon length**
***PMM2* For3**	*PMM2*	Forward	CCGACCCCAAAGAGCATCAC	296 bp
***PMM2* Rev3**	*PMM2*	Reverse	CAAGACACGTTGGCACGAGGAAC	
ChrY For	*TTTY14*	Forward	CCTGAACCCTGTCCTCTTGG	420 bp
ChrY Rev	*TTTY14*	Reverse	GGACTTCCTAAGGGCTGTGT	

The annealing temperature of the *PMM2* For3/Rev3 primer pair was 56°C and the one for ChrY For/Rev 60°C. Primers were purchased from either Integrated DNA Technologies or Thermo Fisher Scientific.

2.4.1.4 Agarose gel electrophoresis

PCR and restriction digest products were analysed on 1-2% agarose gels in TAE (Tris-acetate-EDTA) buffer, depending on the size of the DNA products to analyse. Usually 5 µL per PCR sample were mixed with the 5X Orange G loading dye and brought to a total sample size of in total 19 µL. Restriction digests were complemented with 5X Orange G loading dye before loading. 100 bp or 1 kb molecular weight standards (NEB, see Figure 75, Figure 76) were run alongside each electrophoresis and DNA was visualised by adding 0.25 mg/mL ethidium bromide to the gel. Electrophoresis took place at 120-130 V for 35-45 minutes.

2.4.1.5 Gel purification of DNA

If it was needed, DNA bands were extracted and purified from gel using the GeneJET Gel Extraction Kit (ThermoFisher) according to the manufacturer's instructions. Cut bands were transferred into a 1.5 mL reaction tube and 100 µL/0.1 g gel of binding buffer were added. In the case of 2% gels the volume of binding buffer was doubled. Gel bands were put at 55-60°C on a heating block for 10-20 minutes to melt. Up to 800 µL of molten DNA gel bands were transferred onto the GeneJET purification column and spun in a centrifuge for 1 minute at 13,000 rpm and RT. The flow-through was discarded. This step was repeated if the total sample volume exceeded 800 µL. If low DNA amounts were expected an additional binding step was introduced by adding additional100 µL of binding buffer onto the column silica membrane and the sample was spun as previously described. Next, the bound DNA was washed with ethanol-containing wash buffer (700 µL) and again spun under same conditions. After discarding the flow-through the column was spun one more time to completely remove residual wash buffer. The DNA then was eluted in 15-50 µL sterile ultrapure water, depending on the DNA concentration expected. DNA concentrations were quantified using NanoDrop™ and stored at -20°C.

2.4.1.6 ExoSAP treatment of PCR products

For the enzymatic clean-up of amplified PCR products, prior to sequencing, samples were treated with a combination of exonuclease I and shrimp alkaline phosphatase (ExoSAP, Affymetrix). This allowed the removal of excess primer and nucleotides in the samples which potentially can disturb downstream reactions such as sequencing. Per PCR sample, 0.075 µL exonuclease I, 0.75 µL shrimp alkaline phosphatase, 10X SAP buffer and 0.675 µL dH_2O were added and. Usually, these components were combined in a master mix and 2.2 µL of it added to each PCR sample. Samples were incubated 40 minutes at 37°C, followed by 5 minutes at 80°C and stored at 4°C in a thermocycler.

2.4.1.7 Sanger Sequencing

Sequencing of PCR products or CRISPR constructs (Table 12) was outsourced and completed by Eurofins Genomics using the Mix2Seq Kit. Samples were prepared in provided barcoded tubes. Plasmids were sent in a total reaction volume of 17 µL containing 850 ng of plasmid in ultrapure water and 2 µL of the respective primer (10 µM) (Table 9). ExoSAP-treated PCR products were diluted 1:20 with ultrapure water and 15 µL of the diluted PCR product were mixed with 2 µL of a 10 µM primer (usually only forward primer) and sent for sequencing.

Table 9: Sequencing primer used for verification of *PMM2* ROI, CRISPR constructs and gene editing

Name	Target	Direction	Sequence 5'-3'	Template
***PMM2* For3**	*PMM2*	Forward	CCGACCCCAAAGAGCATCAC	PCR product
U6-F	U6 promoter	Reverse	GAGGGCCTATTTCCCATGATT	Plasmid

2.4.1.8 Analysing Sanger Sequencing

All Sequencing data were analysed in ABI format. Electropherograms showing sequencing traces were reviewed with the free available software SnapGene® Viewer (http://www.snapgene.com/). To confirm insertions and deletions (Indels) in heterozygous sequences after gene editing, the web-based tool TIDE (Tracking of Indels by DEcomposition, https://tide-calculator.nki.nl/) was used.

2.4.2 TIDE

TIDE (Tracking indels by Decomposition) is an easy to use and cost-effective online tool, which allows for quantification of gene editing events generated by CRISPR-Cas9, TALENs or ZFNs by determining sequence alterations and their frequency in a precise manner. The input material needed for a TIDE analysis were two simple Sanger sequencing traces obtained from the edited HepG2 or HEK293T cell pool or later single cell colonies and a wt control sample after PCR amplification of extracted gDNA. It aligns sample and control sequence and then calculates the percentage of the overall efficiency the individual gene editing tool has altered the targeted DNA with. In the case of the CRISPR-Cas9 system, this is significantly dependent on the gRNA target sequence chosen, as gRNA performance can vary tremendously. It also provides the individual indel spectrum showing the resulting predominant deletions and insertions as well as it models the aberrant sequence signal including the predicted cut site. [214].

2.4.3 Plasmid-based CRISPR gene editing methods

2.4.3.1 CRISPR gRNA design

Guide RNA's were designed using the molecular biology software SnapGene® (http://www.snapgene.com/) and two CRISPR design tools, CRISPOR (http://crispor.tefor.net/) [215] and CHOPCHOP v2 (http://chopchop.cbu.uib.no/index.php) [216]. gRNA target regions were chosen directly from the human *PMM2* promoter sequence region of interest to ensure closest proximity to the target site (c.-167), crucial for a CRISPR HDR gene edit. Mandatory for a *S. pyogenes* Cas9 driven approach, the gRNA target sequence has to be 20 nt in length and precedes a 5'-NGG PAM site. Potential gRNA target sequences were chosen and cross referenced with the two CRISPR design tools mentioned above for uniqueness and specificity. Only those gRNAs were used, which have provided a specificity score of above 50 in CRISPOR [217] and above 0.5 in CHOPCHOP. The gRNA sequences chosen to target the *PMM2* promoter are shown in Table 10.

Table 10: gRNA DNA target sequences specific to the CRISPR site within *PMM2* in HepG2 cells to clone into PX458.

Name	**Target**	**Vector**	**Direction**	**Sequence 5'-3'**	**PAM**
gRNA hPMM2*_1	*PMM2*	PX458	Forward	GCGTGATCTGCGTTGCACCC	TGG
gRNA hPMM2*_2	*PMM2*	PX458	Reverse	TTCCTGGACCGCAACTCCCA	GGG

*human PMM2

For cloning into the target vector PX458, gRNA oligonucleotides were designed according to the sequences shown in Table 11.

Table 11: gRNA oligonucleotide sequences for cloning into CRISPR plasmid PX458.
BbsI recognition sites are highlighted in blue.

Name	**Target**	**Direction**	**Sequence 5'-3'**
gRNA hPMM2_1	*PMM2*	Forward	CACCGCGTGATCTGCGTTGCACCC
gRNA hPMM2_1	*PMM2*	Reverse	AAACGGGTGCAACGCAGATCACGC
gRNA hPMM2_2	*PMM2*	Forward	CACCgTTCCTGGACCGCAACTCCCA
gRNA hPMM2_2	*PMM2*	Reverse	AAACTGGGAGTTGCGGTCCAGGAAc

In order to enable ligation into the target vector via *BbsI* restriction sites, 5'-overhangs (CACC-/AAAC-) had to be added to the oligonucleotide sequences. The Cas9 expression plasmid used is controlled by a U6 promoter, which needs a guanine (G) at the transcription start site at the 5' end. A "G" was added to the 5' end of gRNA oligonucleotides when required.

2.4.3.2 Annealing of gRNA oligonucleotides

gRNA oligonucleotides (Table 11) were suspended in ultrapure water and brought to a stock concentration of 100 µM (100 pmol/µl). For annealing, 2 µL of 100 µM forward and reverse primer stock was mixed with 16 µL of TE buffer and the reaction was incubated at 100°C for 5 minutes on a heating block. After a short spin, the reaction tubes were placed back onto the heating block to gradually cool down to RT.

2.4.3.3 Cloning of CRISPR-Cas9 expression constructs

Each gRNA oligo duplex (Table 10) was cloned into the Cas9 expression vector pX458. A Golden Gate inspired approach was used for the cloning procedure, which allowed restriction and ligation in one reaction step [218]. In a final reaction volume of 20 µL, 1 µL of plasmid DNA PX458 (150 ng), 1 µL of annealed gRNA oligo pairs, 1 µL *BbsI*-HF (20 U), 1 µL T4 DNA ligase (400 U) and 10x T4 Ligase buffer were combined. The reaction was incubated in a thermocycler under the following conditions: A sequential incubation at enzyme-specific temperatures (37°C, 5 minutes for *BbsI*-HF and 16°C, 10 minutes for T4 DNA Ligase) was repeated during 10 cycles, followed by two final 5 min incubations at 50°C and 80°C. The reaction was cooled down to 4°C and then used for transformation into *E. coli* DH5α. Both CRISPR constructs had a total size of about 10 kb.

2.4.3.4 Bacterial transformation culturing of positive E. coli clones

Plasmid DNA (Table 12) was transformed into *E. coli* DH5α via heat shock. 5 µL of each CRISPR-Cas9 expression construct were mixed with 100 µL of thawed *E. coli* DH5α and incubated on ice for 20 minutes, following a heat shock at 42°C for 90 seconds and put back on ice for additional 5 minutes. Transformed bacteria were then plated directly on LB agar plates containing 100 µg/mL ampicillin for selection of positive clones ON at 37°C. 1-3 colonies per transformation were picked, inoculated into 2 mL of ampicillin containing LB medium and shaken ON at 37°C for harvest the next day.

2.4.3.5 Plasmid DNA preparation

Plasmid DNA was extracted and purified according to the GeneJET Plasmid Miniprep Kit (ThermoFisher). Transformed *E. coli* DH5α ON cultures were transferred into 1.5 mL reaction tubes and spun at 8,000 rpm for 2 minutes at RT. Pelleted bacteria were then resuspended in 250 µL of RNase A containing chilled resuspension solution and 250 µL of lysis solution were added and the mixtures was thoroughly mixed by inverting the tube 4-6 times resulting in a clear, viscous solution. To separate the vector DNA from gDNA, proteins and cell debris, 350 µL of the neutralisation buffer were added per miniprep and tubes, again, were inverted 4-6 times prior to a 5-minute centrifugation at 13,000 rpm. The plasmid DNA containing supernatant was then added onto the GeneJET Spin columns to harvest plasmid DNA by spinning for 1 minute at 13,000 rpm. The flow-through was discarded and the columns were washed twice with 500 µL of ethanol-containing wash solution, which was removed by centrifugation under conditions mentioned. The columns were placed in new, sterile 1.5 mL collection tubes and plasmid DNA was eluted in 40 µL of sterile, ultrapure water. DNA concentrations were measured using NanoDrop™ and samples were stored at -20°C.

Table 12: *PMM2* target-specific CRISPR constructs generated with pSpCas9(BB)-2A-GFP (pX458)

	Name	**gRNA Sequence 5'-3'**	**PAM**
Construct 1	pX458_hPMM2_1	GCGTGATCTGCGTTGCACCC	TGG
Construct 2	pX458_hPMM2_2	TTCCTGGACCGCAACTCCCA	GGG

2.4.3.6 Plasmid-Safe treatment

Purified CRISPR constructs (Table 12) were treated with an ATP-dependent Plasmid-Safe™ Nuclease (Epicentre) to remove contaminating chromosomal bacterial DNA. The entire cloning reaction was incubated with 1 µL of 10 mM ATP (Epicentre) and 1 µL Plasmid Safe nuclease (10 U) for one hour at 37°C in a thermocycler.

2.4.3.7 Verification of CRISPR-Cas9 expression constructs

Purified plasmids (Table 12) were verified by restriction double digest with *EcoRV* and *BbsI*-HF, both purchased form New England Biolabs, and verification was confirmed by Sanger sequencing.

Double restriction digest helped to distinguish between successfully cloned CRISPR construct and failed cloning due to the loss of the *BbsI* site when the gRNA target sequence was inserted into the CRISPR plasmid pX458. Restriction digests were performed in a final reaction volume of 10 µL. 1 µg of plasmid DNA were digested with 5 units of both *EcoRV* and *BbsI*-HF in 10X CutSmart® buffer and sterile ultrapure water. The results were analysed on a 1% agarose gel alongside a 1 kb DNA ladder (Figure 76) Successfully digested CRISPR constructs were sent for sequence verification using the U6-F forward primer (Table 9) as described in section 2.4.1.7.

2.4.4 Single-stranded donor oligonucleotide design for CRISPR-HDR

To enable a directed HDR-repair after CRISPR, the homologous ssODN which harboured the desired SNP at position c.-167 was synthesised by Eurogentec and co-delivered to the cells. The ssODN was designed according to the leading strand of the target region within the *PMM2* promoter and has a total length of 139 nt with the desired SNP (c.-167G>T) in the centre, flanked by 69 nt long homology arms (Table 13).

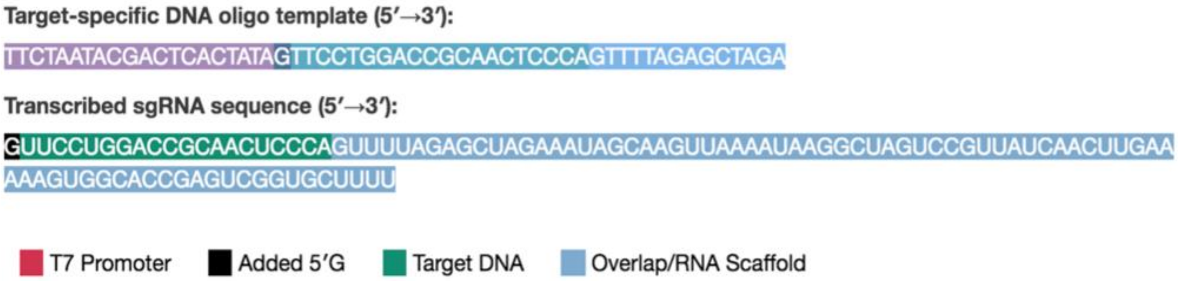
Table 13: Sequence of ssODN used for CRISPR/Cas9 co-transfection and electroporation. Desired SNP highlighted in red.

Target	**Direction**	**ssODN Sequence 5’-3’**
PMM2	Forward	ACCGGCCCCCGGAACCCCTAAGAGCGCATGCTCGAATGTACAAGGCGGGCGTGATCTGCGTTGCACCCTTGGAGTTGCGGTCCAGGAATCGTGGCTGCCGCTACTCCCAGGCGTTATGGGAACGGAGTCCCCTCCTCTT

2.4.5 RNP-based CRISPR gene editing methods

2.4.5.1 sgRNA DNA oligo design

In order to work with CRISPR RNPs, a DNA oligo template (Figure 25) for sgRNA synthesis of gRNA hPMM2_2 (s. Table 10) was designed using the EnGen™ sgRNA Template Oligo Designer according to the target sequence:
5’-TTCCTGGACCGCAACTCCCA-3’.



Target-specific DNA oligo template (5′→3′):

TTCTAATACGACTCACTATAGTTCCTGGACCGCAACTCCCAGTTTTAGAGCTAGA

Transcribed sgRNA sequence (5′→3′):

GUUCCUGGACCGCAACUCCCAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAA
AAAGUGGCACCGAGUCGGUGCUUUU

🟥 T7 Promoter ■ Added 5′G 🟩 Target DNA 🟦 Overlap/RNA Scaffold

Figure 25: Sequence and transcribed sequence of target-specific DNA oligo template for sgRNA hPMM2_2 synthesis.

2.4.5.2 sgRNA synthesis and purification

sgRNA hPMM2_2 was synthesised using the EnGen™ sgRNA synthesis kit (NEB) under RNAse-free conditions according to the manufacturer’s protocol. For synthesis in a 20 µL reaction, the target-specific DNA oligo template 1 µM (5 µL) was combined with nuclease-free water (3 µL), EnGen™ 2x sgRNA Reaction mix (10 µL) and the EnGen™ sgRNA enzyme mix in a PCR reaction tube. The reaction was mixed well, incubated for 2.5 hours at 37°C in a thermocycler and put on ice afterwards. The reaction was brought to 50 µL with nuclease-free water for DNAse I treatment and 2 µL of DNAse I were added (15 minutes, 37°C). After incubation, the reaction was either stored at -20°C or sgRNA was purified straight away.

sgRNA was purified by adding 500 µL phenol-containing TRIzol® reagent (ThermoFisher) to the tube. After mixing, 100 µL of chloroform were added, the mixture inverted several times and spun for 15 minutes at 13,500 rpm to promote phase separation. The RNA-containing aqueous phase was collected carefully in a new 1.5 mL Eppendorf tube and the equal volume of isopropanol was added (pipette up and down). The mixture was then incubated either at -20°C ON or at -80°C for 1 hour to commence precipitation of sgRNA. Following centrifugation at 13,500 for 10 minutes, the isopropanol was discarded and the pelleted sgRNA washed with 70% ethanol and again spun as previously. The ethanol was discarded under sterile conditions (cell culture hood) and the pellet dried for about 20 minutes on ice. The sgRNA was mixed with 30 µL of water (water for embryo transfer, Sigma) and dissolved for 10 minutes on ice. A few µL were taken to a new tube to measure sgRNA concentration with the NanoDrop™ spectrophotometer. The quality of sgRNA was determined by running 1 µg on a 2% agarose gel. sgRNA was diluted to obtain a concentration of 680 ng/µL (20 µM) and stored at -80°C.

2.4.6 Assessment of cutting efficiency after CRISPR and CRISPR-HDR transfection and electroporation

To assess the overall success of each CRISPR experiment in HEK293 and HepG2 cells, about 30% of an edited cell population was lysed and gDNA was extracted as described previously (s. 2.4.1.2) and used as a template for PCR with *PMM2* specific primers (s. 2.4.1.3, Table 8). The approximately 300 bp amplicon was then sent for sequencing to detect whether Indels due to CRISPR/Cas9 activity are visible or not in the resulting electropherogram. In addition, sequences were analysed in TIDE (s. 2.4.1.8) and differences, compared to a control wildtype sequence, quantified. Experiments with an overall CRISPR efficiency below 20% were not considered being successful and have been rejected for further use. CRISPRed cells of a successful experiment were divided in two parts after reaching a decent confluency. One part was used to seed cells in low density to obtain single cell colonies and the rest served as a backup and was kept in culture and later on put in long term storage at -80°C.

2.4.7 Obtaining single-cell colonies from CRISPR-HDR targeted cell populations

CRISPR-HDR targeted HepG2 cells with an overall sufficient editing efficiency were seeded in low density in 10 cm tissue culture dishes to obtain single-cell colonies for genotyping. To guarantee a well separated single cell suspension, cells were split as described previously (s. Culture of HepG2 cells) using a 21G needle and seeded in 10 mL of HepG2 cell culture medium. Although the numbers of cells seeded per 10 cm tissue culture dish varied, a well scattered population was always controlled by microscopy. Low density plates were incubated at 37°C and 5% CO_2 with regular medium changes every 4-5 days. After 10-14 days, single cell colonies were picked using a microscope (stereomicroscope with transillumination) under sterile conditions and each clone was transferred to a separate well of a 96 well plate containing 200 µL of HepG2 cell culture medium for further growth at 37°C and 5% CO_2 for 7-9 days.

2.4.8 Genotyping of HepG2 single cell colonies after CRISPR-HDR

In order to genotype CRISPR-HDR single cell clones, those were first split asymmetrically by transferring 20-30% onto a new 96 well plate. These plates were maintained in culture to have a backup of each clone. The remainder population of each clone on the “old” 96 well plate was used for genotyping after a few more, usually 2-3 days, in culture. When the single clones in each well reached a confluency of at least 70%, cells were lysed in 100 µL of lysis buffer ON and gDNA was extracted directly on the 96 well plate to simplify the process. Lysis and gDNA extraction conditions used previously (s. 2.4.1.2) were adjusted to the dimensions of a 96 well plate as follows: 10 µL (1/10th) of the 8 M LiCl solution and 100 µL of isopropanol (molecular grade) were added subsequently to each lysate. The plate was then mixed for at least 10 minutes on a microplate shaker at 700 rpm and RT. The DNA was then precipitated by spinning for 20 minutes at 2,500 rpm in a centrifuge. The supernatant was removed by inverting the plate over a basin and blotting it onto a dry tissue. DNA was washed with 200 µL per well of a 70% ethanol solution and the plate again spun for 20 minutes at 2,500 rpm. The ethanol was removed as previously described and the plates then put at 37°C for 15-30 minutes to assure complete evaporation of the alcohol. The DNA per well was eluted in 30 µL of sterile ultrapure water and dissolved whilst at least one hour of shaking at 250 rpm and 37°C or during an ON incubation at 4°C prior to usage in downstream reactions. To genotype each clone, 1 µL of each gDNA extract were put in a genotyping PCR as a template and amplification was performed according to the protocol described in section 2.4.1.3 using the *PMM2* For3/Rev3 primer pair (Table 8). As a quick and easy first check whether CRISPR-HDR repaired clones are present, all amplicons were digested with *StyI* (also called Eco130I). The desired G>T SNP at position c.-167 within the *PMM2* promoter coincidentally creates an additional restriction site for *StyI* (ThermoFisher), which is usable to distinguish between HDR and non-HDR clones as cutting results in different band patterns within the region of interest Figure 26.

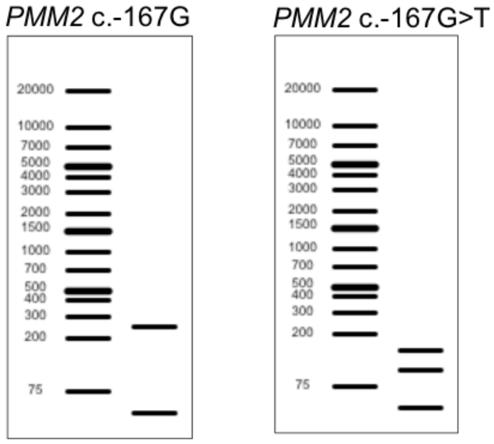


Figure 26: Expected band pattern after digesting CRISPR-HDR targeted single clone amplicons from genotyping PCR with *StyI*.
Left: Bands (250 bp, 50 bp) show pattern as expected when the desired SNP is not introduced at nucleotide position c.-167. Right: Bands (150 bp, 100 bp, 50 bp) show expected band pattern when desired SNP (c.-167G>T) is introduced via CRISPR-HDR.

Each restriction digest reaction was done in a total volume of 20 µL. 5 µL of each PCR product were combined with 15 µL of a previously prepared master mix composed of 2 µL of the enzyme-specific 10X Orange buffer (ThermoFisher), 0.2 µL of *StyI* (2 units) and 12.8 µL of sterile ultrapure water per reaction. Volumes for the master mix were multiplied according to the number of samples digested. The restriction digests were incubated for at least one hour at 37°C and analysed on a 2% agarose gel (2.4.1.4).
Positive clones were further analysed using Sanger sequencing (s. section 2.4.1.7) to confirm the desired gene edit.

2.4.9 Expansion of positive CRISPR-HDR HepG2 single cell clones

Positive clones were expanded from the backup 96 well plates kept in culture after asymmetric splitting. Clones were maintained in HepG2 culture medium and when due to split transferred onto 24 wells. When ready to split, one half of cells was used to extract gDNA for characterisation to ensure the clones taken are the desired ones and the other half was transferred onto a 6-well plate for further expansion. From there, one part of each positive clone was transferred into a T25 flasks for maintenance and growth and the other part was put in cryopreservation at -80°C. From this point, clones were further expanded and kept in culture and several backups were deep-frozen at -80°C to generate an adequate stock of each clone for future experiments.

2.4.10 Checking for off-target effects

In order to investigate whether the gRNA used has created any off-target effects in different regions within the HepG2 genome was tested with PCR (2.4.1.3) and subsequent Sanger sequencing analysis. Those off-targets being predicted most likely by CRISPOR and CHOPCHOP were considered and primers were designed accordingly, spanning a region of 400-500 bp and the PCR products sent for sequencing after the clean-up with the ExoSAP Kit.

Chapter 3: Project background

3.1 Introduction

This chapter shall summarise the beginning of this project and the work included before this PhD was started. This mainly involves clinical observations characterising the phenotype of the patients, genetic studies, some functional assays and bioinformatic analysis, which eventually helped to describe HIPKD as a new polycystic kidney disease with involvement of the pancreas and the liver. This work was done by our group at the Centre for Nephrology in London and in collaboration with colleagues working in universities and hospitals in Spain, Germany, UK, US and Qatar and published in The Journal of the American Society of Nephrology (JASN) in April 2017 [209].

3.2 The aims of this chapter

- Presentation of the clinical phenotype and the patients
- Identification of the mutation
- Functional studies of the disease variant
- Genomic region from bioinformatics point of view
- Hypothesis for disease mechanism

3.3 The patients and the clinical phenotype

The project was born when patients with an ARPKD-like phenotype in combination with an hyperinsulinemic phenotype of unknown origin were seen in clinics. Although, a substantial number of patients underwent genetic testing focussing on mutations in causative gene(s) for ARPKD (*PKHD1*) and hyperinsulinemic hypoglycemia (*ABCC8*, *KCNJ11*), no pathogenic mutations were identified, and the genetic cause of the phenotype remained unclear at that point.

To date, we know 21 patients from 13 unrelated families with Caucasian background showing the phenotypic combination of PKD and HI (HIPKD). Four of them are offspring of a consanguineous family. From the pedigrees of five of the affected families we could conclude an autosomal recessive inheritance of the disease as shown in Figure 27 and which was compatible with the inheritance in the remaining families (data not shown).

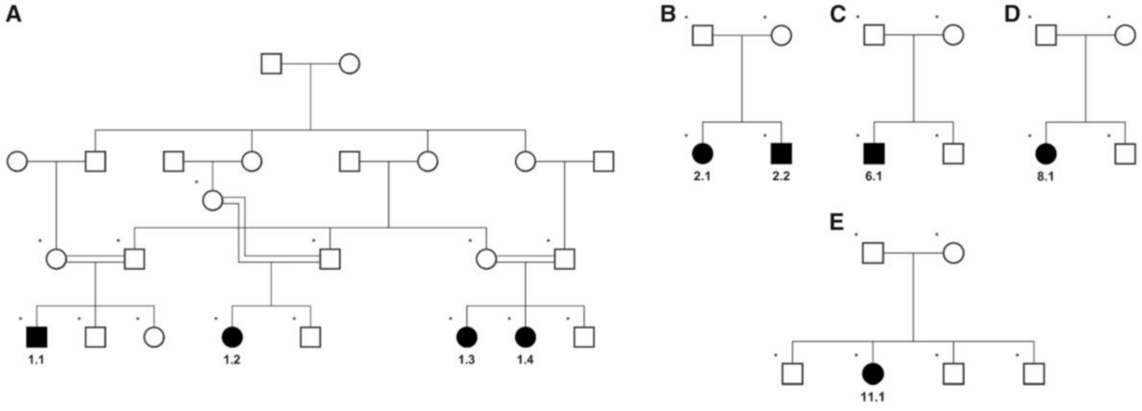


Figure 27: Pedigrees of five informative families included in the study showing the origin of nine patients with HIPKD.
Dark symbols indicate affected individuals. Empty symbols indicate unaffected individuals. Circles: Female. Squares: Male. A small dot above the left-hand corner of the symbols indicates confirmed carrier status of the promoter mutation in PMM2 c.-167G>T.

The physical manifestations of the disease in all patients were restricted to the kidneys and the pancreas and in some patients also in the liver. In all patients, enlarged and cystic kidneys were clearly identified (Figure 28 A–G). The morphologic appearances of the kidneys typically prompted an initial clinical diagnosis of ARPKD before the genetic testing was done. Histology was available in one patient only and showed cysts predominantly or perhaps even exclusively of glomerular nature consistent with a glomerulocystic disorder (Figure 28 H and I). This excluded a diagnosis of ARPKD, as cysts in this disorder exclusively localise to the collecting duct. Since histology is only available in one patient it is currently unclear whether all patients have glomerular cysts, or whether this was unique to this particular patient.

In addition to the renal malformations an excess secretion of insulin was noted in all patients. In eight patients, small liver cysts were also seen. Of note, one patient without detectable liver cysts on imaging had a liver biopsy performed. This was done at the time to investigate the suspected diagnosis of ARPKD, which has obligated liver involvement in the form of ductal plate malformation, a developmental abnormality of the intrahepatic bile ducts. Interestingly, the biopsy indeed showed ductal plate malformation. Since this biopsy was done in a patient without apparent liver abnormalities on routine liver imaging, it raises the question whether ductal plate malformation is also an obligate manifestation of HIPKD, or whether this patient was an exception. Figure 28 and Figure 29 represent the appearance of kidneys and liver in HIPKD on MRI, ultrasound and histology images.

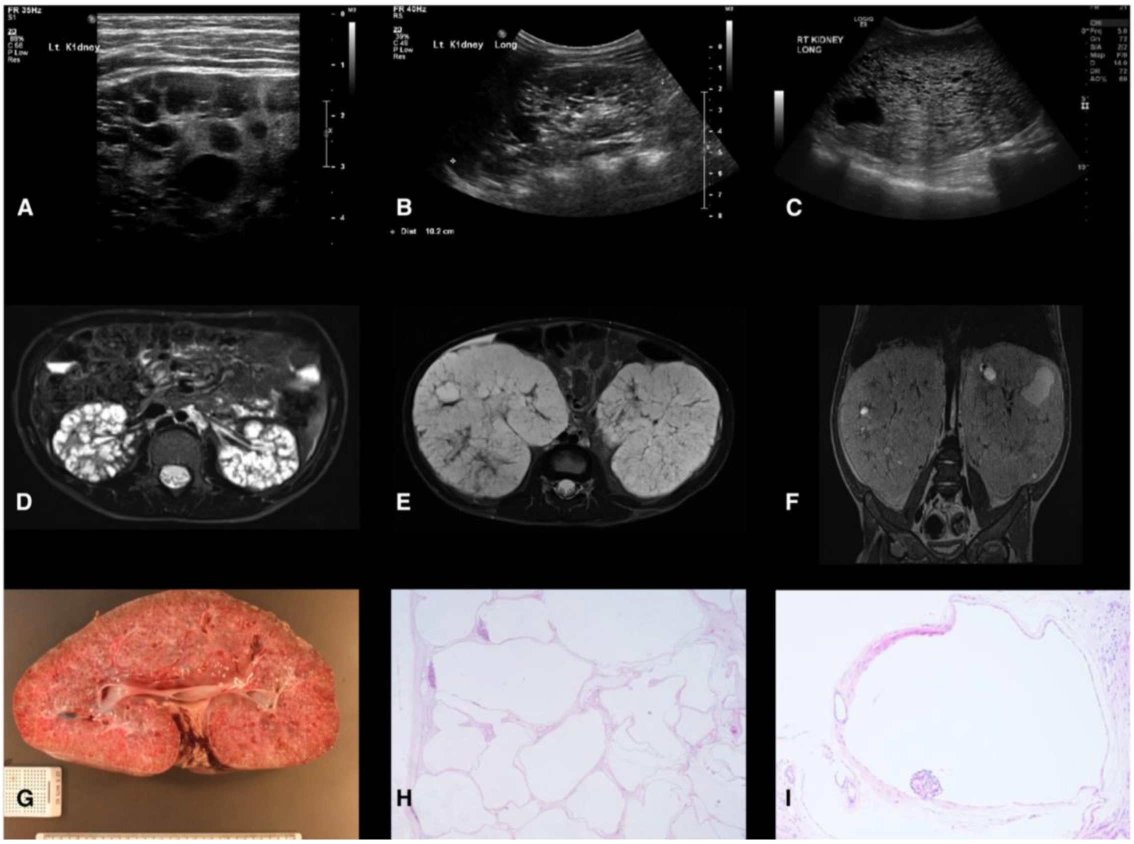


Figure 28: Renal phenotype in patients with HIPKD.
A-C: Ultrasound of three cystic kidneys. A: Left kidney, patient 11 years old. B: Right kidney, patient 4 years old. C: Right kidney, patient 2 years old. D-E: MRIs of two cystic kidneys. D: Axial MRI, patient 11 years old. E-F: Axial and coronal MRIs (without contrast), patient 2 years old. G: Nephrectomy specimen of enlarged, cystic HIPKD kidney, patient 2 years old. H and I: Histology (hematoxylin and eosin staining) of the same kidney showing multiple cysts lined with a thin epithelium and some enclosing the glomeruli.

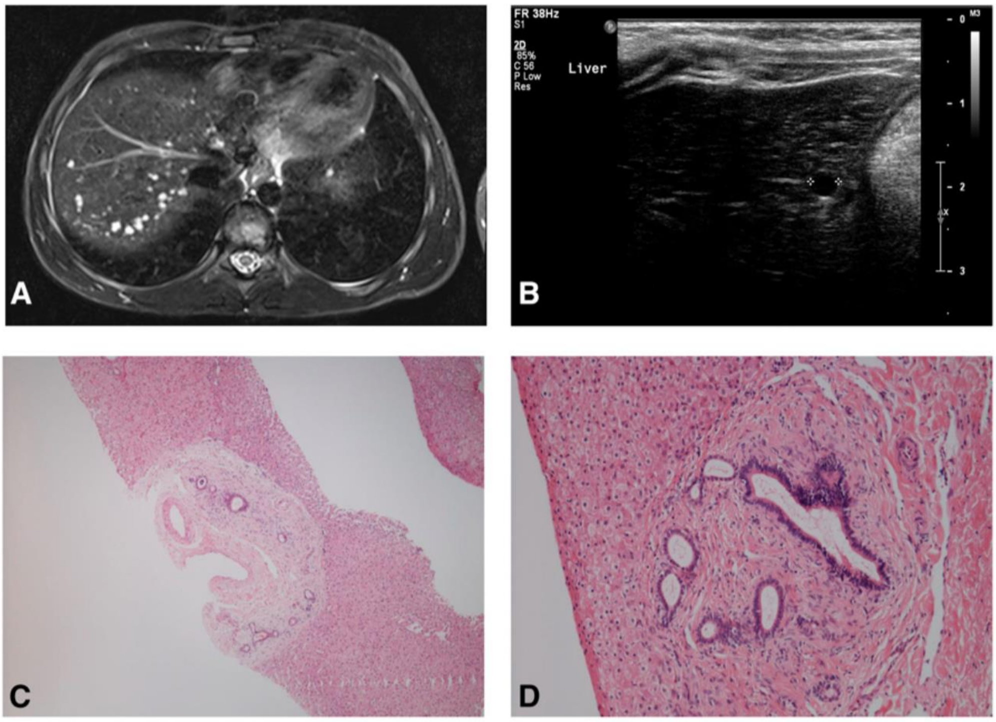
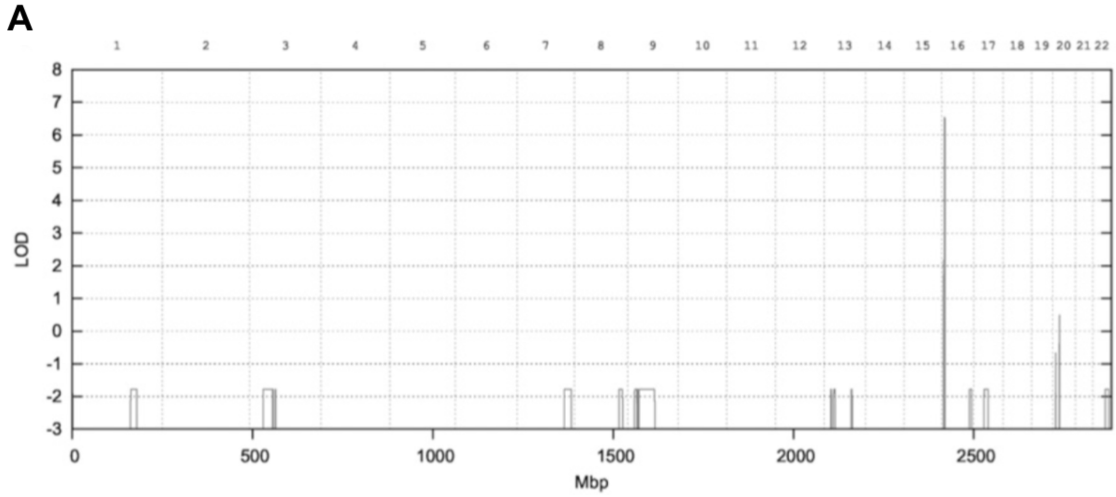


Figure 29: Hepatic phenotype in HIPKD.
Note that liver abnormalities are not expressed in all patients. A: Axial MRI of HIPKD liver with multiple small cysts, patient 11 years old. B: Ultrasound of HIPKD liver showing a cyst, 4 years old. C-D: Histology (hematoxylin and eosin staining) showing ductal plate malformation, patient 1 year old. No liver cysts were noted in this patient.

The children often presented hypoglycemic seizures as a consequence of hyperinsulinism, which was diagnosed at an age of 10 months on average. There was no evidence for the involvement of other organs.

3.4 Identification of the disease-causing mutation and description of the locus

An elaborate genetic analysis was done in order to find the HIPKD-causing mutation and the results are shown in Figure 30. To be able to identify the precise locus of the mutation within the genomes included, a whole-genome multipoint parametric linkage analysis was performed. The analysis was done on the basis of the five pedigrees shown in Figure 27 detecting a distinct single locus on chromosome 16p.13.2 at 2.3 Mb with a combined LOD score of 6.5. With next-generation sequencing we finally identified a point mutation within the promoter region of *PMM2*, encoding the phosphomannomutase 2. The mutation is a guanine to thymine single nucleotide polymorphism (SNP) located 167 nucleotides upstream (c.-167G>T) of the start codon, which has not been reported in any database yet. The point mutation was shared by all 17 patients as confirmed with Sanger sequencing (Figure 30 B), either in homozygous state, as it was the case for the four patients descending from the consanguineous family, or in compound heterozygous state. The latter was found in all other 13 patients with *PMM2* coding mutations on one allele, known to be causative for the severe multisystem disorder CDG-1a described previously 1.8.1.



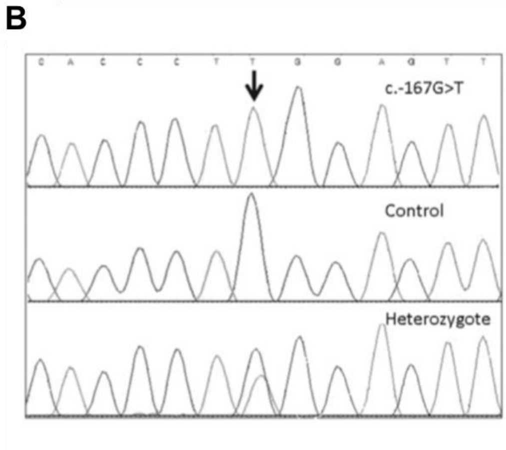


Figure 30: Genetic studies identifying the HIPKD causing mutation.
A: Parametric multipoint linkage analysis. Single peak on chromosome 16 with combined LOD score of 6.5. B: Sequencing result of the linked region showing the *PMM2* promoter mutation c.-167G>T. Depicted are Sanger sequencing traces from position c.-173_-161. Top: Homozygous patient. Middle wt control. Bottom: Heterozygous parent.

Knowing the mutation lies within the non-regulatory part of *PMM2*, we were able to investigate this further. The literature provided us with some insightful information around the *PMM2* promoter. A major characteristic of this promoter is its bidirectionality making it responsible for the regulation of two genes, *PMM2* and *TMEM186.* To drive its regulatory function the *PMM2* promoter hosts specific binding sites (SBS) were transcription factors bind to activate or block the transcription of the genes. In this context, Anno et al. have described the existence and functional importance of three SBSs (1-3) within the *PMM2* promoter and have shown reduced expression levels of both genes when any of the SBS was deleted. They also described ZNF143, a ubiquitously expressed protein, as being an essential transcription factor for bidirectional transcription, having strong affinity to bind at SBS1 within the *PMM2* promoter [219]. Our group analysed the SBS independently and localised the c.-167G>T promoter SNP within SBS 1, which contains a binding motif for ZNF143.

3.5 Functional studies of the *PMM2* promoter variant

We were able to confirm ZNF143 binding to SBS1 within the *PMM2* promoter and most importantly we could show significantly reduced binding of ZNF143 to its binding motif when the *PMM2* promoter mutation was present. This was confirmed by a densitometry analysis of the two bands appearing in the immunoblot representing ZNF143 binding. The intensity of the band resulting from ZNF143 binding to the mutant promoter was 70% reduced compared to the band from ZNF143 interaction with the wt promoter. The results are shown in Figure 31.

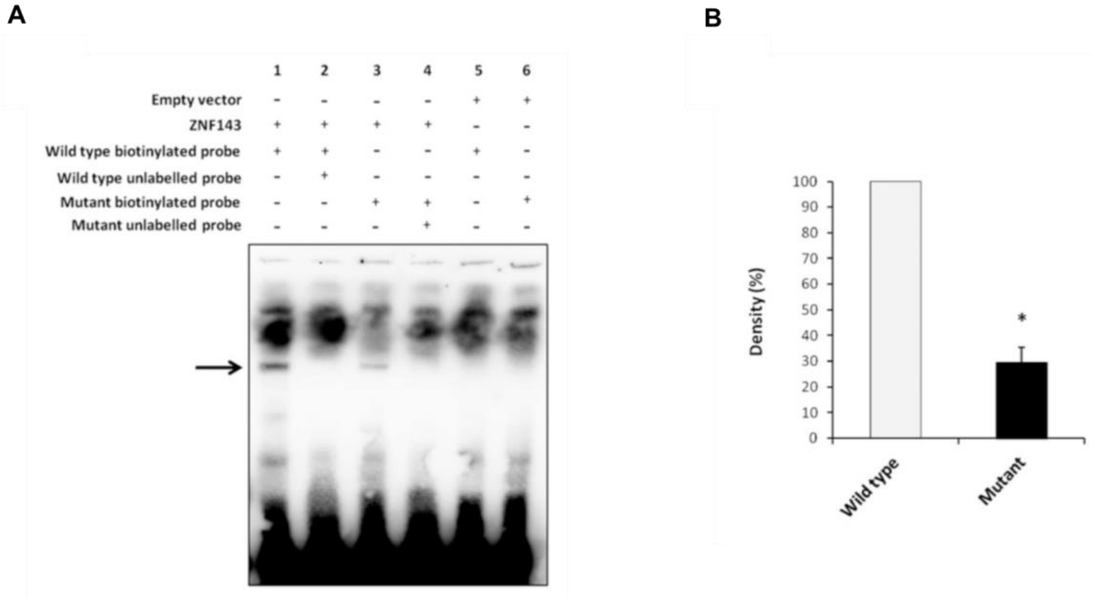


Figure 31: EMSA assay demonstrating ZNF143 binding to mutant and wt *PMM2* promoter. Black arrow indicates ZNF143 binding. Lane 1: ZNF143 binding to the wt *PMM2* promoter (labelled probe). Lane 2: Competition with excess of wt unlabelled probe. Lane 3: ZNF143 binding to the mutant *PMM2* promoter (labelled probe). Lane 4: Competition with excess of mutant unlabelled probe. Lane 5: Negative control without ZNF143, with empty vector and wt labelled probe. Lane 6: Negative control without ZNF143, empty vector and mutant labelled probe. B: Densitometry analysis comparing band intensities (Lane 1 vs. Lane 3) in percent.

ZNF143 and *PMM2* promoter interaction was assessed with an electrophoretic mobility shift assay (EMSA). ZNF143 was generated *in vitro* in a cell free system (TNT T7 Quick Coupled Transcription/Translation System, Promega) with an expression vector and two biotinylated oligonucleotide probes containing either the wt or mutant *PMM2* promoter sequence c.150_-178 were used to provide the ZNF143 binding site for the assay.

To address the expression levels of *PMM2* in HIPKD patients, we expected these to be reduced due to the promoter mutation and have tested this experimentally *in vitro* performing a luciferase assay using two commercially available cell lines; the human renal proximal tubular cell line RPTEC/TERT1 (LGC standards, ATCC-CRL- 4031, UK) and the human pancreatic β-cell line 1.1B4 (Sigma Aldrich/ECACC, 10012801-1VL, UK). The Renilla luciferase gene on a promoter reporter construct was under control of either wt *PMM2* promoter, mutant *PMM2* promoter (c.-167G>T) or SBS1-deleted *PMM2* promoter (c.-163_-180del). The targeted cell lines were transfected with these constructs. The results are shown in Figure 32.

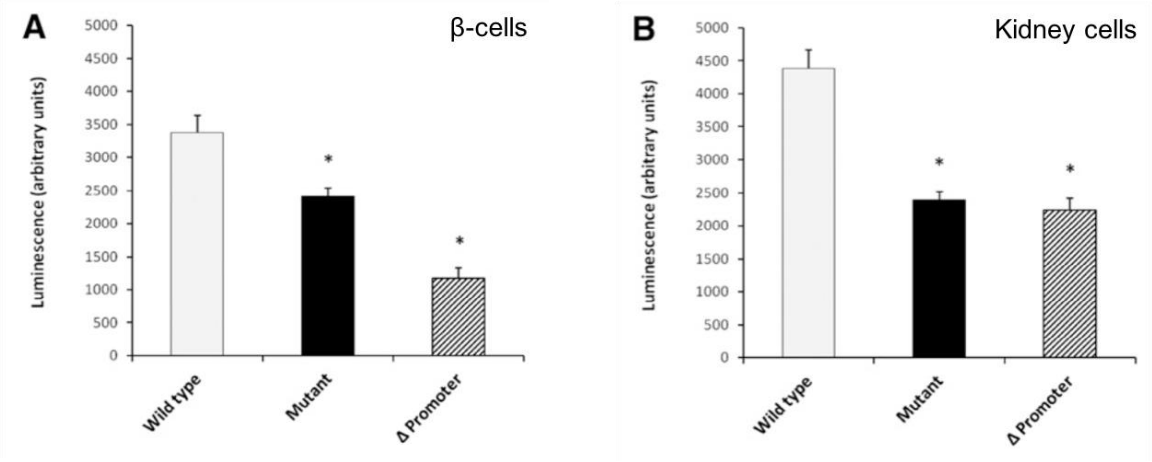


Figure 32: Luciferase assays showing luciferase expression controlled by the *PMM2* promoter. Light grey bars: Luciferase activity with wt promoter (c.-167G). Black bars: Luciferase activity with mutant promoter (c.-167G>T). Crosshatched bars: Luciferase activity with ZNF143 binding site deleted in promoter (c.-163_-180del). A: Pancreatic human β-cell line. B: Human kidney cell line. Asterisk indicates statistical significance ($P<0.01$) of wt compared to mutant/ΔPromoter. Luminescence in arbitrary units.

In both cell lines the luciferase expression was significantly reduced when the gene was regulated by the mutant and SBS1-deleted *PMM2* promoter. The latter turned down luciferase expression even more in β-cells whereas expression with mutant promoters in kidney cells remained the same.

3.6 *PMM2* and Congenital disorder of glycosylation

The finding of mutations in *PMM2* as the cause of HIPKD was surprising, as *PMM2* encodes a key enzyme in N-glycosylation and is a recognised disease gene for CDG-1a, a severe multisystem disorder caused by biallelic *PMM2* coding mutations as described in the introduction. In contrast, our patients appeared to have isolated involvement of kidneys, pancreatic beta cells and liver. The standard clinical test to assess for CDG-1a is isoelectric focusing of transferrin. Transferrin is an iron-binding blood plasma protein predominantly produced in the liver and a major distributor of iron to all tissues in the body [220]. It is normally strongly N-glycosylated and defects in glycosylation thus result in a shift of the isoelectric point, which is detectable [221]. Interestingly, this test was normal in patients with HIPKD (Figure 33 C) arguing against a systemic defect in N-glycosylation. The transferrin N-glycosylation pattern of three HIPKD patients tested did not differ to the normal control Figure 33 C.

We therefore hypothesised that the promoter mutation in *PMM2* causes an organ-specific defect in glycosylation in kidneys, pancreatic β-cells and liver only, in contrast to biallelic coding mutations in *PMM2* which cause a generalised defect in glycosylation, affecting all organ systems. This would be consistent with the concept of pleiotropy: that mutations in one gene can cause two or more different disorders as is the case for example for the *OCRL* gene [222]. To prove this hypothesis, we aimed to assess glycosylation status of affected tissue. The only available patient material from affected organ systems were cells grown in primary culture and derived from the end-stage kidney of a patient who had nephrectomies performed at the time of transplant. We assessed the N-glycosylation status of TRPP2 immunoprecipitates from patient cells used in a deglycosylation assay with PNGase F and Endo H by western blot. TRPP2 (transient receptor potential channel 2, *PKD2* gene) is a non-selective cation channel in the kidney which is heavily N-glycosylated. Mutations in TRPP2 are found in ARPKD, some of which can affect aminoacids that represent N-glycosylation sites (Asn-X-Ser/Thr; X can be any amino acid but prolin), consistent with the idea that abnormal TRPP2 glycosylation can cause cystic kidney disease [223]. The results of these assays are shown in Figure 33 A and B.

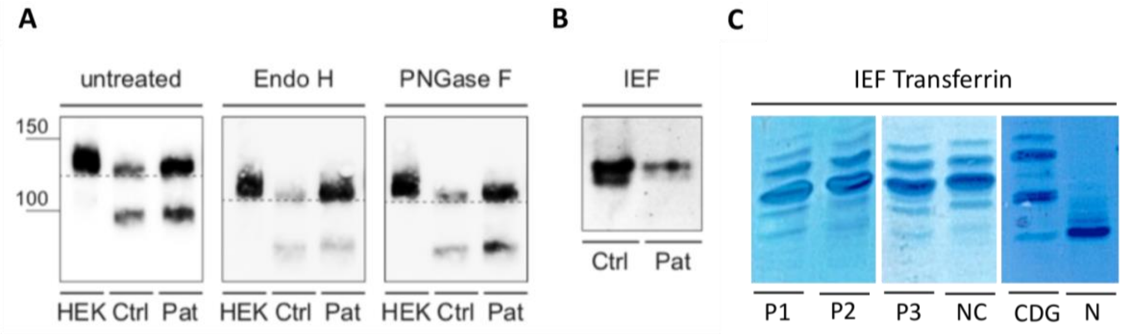


Figure 33: Glycosylation levels of TRPP2 and serum transferrin in HIPKD patient primary cells. A: Enzymatic deglycosylation of endogenous TRPP2 immunoprecipitates with either Endo H or PNGase F from HIPKD patient cells compared to control primary kidney cells and HEK cells on western blot detected with an anti-TRPP2 antibody. HEK: HEK293T transfected with TRPP2. Ctrl: Endogenous expressed TRPP2 in control primary kidney cells. Pat: Endogenous expressed TRPP2 in HIPKD patient primary kidney cells. 150: 150 kDa. 100: 100 kDa. Left panel: Western blot of non-glycosylated TRPP2. Middle panel: Western blot of TRPP2 treated with Endo H. Right panel: Western blot of TRPP2 treated with PNGase F. B: Isoelectric focusing (IEF) of TRRP2 in control primary kidney cells (Ctrl) and patient primary kidney cells (Pat). C: IEF of serum transferrin of three patients compared to three controls. P1-P3: Patients 1-3. NC: Normal wt control. CDG: Congenital disorder of glycosylation. N: Deglycosylated Control (PNGase F & EndoH).

Differently to heterologous expressed TRPP2, endogenous TRPP2 appears with two distinct bands on western blot as was observed before [223]. The bands of immunoprecipitated, normally glycosylated TRPP2 for patient and control sample run at around 120 kDa and just below 100 kDa consistent with the results of Hofherr et al. As it was expected, immunoprecipitates after deglycosylation with Endo H and PNGase F run lower in the blot at around 105 and 80 kDa due to the removal of N-glycans (Figure 33 A). However, in both cases no shift between patient and control samples was detectable and this confirmed an overall glycosylation defect being absent in HIPKD. This was also confirmed by isoelectric focussing of TRRP2 in HIPKD patient compared to control cells with no difference in the band pattern being detected (Figure 33 B). This result was inconsistent with our hypothesis of organ-specific dysfunction of glycosylation. Our current best explanation is that the cells which were obtained from an end-stage kidney do not represent healthy kidney tissue but mostly fibrotic scar tissue and therefore are not representative. Further investigations are needed to conclusively prove or disprove our hypothesis.

3.7 Impact of impaired glycosylation on insulin secretion

Knowing PMM2 plays a major in N-glycosylation, it was expected that impaired *PMM2* expression has an effect on insulin secretion based on a dysfunctional performance of PMM2 in HIPKD patients. As an excess secretion of insulin has been observed in affected children, it was of interest to investigate whether deglycosylation enhances insulin secretion. This was tested in a glucose-stimulated insulin secretion assay using the murine β-cell line MIN6. The cells were treated with two glycosidases, Endo H and PNGase F, beforehand to provoke deglycosylation of extracellularly accessible proteins. This was meant to mimic PMM2 dysfunction.

Glycosidase-treated MIN6 cells were compared to a control group not treated with any of the enzymes. Cells were split in three groups and after enzyme-incubation either stimulated with very basal glucose levels (2 mM), or high glucose levels (20 mM) or high glucose (20 mM) in combination with the PKC-enhancer PMA. Secreted insulin was assessed using a human insulin ELISA. Figure 34 shows the insulin measurement.

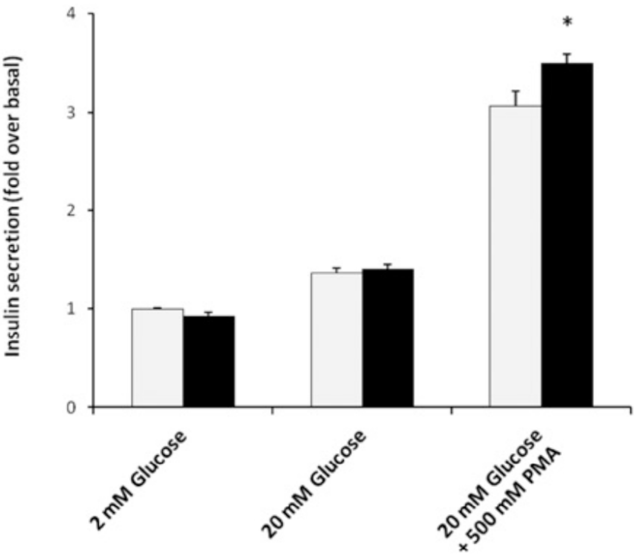


Figure 34: Insulin secretion of deglycosylated MIN6 β-cells. Stimulation with different glucose concentrations (2 mM, 20 mM, 20 mM + PMA). Light grey bars: Control group non-deglycosylated. Black bars: Deglycosylated cells treated with Endo H and PNGase F. Insulin secretion levels are provided in fold over basal. Asterisk indicates statistical significance (P<0.01) of deglycosylated cells at 20 mM glucose + PMA compared to the control group.

Insulin release was significantly higher from deglycosylated cells when treated with PMA at high glucose concentrations compared to the control group and providing a relation between glycosylation and insulin secretion.

3.8 Bioinformatic analysis of the *PMM2* promoter region

The fact that the HIPKD mutation lies within the regulatory region of *PMM2* and is located in a transcription factor binding motif for ZNF143 gave rise to look into the genomic region on chromosome 16 in more detail. Our bioinformaticians conducted an elaborate *in silico* analysis within the UCSC genome browser (Hg19) and identified a chromatin loop of about 220 kb that includes *PMM2* and three additional genes (*TMEM186, ABAT, CARHSP1*) and is limited by CTCF motifs. Chromatin segmentation and transcription factor binding sites were assessed within that predicted looping region in different human cell lines. These data sets (ChIP-Seq, Hi-C and ChIA-PET) were extracted from ENCODE (Encyclopedia of DNA elements). In particular, binding sites for CTCF, ZNF143 and HNF4A were of interest, as 1) CTCF motifs are known for their loop limiting function [224] and 2) to confirm the previous results of ZNF143 binding to the *PMM2* promoter. HNF4A was of certain interest, because of its predominant expression in kidney, pancreas and liver – the three affected organs in HIPKD – consistent with the idea of a tissue-specific pathomechanism in HIPKD.

Figure 35 shows the result of the *in silico* analysis. CTCF motifs at the outer part mark the loop and face each other – a typical feature of CTCF motifs [225]. A strong ZNF143 binding site within the *PMM2* promoter matches the EMSA data nicely and there are several clusters of HNF4A sites of different strength towards the outer part of the chromatin loop. Interestingly, the chromatin throughout the looping region is active exclusively in HepG2 cells, which is a liver carcinoma cell line with the *PMM2* promoter located almost in the centre. In the other five cell lines the chromatin seems to be generally active only downstream of the *PMM2* promoter.

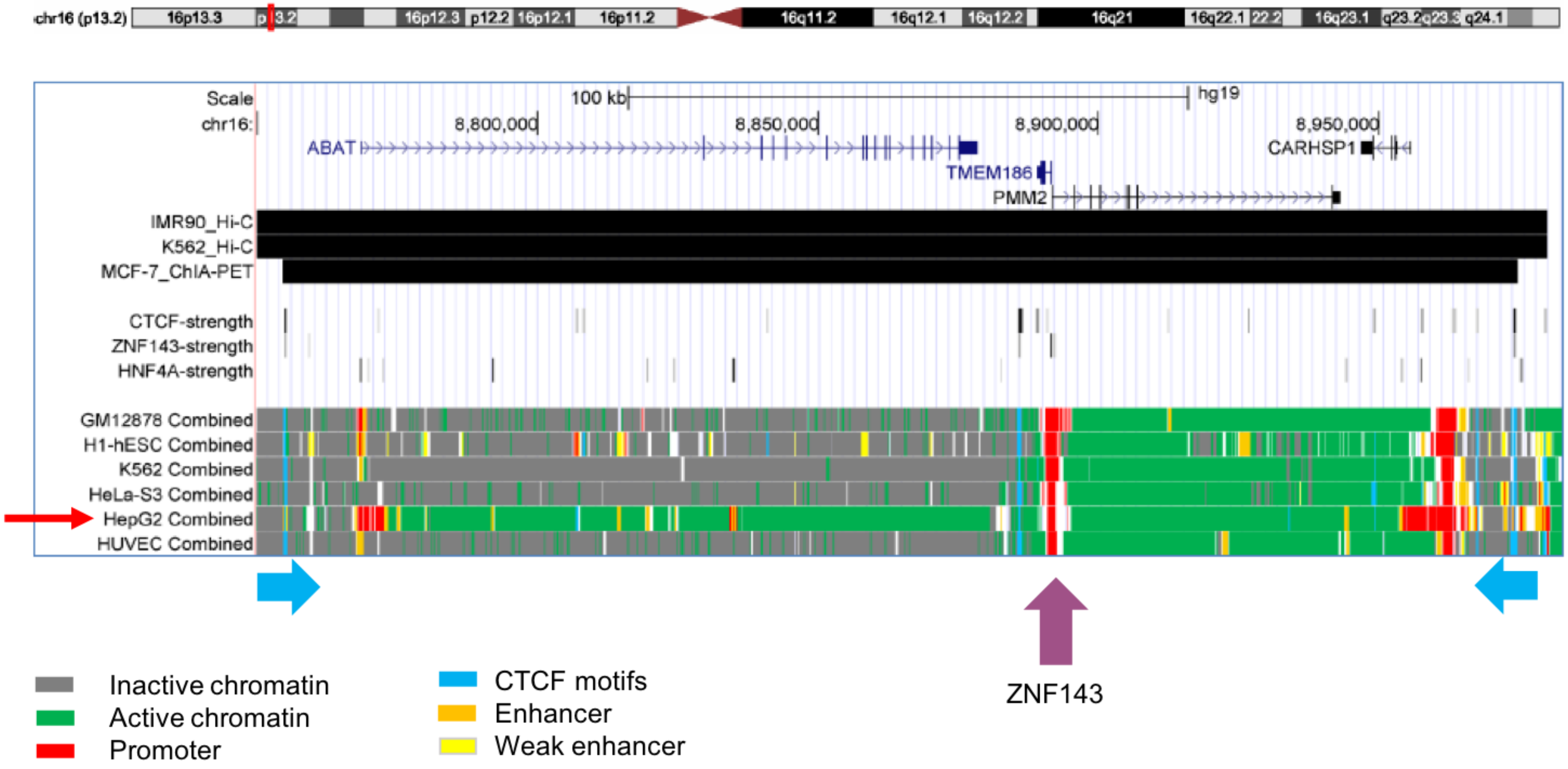


Figure 35: *In silico* analysis of the genomic region around *PMM2* on chromosome 16.
A chromatin loop of around 220 kb is indicated. Top: Tracks indicating the genes within the region. Middle: Tracks indicating CTCF, ZNF143 and HNF4A binding sites and strength according to grayscale. Bottom: Chromatin segmentation of the region in six human cells lines (GM12878, Blood; H1-hESC, embryonic stem cells; K562, leukocytes; HeLa-S3, Cervix; HepG2, liver; HUVEC, Umbilical vein). The colour code indicates important characteristics of the chromatin: Grey: Inactive chromatin. Green: Active chromatin. Red: Promoter regions. Blue: CTCF motifs. Orange: Enhancers. Yellow: Weak enhancers.

3.9 Summary of project background and conclusion

HIPKD is a combination of polycystic kidney disease, hyperinsulinism with partial liver involvement and of autosomal-recessive inheritance, that had not been described before. The disease-causing mutation was identified to be a G to T substitution (c.-167G>T) in the regulatory promoter region of the *PMM2* gene, located within a transcription factor binding site for ZNF143 – a ubiquitously expressed protein known to be crucially involved in 3D chromatin conformation [226]. The affected patients all share this noncoding mutation either in homozygous or compound heterozygous state. Impaired ZNF143 binding to the mutant promoter and reduced gene expression when under control of the same promoter were confirmed *in vitro.* The absence of an overall glycosylation defect in our patient and the distinct manifestation in three organs clearly separate the HIPKD phenotype from the well-described CDG-1a due to biallelic coding mutations in *PMM2*. However, based on the HIPKD phenotype there is strong evidence that glycosylation has a major impact on cyst formation and increased insulin release – the latter was experimentally confirmed with MIN6 β-cells.

The fact that HIPKD is caused by a noncoding *PMM2* mutation clearly implied, that the disease mechanism is a gene regulatory problem. Knowing of ZNF143 and its function and impaired binding to the *PMM2* promoter in HIPKD raised the idea of an altered 3D chromatin conformation in our patients. *In silico* analysis confirmed the presence of convergent orientated CTCF motifs flanking the active chromatin in HepG2 cells, indicating possible anchors for a chromatin loop [227]. It is well established, that ZNF143 mediates chromatin loop formation by interacting with CTCF motifs and gene promoters [228] suggesting, that ZNF143 facilitates a change of the loop formation by binding to both, the *PMM2* promoter and CTCF motifs simultaneously to allow interactions between promoter with distal regulatory elements at the anchors of the loop. Moreover, knowing of HNF4A and its predominant expression in kidney, liver and pancreas we linked these facts and suspect a possible distinct 3D chromatin conformation that affects regulation of *PMM2* expression by HNF4A and therefore creating the limitation of the phenotype to these three organ systems. Thus, in HIPKD, the promoter mutation causes impaired ZNF143 binding to the promoter, preventing this particular loop formation to happen, consequently disturbing HNF4A-mediated *PMM2* transcription in kidney, liver and pancreas. Figure 36 depicts this proposed disease mechanism.

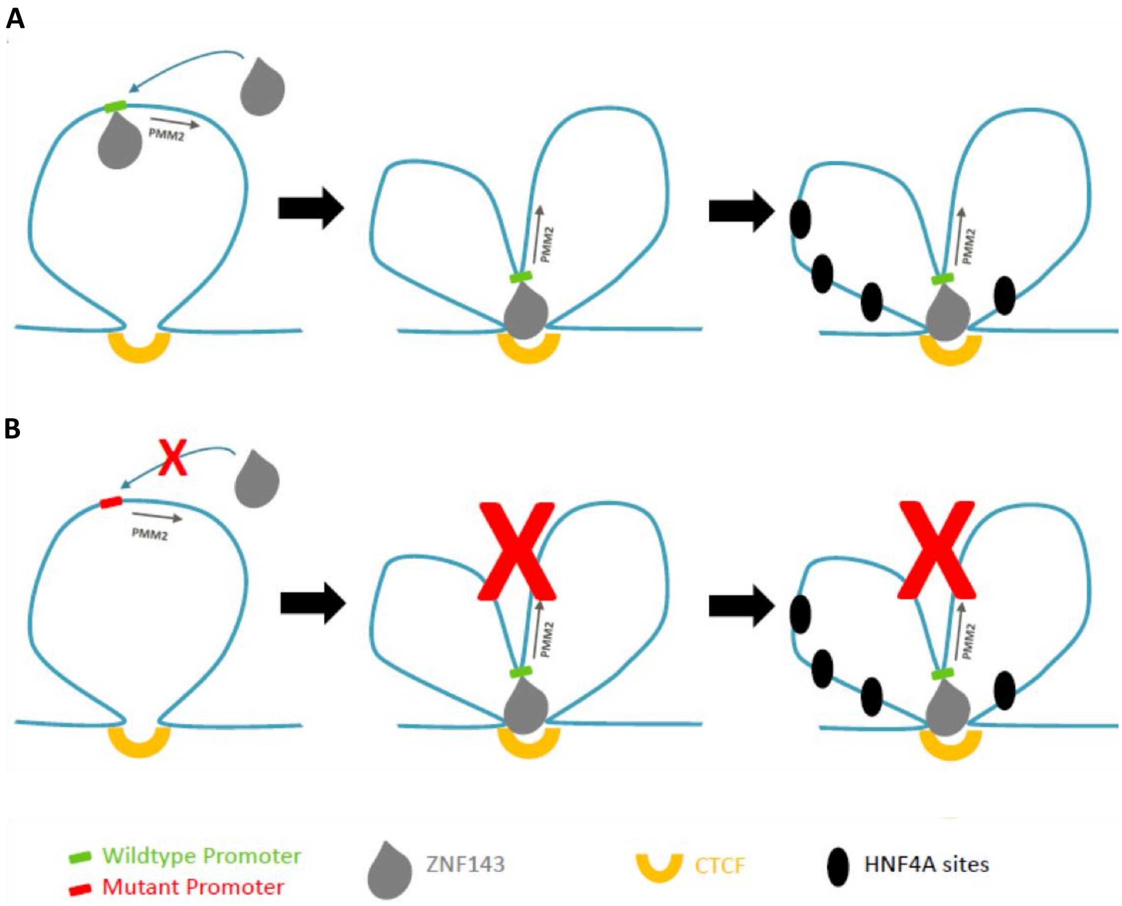


Figure 36: Simplified cartoon demonstrating tissue-specific chromatin loop formation with wt and mutant *PMM2* promoter.
A: Wt promoter. A big structural loop is formed with the *PMM2* promoter (green) on top and limited by CTCF motifs (yellow). ZNF143 binding to the *PMM2* promoter pulls down the loop, changing its 3D architecture and allowing close proximity to CTCF (functional loop) and interaction with HNF4A (black). B: Mutant promoter: ZNF143 binding to the PMM2 promoter is impaired preventing functional loop formation and interaction of ZNF143 and promoter with CTCF and HNF4A causing disturbance of *PMM2* gene regulation/transcription.

3.10 Next steps

The absence of a suitable cell or tissue model motivated us to develop such to facilitate the investigation of chromatin structure and looping. To address this and to investigate whether our hypothesis is right or wrong, a HIPKD cell model was to be generated with CRISPR-Cas9. In addition, it was of interest to show dysfunctional *PMM2* expression in human. For this purpose, gene expression studies with patient cells from one of the compound heterozygous patient were performed to confirm the gene expression results obtained *in vitro.* These cells also allowed us to measure the expression level of each

PMM2 allele separately in an allele-specific assay to investigate whether the expression of the allele under control of the mutated promoter was reduced. Furthermore, the effect of impaired glycosylation in HIPKD was investigated using human pancreatic islets to confirm the result obtained from mouse pancreatic β-cells. All this work was addressed during this PhD.

Chapter 4: The expression of *PMM2* in HIPKD

4.1 Introduction

This chapter focusses on *PMM2* expression levels in cells in primary culture from the nephrectomy specimen of a compound heterozygous patient (c.-167G>T/c.422G>A), to investigate whether the promoter mutation has an influence on *PMM2* expression in HIPKD or not. In addition to the promoter mutation (c.-167G>T) located on one allele, this patient also carries a point mutation within the coding region of *PMM2* at nucleotide position c.422G>A (p.Arg141His) on the other *PMM2* allele.

In particular of interest was to determine the overall gene expression of *PMM2* in HIPKD patients compared to controls, but also the expression of the individual *PMM2* alleles (mutant vs. wildtype promoter) in comparison to each other. In addition, the expression of the neighbouring genes of *PMM2*, *TMEM186*, *ABAT* and *CARHSP1* and two control genes (*PPIA*, *CDH1*) were also included in the study.

TMEM186 encodes for a potential transmembrane protein (RefSeq NM_015421) and not much more is known about it, except that is shares the promoter with *PMM2*. *ABAT* encodes the enzyme 4-aminobutyrate amino transferase (RefSeq NM_020686) and converts the neurotransmitter gamma-aminobutyric acid (GABA) into semialdehyde. GABA primarily exerts inhibitory functions within the central nervous system. ABAT is active in about one third of synapses in the human body and expressed in several tissues, predominantly in the liver [229]. Mutations in the gene cause GABA-transaminase deficiency, a neurological disorder presenting with psychomotor retardation, hypotonia, hyperreflexia and seizures [230]. *CARHSP1* translates into the calcium regulated heat stable protein 1 (RefSeq NM_001042476), a cytoplasmic protein which enhances mRNA stability. Using the example of tumour necrosis factor α (TNF-α) transcription, *CARHSP1* is presumed to prevent mRNA degradation, indirectly promoting translation by enabling the mRNA to re-enter the translational pool [231].

4.2 The aims of this chapter

- Performing gene expression studies with quantitative digital PCR
- Measure whether there is allelic discrimination detectable between expression levels of the two mutant allele variants of *PMM2* in the compound heterozygous patient
- Study the expression levels of *PMM2* and surrounding genes in specific gene expression assays

4.3 Results

4.3.1 Allelic discrimination of the *PMM2* variant alleles

The expression levels of the *PMM2* alleles in the compound heterozygous patient were measured in an allele specific discrimination assay (ASA) using quantitative dPCR. Expression levels of the *PMM2* alleles in patient and control samples were determined using a custom allele specific dPCR assay (Assay-ID: AHGJ9ED, (2.2.3.1).

In this case, two probes were needed to discriminate the *PMM2* alleles and both were part of the TaqMan® Assay used. The probes were differently labelled with a fluorophore (FAM™ or VIC®) for discrimination of the expression levels of each *PMM2* allele (Alleles 1 and 2). Both probes were designed to span the nucleotide position 422 (Ref.-Seq.: NM_000303.2) of *PMM2* within the cDNA of patient and control. In this way, discrimination of the two *PMM2* alleles was possible due to a variant (c.422G>A) on the allele (Allele 1) not carrying the promoter mutation, whereas the allele having the promoter mutation (c.-167G>T, Allele 2) is wildtype for this variant (c.422G). Therefore, the FAM™-labelled probe was used to detect the allele with the promoter mutation (Allele 2) and the VIC®-labelled probe was specific for the allele carrying the c.422G>A genetic variant (Allele 1).

Prior to the actual experiment, defined amounts of patient (10ng, 25 ng) and control cDNA (20 pg-100 fg) were tested to assess the required amounts to be within the desired range of target sequence (200-2,000 copies/µL). Based on this, a final input of 25 ng and 200 fg of patient cDNA and control construct mix, respectively (data not shown), was used. Negative controls (No reverse transcriptase patient control, no template control, control constructs 1 and 2 individually) were included to assess assay specificity. This is reflected by Figure 37, showing the allele fraction (%) of the two constructs measured individually.

To assess allelic discrimination, each cDNA sample was measured once (n=3) and for the control, the 1:1 construct mix was prepared three times independently (n=3), each also measured once. A ratio based on the fluorescent signals (FAM™ or VIC®) detected for each allele in patient and control samples divided over the total fluorescence (FAM™ + VIC®) measured was calculated to identify differences in expression levels of the *PMM2* alleles in percent (Allele fraction, Figure 38). The values plotted are means±SEM.

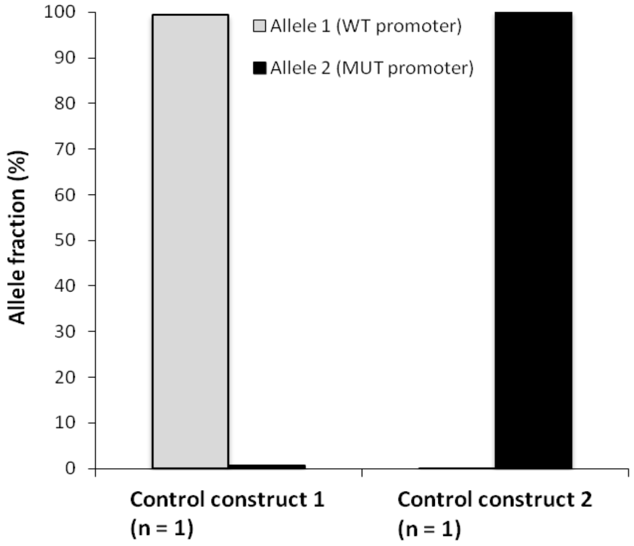


Figure 37: Allele fraction (%) of control constructs 1 and 2 determined by quantitative dPCR. The percentages of the fluorescent signals detected for the expression of each construct over the total signal measured are shown. Grey bars indicate the detection of Control construct 1 (mimics *PMM2* Allele 1) carrying the c.422G>A SNP (VIC®-labelled probe). Black bars indicate the expression of Control construct 2 (mimics *PMM2* Allele 2), wt at nucleotide position 422 (Ref.-Seq.: NM_000303.2) (FAM®-labelled probe). Each construct was measured once with a total input of plasmid DNA of 200 fg/sample.

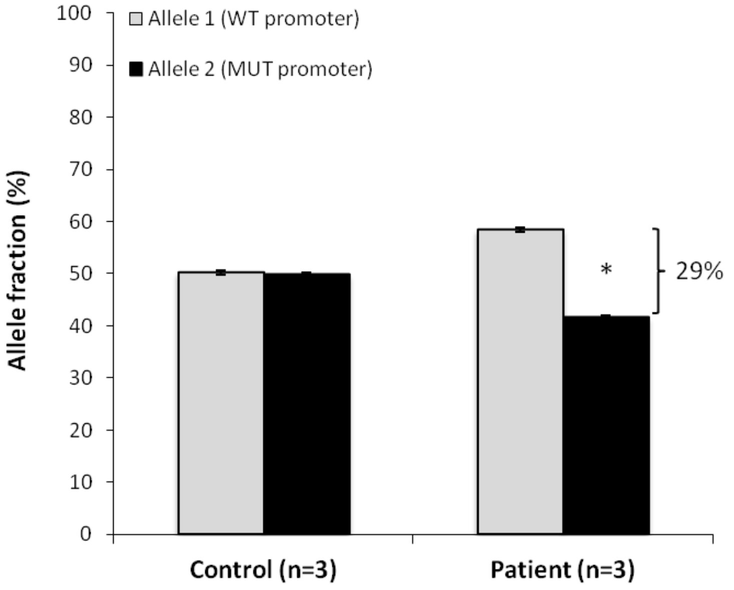


Figure 38: Allelic discrimination of *PMM2* alleles determined by quantitative dPCR.
The percentages of the fluorescent signals detected for the expression of each allele over the total signal measured for control (1:1 construct mix) and patient cDNA samples are shown (Allele fraction). Grey bars indicate the detection of Allele 1 negative for the promoter mutation (VIC®-labelled probe). Black bars indicate the expression of the *PMM2* Allele 2 carrying the promoter mutation (FAM®-labelled probe). Three patient cDNA samples (equal to 25 ng total RNA) were measured 1x (technical replicates, n=3). The control (1:1 construct mix, 200 fg) was assayed 3x individually (technical replicates, n=3). Samples were within the appropriate range of target numbers (200 – 2,000 copies/µL). Values are means±SEM. *p-value < 0.05.

Overall, the expression levels of the *PMM2* alleles in the patient are clearly unequal compared to the control (1:1 plasmid construct mix) Figure 38. As it was expected, in the control sample alleles 1 and 2 were expressed in equal amounts with 51% and 49%, respectively. In contrast, the total expression in the patient measured for Allele 1 (wt promoter) accounted for 58±0.33% versus 42±0.33% of that determined for the allele under control of the mutated promoter (Allele 2, MUT promoter). In relation to each other, this implies a 29% reduction of expression levels due to the promoter mutation in Allele 2 compared to Allele 1 of *PMM2* in the compound heterozygous patient. A two-tailed and unpaired student’s t-test verified the significance of the described allele expression differences statistically (p-value < 0.05).

4.3.2 Gene expression of *PMM2* and its neighbours in compound heterozygous HIPKD patient and wildtype cells

To evaluate the overall expression levels of the genes of interest, *PMM2, TMEM186, CARHSP1 and ABAT,* a dPCR gene expression assay was performed. In addition, two control genes, *PPIA* and *CDH1*, which are not in the neighbourhood of the promoter and thus expected to be expressed equally in patient and control, were also assayed for direct comparison to the results of the four genes of interest.

For detection and quantification of the six genes, specific TaqMan® gene expression assays containing forward and reverse primers and fluorophore-labelled probes were used (Table 7). Absolute quantification was measured and provided by the dPCR system as copies/µL and analysed and processed as such. The plotted data are the means±SEM of two to four (n = 2, n = 3, n = 4, Figure 39) cDNA samples for both, patient and control (2.2.2), each measured in duplicate (*ABAT, CDH1*) or triplicate (*PMM2, TMEM186, CARHSP1, PPIA)*. Statistical significance was assessed with a two-tailed and unpaired student's t-test.

Optimal cDNA input was also assessed prior to the final dPCR gene expression series (Data not shown). This suggested a final cDNA input per sample (per chip) and assay equivalent to 25 ng total RNA for all genes aside from *TMEM186* (50 ng) and *ABAT* (200 ng). A no reverse transcriptase (RT) negative control for each assay was tested in parallel for both, patient and control cDNA to successfully exclude possible DNA contamination (Data not shown).

Figure 39 represents the expression levels of all six genes. Expression levels of *PMM2, TMEM186, PPIA* and *CDH1* were not significantly different in the patient compared to the control. Yet, *PMM2, TMEM 186* and *CDH1* show a trend for lower expression in the patient. As expected, *PPIA* was equally expressed, with about 1,000 copies/µL in both groups tested. Even though *ABAT* expression was not significantly different as well, its expression trended towards a decrease in the control sample. *CARHSP1* instead, was the only gene within the chromatin loop assayed, which showed significantly reduced expression in the patient, with a decrease of approximately 30% (913 copies/µL in control versus 540 copies/µL in patient).

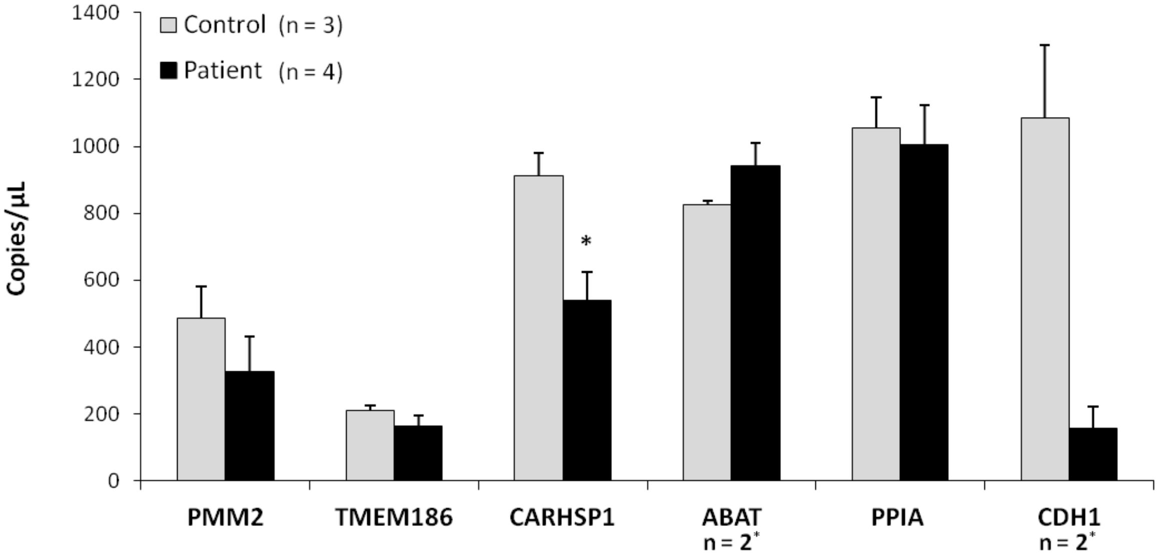


Figure 39: Expression levels of *PMM2, TMEM186, CARHSP1, ABAT, PPIA and CDH1* in patient and control samples determined by quantitative digital PCR.
Absolute quantification (copies/µL) of the expression levels of *PMM2, TMEM186, CARHSP1, ABAT, PPIA and CDH1* in patient (Black bars) and control (Grey bars). cDNA input was equal to 25 ng (*PMM2*, *CARHSP1*, *PPIA* and CDH1), 50 ng (*TMEM186*) and 200 ng (*ABAT*) of total RNA for patient and control samples. In total four patient (n=4) and three control (n=3) cDNA samples for *PMM2, TMEM186, CARHSP1* and *PPIA* were measured three-times each and the means were taken. *ABAT* and *CDH1* were determined using two cDNA samples for patient and control, each measured twice and the means were taken (n=2*). Values are means±SEM. *p < 0.05.

In general, there clearly is some variation in the average expression levels of the genes tested, as is evident from the controls. *PMM2* and *TMEM186* are expressed the lowest, with 500 and 200 copies/µL, respectively, followed by *CARHSP1* and *ABAT* being expressed much higher with values in the range of 900 to 950 copies/µL and finally *PPIA* and *CDH1* slightly surpass the two previous ones with copy numbers of 1,000 to 1,100 per micro litre.

4.4 Discussion

Gene expression studies based on quantitative PCR are a well-established tool within functional genomics to quantify RNAs. It covers a broad range of applications in many fields of research including the quantification of gene transcription by measuring mRNA levels of particular genes [232]. The significant advantage of qPCR is its high sensitivity and precision, making the detection of very small RNA amounts possible. Besides, its utilisation is relatively simple and screening of large sample numbers and different genes in the same experiment is possible, providing higher flexibility compared to more traditional techniques, such as Northern blot and in-situ hybridisation where input/output relation (dynamic range) is more limited [233].

Apart from the well-established *real-time* qPCR for quantification of DNA/RNA sequences, a novel approach has settled in modern research, improving the possibilities of qPCR due to the ability of “digital” measurements [234]. Quantitative dPCR simply counts the number of a gene transcript at a given time precisely and with excellent sensitivity and specificity. It provides absolute values and higher accuracy compared to classic *real-time* qPCR. The high precision and accuracy dPCR results from diluting and partitioning the sample into thousands of separate reaction chambers so that each contains one or no copy of the target sequence. Hence, positive and negative partitions can simply be counted resulting in exact copy numbers (absolute values) and absolute quantification is guaranteed making the use of reference genes of known concentration redundant. Furthermore, dPCR allows the detection of 30% or smaller differences in expression levels. It can determine whether a variant occurs in five out of six copies and it can detect one different allele within thousands. Thus, dPCR is helpful where qPCR has its limitations, being unable to distinguish expression differences or copy number variants below two-fold [235].

First tested and applied by Vogelstein and Kienzler in 1999, they successfully could identify the *KRAS* mutation causative for colorectal cancer in a stool specimen. They set a genetic hallmark for early detection of tumorgenesis in a mixed cell population [236], reflecting the opportunities dPCR has in modern molecular research and especially clinical applications to detect rare mutations at an early stage crucial for diagnosis and treatment [237] as well as the quantitative measurement of allelic imbalances [238,239] . The ability to detect genetic variants expressed in very low copy numbers makes digital PCR a powerful tool in clinical diagnostics and optimally meets the requirements for the here described study.

As expected, a reduced expression of the *PMM2* allele with the mutant promoter was demonstrated by the allelic discrimination assay (Figure 38). The data repeatedly demonstrate significant reduction in expression levels of the *PMM2* allele under control of the mutant promoter, which fundamentally strengthens our hypothesis of a dysfunctional *PMM2* gene regulation causative for the disease. Yet, the overall expression of *PMM2* and also *TMEM186* turned out to be not as clear after differential expression analysis (Figure 39). However, a trend for a reduction in expression levels of both genes in patient compared to control (wt) samples could be detected, thus supporting the hypothesis of disturbed bidirectional promoter activity of the promoter variant.

The results for the additional genes tested in the gene expression analysis are difficult to interpret. We tested all genes considered to be located within the chromatin loop containing *PMM2* (Figure 35) to get a first overall picture of this region. Furthermore, to get an idea of what might be happening on the regulatory level with respect to the *PMM2* promoter mutation.

As outlined above (Figure 36), according to our loop hypothesis, bioinformatical predictions and ChIP-Seq data pointed towards the existence of HNF4A binding sites close to the linked CTCF motifs suggesting ZNF143 binding to the *PMM2* promoter is crucial to bring promoter and HNF4A into close proximity to activate *PMM2* and *TMEM186* expression. In contrast, we concluded, that *CARHSP1* and *ABAT* expression was independent from *PMM2* promoter activity. This in turn implied equal expression levels of these genes in patient and control. However, *CARHSP1* and *ABAT* expression revealed a different expression pattern in patient and control, although this was statistically significant only for *CARHSP1*. In addition, the results generated in the same experiment for *PPIA* and *CDH1*, used as reference genes, did also not completely meet the expectations of being expressed equally. While *PPIA,* as expected, was expressed equally in both patient and control kidney [240], for *CDH1* the opposite effect was observed with less expression in the patient. *CDH1* encodes for E-cadherin, one of the major adhesion molecules crucially involved in cell-cell adhesion in renal epithelial tissue [241]. It forms a multiprotein complex with polycystin-1, polycystin-2 and β-catenin, which is disrupted in polycystic kidney disease cells. In ADPKD for example, it was shown that E-cadherin expression was significantly reduced and not part of the protein complex anymore affecting cellular stabilisation [242]. Due to its interaction with β-catenin, E-cadherin is also a modulator in signal transduction having an impact on transcriptional regulation and cell cycle control. β-catenin is heavily involved in regulation of cell proliferation, apoptosis and tumorigenesis (e.g. via Wnt signalling pathway). This indicates a direct correlation of reduced E-cadherin synthesis with uncontrolled cell growth and metastasis as β-catenin dissociates from E-cadherin to exert its transcriptional activity under normal circumstances [243], thus likely contributing to cyst formation. As the patients' kidney was in its end-stage, traversed by cysts with functional renal tissue replaced by fibrotic scar tissue at the time of nephrectomy, *CDH1* gene expression was likely seriously reduced in the HIPKD patient sample due to the loss of epithelial integrity [244]. Moreover, as the RNA used was obtained from a heterogeneous sample (containing fibroblasts besides renal epithelial cells) of primary renal cells, grown in different wells, this likely contributed to the variability. Further, RNA extraction was done partly after different passages, which can also alter gene expression patterns [245]. All these potential contributors need to be kept in mind when analysing the data.

Other factors influencing qPCR outcome are technical hurdles as variability increases if experimental performance is lacking accuracy [233]. The efficiency of RT PCR based cDNA synthesis is the most important but also potentially inaccurate step in expression analysis. It can introduce errors generated by effects of mRNA secondary and tertiary structures, variation in priming efficiency, and properties of the reverse transcriptase [246]. It is crucial to always use the same amount of RNA and ideally generate cDNA in the same reaction, to keep results comparable. Varying mRNA degradation processes are considered being major control points of gene expression, significantly contributing to gene regulation in eukaryotes [247,248]. Different stimuli from developmental- (proliferation), environmental- (e.g. temperature) or such of disease-related origin like fibrosis, as it was likely the case in the patients' kidney, can influence mRNA degradation.

Chapter 5: Insulin secretion studies with human pancreatic islets

5.1 Introduction

To investigate a possible impact of the *PMM2* promoter mutation on insulin secretion, two sets of experiments (Assay 1, Assay 2) were performed to test whether deglycosylation affects the functionality of β-cells within the pancreatic islets cell cluster. For these experiments it was necessary to mimic a dysfunction of the PMM2 protein, which normally acts at an early state within the protein glycosylation pathway during posttranslational modification. In our experiments, however, we opted for deglycosylation of glycoproteins, using commercially available recombinant deglycosylation enzymes (glycosidases) which are able to remove sugar residues attached to various proteins.

To investigate the effect of deglycosylation on insulin secretion, two batches (Batch 1 and 2) of human pancreatic islets were kindly provided by the Oxford Consortium for Islet Transplantation (OXCIT) and used in two differently designed deglycosylation assays. Secreted (extracellular) and intracellular insulin concentration produced by these islets were measured with a human insulin ELISA assay.

ELISA represents a suitable and common method to detect a protein of interest in a liquid sample based on the principle of immunobiochemistry: Insulin is captured by an immobilised specific mouse monoclonal anti-Human Insulin antibody and can be detected due to the addition of an enzyme (Streptavidin-Horseradish Peroxidase Conjugate) and an enzyme-linked secondary detection antibody (Biotinylated Monoclonal Mouse anti-Human Insulin Antibody), which builds a complex with the antigen (Insulin). The addition of a chromogenic enzyme substrate (TMB, 3, 3’,5,5’-tetramethylbenzidine) causes a chromogenic enzyme-substrate reaction resulting in a colour change (blue). The addition of HCl (acidification) stops the reaction and turns the blue colour into yellow allowing to measure the ELISA reaction colorimetric in an appropriate spectrophotometric plate reader at an absorbance of 450 nm. The colour intensity is directly proportional to the insulin concentration in the samples allowing the detection and quantification of the insulin content. Therefore, the unknown insulin concentrations (µU/mL) could be extrapolated by interpolation of the absorbance values from a standard curve (standards: 5, 20, 50 100 and 200 µU/mL) generated in the same assay.

The first batch of human pancreatic islets (Batch 1) received was used to conduct a glucose-stimulated insulin secretion (GSIS) with a subsequent Peptide N-Glycosidase F

(PNGase F) treatment series (2.3.1). The experimental set up was based on the standard operating procedure of the National Institute of Health (NIH) applied for the functional assessment of purified human pancreatic islets within the clinical islet transplantation consortium and adjusted to the purposes of this experiment. PNGase F (Amidohydrolase) is a commonly and widely used deglycosylation enzyme purified from *Flavobacterium meningosepticum*. It cleaves the GlcNac-ASN bond of all three major N-Glycans, regardless of the oligosaccharide class attached [249,250].

With the second batch of human pancreatic islets (Batch 2) the impact of deglycosylation on insulin secretion was approached differently by conducting the experiment (2.3.2) according to the protocol used by a collaborating group which addressed this question in a deglycosylation experiment with MIN6 mouse β-cells [209]. In this setting the deglycosylation of human pancreatic islets took place prior to glucose stimulation. In addition to PNGase F, a second deglycosylation enzyme, Endoglycosidase H (Endo H), was used. Endo H is a recombinant glycosidase which specifically cleaves at the dimer core (chitobiose) of high mannose and some hybrid types of N-linked sugars [251]. To enhance the predicted effect of an increased insulin release, the protein kinase C (PKC) activator PMA (phorbol- 12 myristate-13 acetate) was added to the high glucose solution as a third condition. PMA mimics the function of GLP-1 (incretin hormone glucagon-like peptide), a hormone produced in the distal intestine that stimulates insulin secretion in response to meals. Effectively, like GLP-1, PMA-dependent PKC activation via diacylglycerol (DAG) stimulates insulin secretion by activating L-type Ca^{2+} channels and the increasing Ca^{2+} concentration eventually leads to insulin synthesis and secretion [155,252].

5.2 The aims of this chapter

- Study insulin secretion in human pancreatic islets
- Mimic *PMM2* dysfunction using glycosidases (PNGase F and Endo H)
- Study the effect of deglycosylation on beta cells within human pancreatic islets
- Investigate whether deglycosylation increases insulin secretion

5.3 Results

5.3.1 Insulin measurement of glucose-stimulated and PNGase F treated human pancreatic islets (Assay 1)

To investigate the effect of deglycosylation on insulin secretion, human pancreatic islets were treated with 2,000 U of PNGase F after a series of glucose stimulation (GSIS) with low (2.8 mM) and high (28 mM) concentrated glucose solutions to further measure Insulin secretion using ELISA (2.3.3). The GSIS series was performed to test the eligibility of the islets used for the experiment and to investigate whether PNGase F treatment induces insulin secretion to the same or even a higher level.

The experiment was performed twice using the same islets batch. After equilibration, all samples were incorporated into the GSIS series. The final PNGase F treatment was done with three of the samples and one served as a control only treated with the enzyme buffer (Na_3PO_4, 50 mM final concentration).

Insulin secretion was determined measuring the accumulated insulin in the test medium (KRBB) using the human insulin ELISA assay. For this, the medium samples before (T_0) and after incubation (T_{60}) of each test condition were collected during the experiment. In addition, the islets were harvested in RIPA buffer immediately at the end of each experiment for lysis. The islets lysates were also measured in the ELISA assay to assess the intracellular insulin content of each sample.

Insulin concentrations measured in ELISA were converted into µU/mL using the standard equation obtained from the standard curve measured in each ELISA assay. To evaluate the response of the islets to the GSIS and PNGase F treatment the ratio reflecting the difference (increase or decrease) of the secreted insulin concentration between T_0 and T_{60} of each condition was calculated for each sample and the means were taken. One-way ANOVA and a post hoc Bonferroni test were used to assess statistical significance. The measurement of intracellular insulin from islets lysates and also the absolute concentrations (µU/mL) of means and single values of secreted insulin are provided in the appendix for both experiments (see appendix Figure 72, Figure 73, Table 16, Table 17)

The temporal difference between the two experiments was seven days and analyses were therefore kept separately. The results of both experiments are shown in Figure 40 and Figure 41.

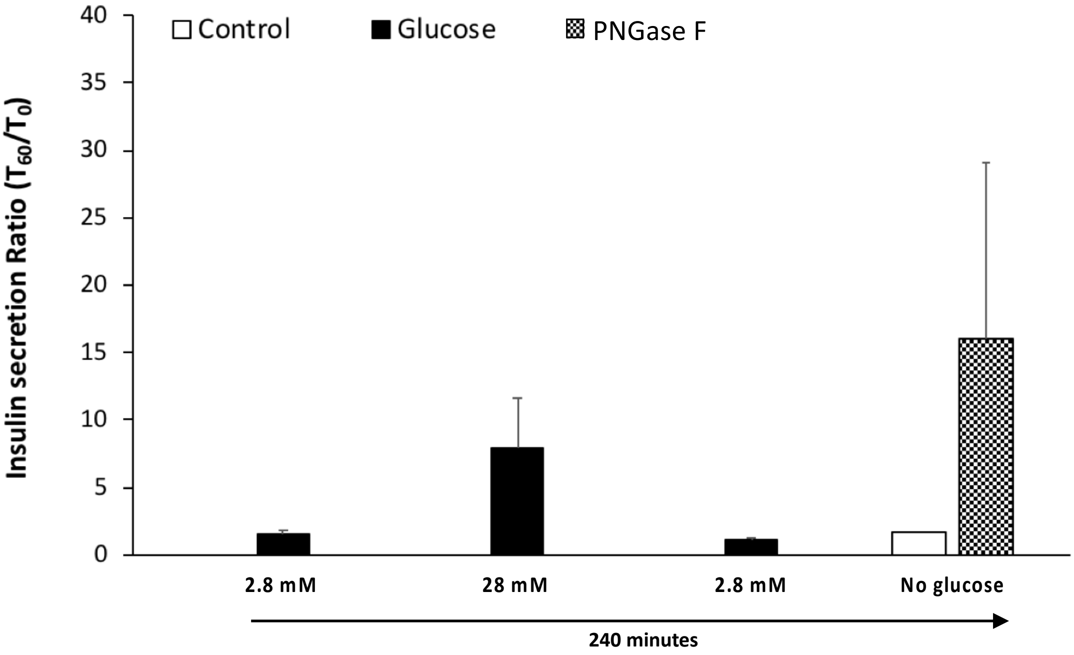


Figure 40: Assay 1: Experiment 1 human pancreatic islets Batch 1.
Differences of Glucose-stimulated insulin secretion (GSIS) and subsequent PNGase F treatment for deglycosylation are shown. The ratios (T_{60}/T_0) of secreted insulin are plotted and were calculated from absolute concentrations obtained from ELISA. The ratio reflects the change in insulin secretion per condition over an incubation period of 60 minutes. The entire experiment lasted for 240 minutes. The experiment was performed in 24-well-plates. Four islets samples were included in the experiment. For each sample 300 IEQ (theoretical number) were seeded in one well. In total, four different conditions were tested and completed in KRBB (Krebs Ringer bicarbonate buffer). All four islets samples were included in the GSIS series (n=4) and stimulated with 2.8 mM, 28 mM and 2.8 mM glucose (Black bars). After GSIS three of the samples were incubated with 2,000 U PNGase F pre-prepared in its enzyme buffer (Na_3PO_4, 50 mM final concentration) in 0 mM glucose (n=3, chequered bar). One sample was used as a control only incubated with enzyme buffer (Na_3PO_4, 50 mM final concentration) also in 0 mM glucose (n=1, white bar). Standards were measured from 5-100 µU/mL (data not shown). Values are means±SEM when applicable. Statistical significance ($P<0.05$) was tested using a one-way ANOVA followed by a Bonferroni correction for multiple comparisons to compare between the different conditions

The difference in insulin secretion during glucose induced insulin secretion series (GSIS) in the first experiment (Figure 40) turned out not to be statistically significant. However, a trend of a higher increase in insulin release when treated with 28 mM glucose (Ratio T_{60}/T_0 7.91±3.77) compared to low glucose incubation (Ratio T_{60}/T_0 1.51±0.26) was observed. Likewise, a decrease of insulin secretion by reducing the glucose concentration back to 2.8 mM (Ratio T_{60}/T_0 1.15±0.18) seems to be likely. When treated with PNGase F in 0 mM glucose KRBB, the change in insulin concentration over 60 minutes was not significant compared to the control sample (Ratio T_{60}/T_0 1.67). There was a trend of an increase, which was based on one out of the three samples with a very low T_0 concentration compared to the other two (Table 16). This explains the strong variation with ratios of a range from 2.67 to 42.27 (16±13.15). Similarly, compared to the high glucose stimulation and also both low glucose stimulation, no significant difference was observed when islets were treated with PNGase F due to the high fluctuation.

The absolute concentrations of single values ranged between approximately 900 and 2,000 µU/mL for secreted insulin, whereas this was higher in islets lysates with insulin concentrations of 13,500 to 46,000 µU/mL (see appendix Figure 72, Table 16).

The quality controls of the assay were within the predefined range (2.3.3) and accounted for 8.53 µU/mL and 25.86 µU/mL for low and high quality control, respectively.

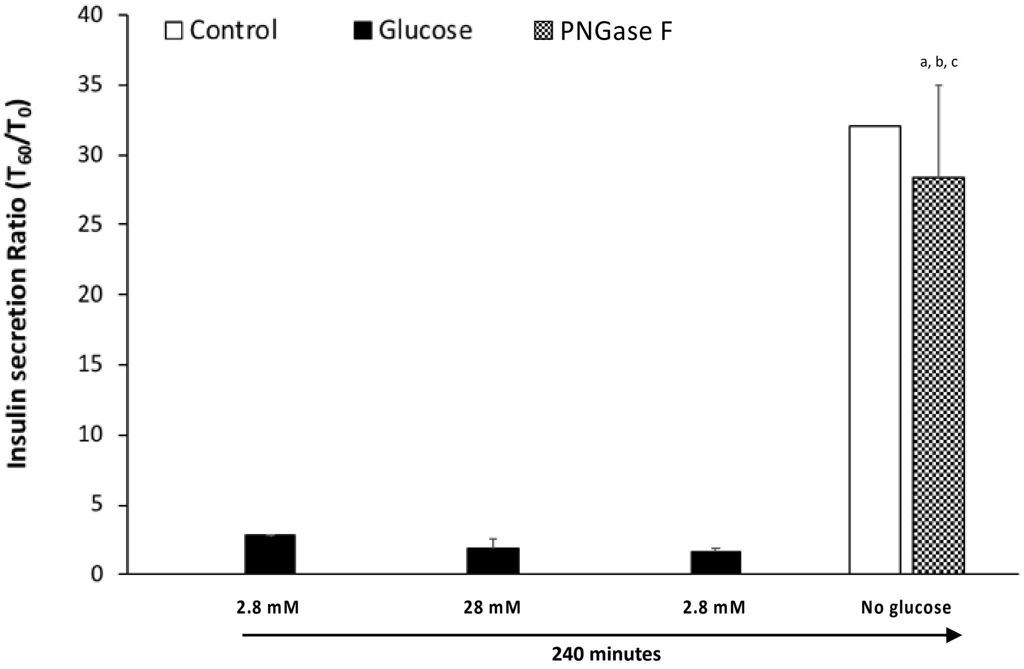


Figure 41: Assay 1: Experiment 2 human pancreatic islets Batch 1.

Repetition of experiment 1 seven days later. Differences of Glucose-stimulated insulin secretion (GSIS) and subsequent PNGase F treatment for deglycosylation are shown. The ratios (T_{60}/T_0) of secreted insulin are plotted and were calculated from absolute concentrations obtained from ELISA. The ratio reflects the change in insulin secretion per condition over an incubation period of 60 minutes. The entire experiment lasted for 240 minutes. The experiment was performed in 24-well-plates. Four islets samples were included in the experiment. For each sample 300 IEQ (theoretical number) were seeded in one well. In total, four different conditions were tested and completed in KRBB (Krebs Ringer bicarbonate buffer). All four islets samples were included in the GSIS series (n=4) and stimulated with 2.8 mM, 28 mM and 2.8 mM glucose (Black bars). After GSIS three of the samples were incubated with 2,000 U PNGase F pre-prepared in its enzyme buffer (Na_3PO_4, 50 mM final concentration) in 0 mM glucose (n=3, chequered bar). One sample was used as a control only incubated with enzyme buffer (Na_3PO_4, 50 mM final concentration) also in 0 mM glucose (n=1, white bar). Standards were measured from 5-200 µU/mL (data not shown). Values are means±SEM when applicable. Statistical significance was tested using a one-way ANOVA followed by a Bonferroni correction for multiple comparisons to compare between the different conditions. “a” indicates a significant difference (P<0.05) of the PNGase F-treated sample group to 2.8 mM, “b” to 28 mM and “c” to again 2.8 mM glucose.

Experiment 2 (Figure 41) represents the repetition of experiment 1 (Figure 40). Rising insulin secretion (Ratio T_{60}/T_0) during GSIS was still measurable but overall to a lower extent (Ratios 1.62±0.33 to 2.76±0.07) compared to experiment 1 (Ratios 1.15±0.18-7.91±3.77). Moreover, the trend of an increasing insulin concentration when stimulated with 28 mM glucose compared to 2.8 mM glucose could not be observed anymore. Similarly, the expected drop in insulin secretion when returning the islets back into 2.8 mM glucose failed to appear.

Even though the rise of insulin secretion during the 60-minute PNGase F treatment was significantly higher, with a factor of 28.40±6.64 compared to the GSIS series, PNGase F influence did not significantly alter insulin secretion compared to the control, which showed a 32.1-times increase after the 60-minute incubation.

Overall, the extra- and intracellular insulin concentrations measured in experiment 2 were much lower compared to experiment 1. The absolute values for both, secreted and intracellular insulin can be reviewed in the appendix (Figure 73, Table 17). The quality controls of the assay were again within the predefined range (2.3.3) and accounted for 8.06 µU/mL and 20.81 µU/mL for low and high quality control, respectively.

5.3.2 Response of deglycosylated human pancreatic islets to glucose and PMA stimulation (Assay 2)

A second, differently designed assay (Assay 2) was conducted to continue investigating the influence of deglycosylation on insulin secretion from human pancreatic islets. In this setting, islets were treated with two deglycosylation enzymes (PNGase F and Endo H) prior to glucose and PMA stimulation and the results were compared to a non-glycosylated control group. In total, three conditions were tested. The glucose concentrations changed from 2.8 mM to 2 mM for the low glucose condition and from 28 mM to 20 mM for the high glucose condition. In a third condition, islets were treated with 500 nM PMA in addition to 20 mM glucose.

As before, insulin concentrations measured in ELISA were converted into µU/mL using the standard equation obtained from the standard curve measured in each ELISA assay. The full experiment was conducted twice within two days and the results were combined. The response of the islets to the different treatments was expressed in fold over basal (fold change during treatment to the basal level of insulin secretion) and the means were plotted. The basal level of insulin secretion was calculated from values measured for non-glycosylated control samples at 2 mM glucose. To assess statistical significance, a two-way ANOVA and a post hoc two-tailed, unpaired student's t-test were applied to the data. The resulting bar chart is shown in Figure 42.

The absolute concentrations (µU/mL) of means and single values of secreted and intracellular insulin are provided in the appendix for both experiments (see appendix Figure 74, Table 18, Table 19).

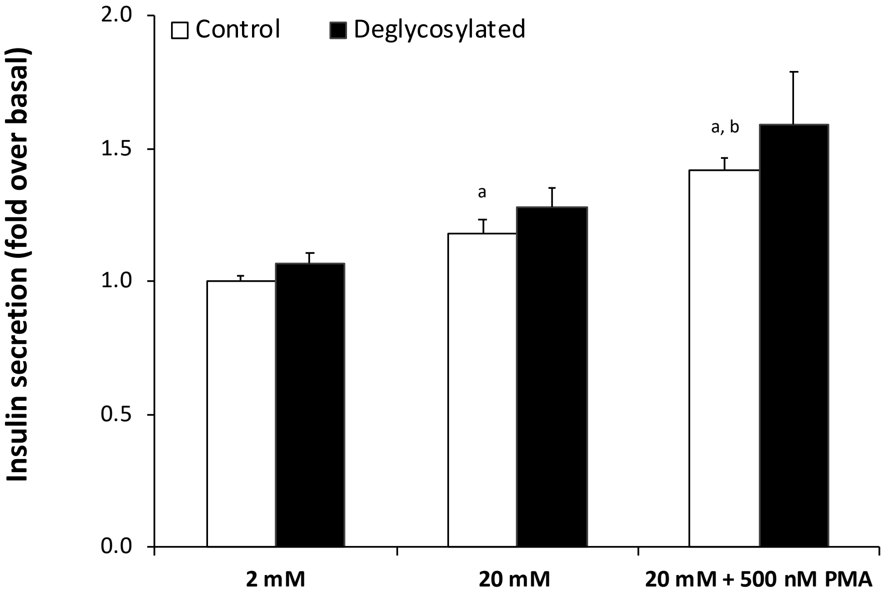


Figure 42: Assay 2: Insulin secretion from human pancreatic islets Batch 2.
Deglycoylated (black bars) vs. non-deglycosylated (control, white bars) samples from two experiments. Each experiment was performed in 24-well-plates. Twelve islets samples were included per experiment. For each sample 150 IEQ (theoretical number) were seeded in one well. Six samples were treated with PNGase F and Endo H for deglycosylation and six were kept glycosylated in a first incubation step at 2 mM glucose (120 minutes). In total, three different conditions, 2 mM glucose, 20 mM glucose and 20 mM glucose plus 500 nM PMA were tested and completed in BSS (Balanced salt solution) in a second step (60 minutes incubation). Two islets samples were included per condition per experiment. The data shown represent the mean±SEM from both experiments per condition (n=4) and group (control samples, deglycosylated samples). In total two outliers were omitted from the deglycosylated group and n-numbers reduced accordingly. One value from the 2 mM glucose stimulation (n=3) and one from the 20 mM glucose plus PMA treatment (n=3) were omitted. Standards were measured from 5-200 µU/mL (data not shown). Statistical significance was tested using a two-way ANOVA followed by a two-tailed, unpaired student’s t-test to compare between the different conditions and groups. “a” indicates a significant difference ($P<0.05$) of the 2 mM treatment compared to the other conditions. “b” indicates a significant difference ($P<0.05$) of the 20 mM treatment compared to the other conditions.

Within the control group, increasing insulin secretion was significant with rising glucose concentrations and additional PMA treatment. Insulin concentrations measured, changed by 1.18-fold (20 mM) and 1.42-fold (20 mM plus PMA) during the 60-minute incubation compared to the basal level (2 mM non-glycosylated control). A similar trend was observed within the group of deglycosylated islets between the different conditions. However, changes were not significant compared to the basal level nor between the three conditions. Compared to the control group, deglycosylated islets appear to secrete slightly more insulin within each condition but this was not significant. In particular, when treated with PMA the response of deglycosylated islets did not show a distinct increase compared to the control as hypothesised a priori.

The quality controls accounted for 8.53 µU/mL (low QC) and 25.86 µU/mL (high QC) and therefore were within the defined range indicating appropriate performance of the assay (2.3.3).

5.4 Discussion

Glucose–induced insulin secretion (GSIS) is commonly used for the assessment of quality and viability of human pancreatic islets in islets transplant centres [253]. Although the islets used for Assay 1 were of relatively poor quality (Batch 1), the data obtained in the first experiment provide initial evidence that the islets have responded to the GSIS (Figure 40), reflected by a trend of increased insulin secretion when stimulated with 28 mM Glucose. Yet, this was not statistically significant. This pattern changed in the second experiment conducted seven days later, indicating no distinct stimulation of insulin release during the GSIS series. The differences in insulin secretion levels during GSIS, when comparing experiment 1 and 2 of Assay 1 (Figure 40, Figure 41), are likely a result of decreasing islets viability due to the interval of seven days between the experiments. Islet quality decreases with time and even more rapidly when they are not in good condition after isolation, as it was the case for the batch used as pointed out before [254]. Islet quality during and after isolation at transplant centres is regularly being tested extensively *in vitro* and *in vivo* in order to assess islets identity, purity, viability, and potency to predict post-transplant functional performance [255]. In this respect, it has also been reported, that islet viability suffers when physical, chemical or temperature stress is applied [256], also indicating that the experimental setting applied may needs improvement to address a reduction of these stress factors. Changing the experimental set up as done for Assay 2 indeed has partially improved islets response during GSIS. The non-deglycosylated control group showed a significant increase of insulin secretion with rising glucose concentration and also additional PMA treatment as it was expected initially. The batch of islets used here was also of better quality compared to the previous batch (please see 2.1.2.1) and this might explain the difference here.

Overall, the results of both assays reflect, that insulin secretion was measurable using human pancreatic islets and those seem to potentially be a suitable model to study β-cell-dependent insulin secretion in human disease.

The purpose of this study was to investigate the impact of the *PMM2* mutation and related impairment of N-glycosylation on insulin secretion. As detailed previously (1.6.1), key proteins for insulin secretion are the sulfonylurea receptor 1 (SUR1) and the ATP-dependent potassium channel KCNJ11 and previous work has shown that mutations affecting N-glycosylation of SUR1 and KCNJ11 can cause hyperinsulinism [257,258].

The treatment with deglycosylation enzymes was meant to induce deglycosylation of membrane proteins to assess the effect on insulin secretion. However, deglycosylated islets did not respond with a significant increase in insulin release in both settings (Assay 1, Assay 2) compared to the control samples not treated with PNGase F or Endo H. Either the variations within the deglycosylated sample group were too high or the non-glycosylated control behaved similarly as was the case in Assay 2, so that a distinct effect could not be observed. Although PMA did enhance insulin release in Assay 2, a significant difference between deglycosylated and non-deglycosylated sample groups was not observed and the expected effect of increased insulin release due to impaired N-glycosylation was not proven. Furthermore, instead of working with PMA concentrations within the physiological range, PMA was added in way higher concentrations within the nanomolar range to mimic GLP-1 stimulated insulin secretion. As outlined in the introduction of this thesis, differing GLP-1 concentrations stimulate insulin secretion through different pathways which might change the levels of insulin secretion. However, the reason why GLP-1 employs different pathways depending on concentration is still not completely understood [154].

It has to be noted, that the experimental set up is difficult and results therefore have to be analysed with caution. First of all, a problem occurred whilst performing Assay 1. A white precipitation was observed both times after adding the PNGase F in its enzyme buffer and also whilst adding the enzyme buffer only to the control. This was likely a result of salt formation between components of the KRBB solution and the enzyme buffer. The latter was a highly concentrated Na_3PO_4 solution which yielded a final concentration of 50 mM. It is very likely, that the phosphate ions have formed a salt with the calcium and/or magnesium ions from the KRBB solution and either $Ca_3(PO_4)_2$ and/or $Mg_3(PO_4)_2$ did precipitate. This phenomenon is explained by the Le Chatelier principle, which states that when stress (e.g. temperature-, pH-, concentration-change) moves the reaction away from equilibrium, the reaction will try to adjust to get back to equilibrium. Whether a precipitate will form depends on the relation of reaction quotient (Q) and solubility constant (K_{sp}). This can be determined by comparing Q and K_{sp} of the predicted precipitate formed: If Q is greater than K_{sp}, excess product will precipitate [259]. This is what likely has happened in our experiment. The solubility of $Ca_3(PO_4)_2$ and $Mg_3(PO_4)_2$ in aqueous solution is very low; whereas sodium-salts are highly soluble resulting in the precipitation observed due to the highly concentrated Na_3PO_4 solution.

Associated with the enzyme buffer addition, islets clusters dissociated within the precipitate and cells appeared to be swollen as observed by microscopy (data not shown), potentially altering their secretory function. Consequently, for Assay 2 the deglycosylation enzymes were added directly to the wells without the enzyme buffer this has solved the problem of precipitation. Disintegration of islets structure changes insulin secretion negatively as it is a multicellular process dependent on the interplay of many cell types rather than the performance of β-cells only. This was tested in experiments looking at the effect of islets dispersion on insulin secretion showing that reaggregation of islets resulted in a rapid improvement of insulin release [260]. The 3D architecture and cellular coordination of those “microorgans” is of vast importance, enabling communication and cross-talk between the different cell types, crucial for appropriate islets function. Besides communicating in an indirect way through, for example hormones, neurotransmitter and other signalling molecules, which do not require cells to be within close proximity, also direct cell-to-cell communication is essential for β-cells to release insulin and relies on the distinct islets 3D organisation. Many different adhesion glycoproteins called CAMs (cell adhesion molecule) maintain cell cohesion and mediate direct cell-to-cell communication between neighbouring cells within pancreatic islets [261,262]. This implies that deglycosylation of the membrane-bound cell surface proteins likely alters islets structure and consequently β-cell behaviour and response to glucose stimulation. This may explain that no distinct effect was observed in all three experiments conducted when islets were deglycosylated.

A major issue whilst conducting these assays was a large variation in insulin measurement within the sample groups. Increasing the sample number would be one option to try to overcome the problem and reduce errors. However, some technical issues related to the handling of the islets are contributing to variation, which are not easy to solve. Those variables possibly affecting experimental data are:

1) Difficulties to control equal numbers of islets seeded per well
2) Islet quality might differ across wells
3) Distribution of β-cells per islet
4) Number of β-cells per islet likely varies
5) Activity of β-cells per islet

All these factors likely have an impact on insulin measurement. Moreover, particularly in Assay 1, insulin might have been transferred from one well to the other between the different conditions. It is likely that secreted insulin accumulated within the islets cluster and simply diffused into the solution after transfer to the next well.

Some additional issues are also related to the enzyme addition itself to facilitate deglycosylation of the islets. First of all, for both assays the enzymes have not been tested for their functionality beforehand and also using different concentrations of the enzymes were not compared. Moreover, adding the enzymes effectively changes the surrounding environment the islets are situated in, which the non-glycosylated control groups do not experience, this might have a general impact on islets behaviour and physiology, which is not related to an effect of deglycosylation. This can be either inhibition but also the enhancement of insulin release as different stressors have different effects on β-cell function [263]. Some disintegration was also observed when treating the islets in Assay 2 without the addition of the enzyme buffer. As pointed out before, this does have an impact on islets and in particular β-cell function. The natural environment of β-cells is living in these cell clusters with four other endocrine cell types. If this changes, β-cells behaviour and function very likely changes as well. In addition, this also causes an error within the experimental set up as the non-deglycosylated control group did not experience such changes. Effectively the conditions both sample groups were exposed to were technically different, thus unfavourable for a proper comparison.

Pancreatic islets are quite sensitive and prone to stress in general. During the islets isolation procedure, for example elevated apoptosis and necrosis rates are measurable as studies have shown: Two major cellular stress-induced pathways, NF-κB (Nuclear Factor kappa-light-chain-enhancer of activated B cells) and PARP (poly(ADP-ribose) polymerase) are active during isolation and related signalling molecules such as interleukin-6 and -8 (proinflammatory cytokines) are still detectable days after isolation. Significant numbers of islets die post isolation, thus lowering the yield of functional islets dramatically and this still represents a major obstacle to overcome for islets transplant therapy [264].

In conclusion, glucose-stimulated insulin secretion (GSIS) overall has worked nicely, and this shows that human pancreatic islets are a possible model to investigate insulin secretion. However, due to the restricted availability of human pancreatic islets conducting appropriate experiments will remain difficult. Moreover, the experimental settings itself require significant improvement in order to generate reliable data as the settings as such are too error prone and this likely affects insulin measurement (see 8.3).

Chapter 6: Development of the HIPKD HepG2 cell model with CRISPR-Cas9

6.1 Introduction

This chapter describes the generation of an HIPKD cell model, which harbours the disease-causing mutation (c.-167G>T) in homozygous state. The HepG2 cell line from ATCC (ATCC® HB-8065™) was chosen to be directly targeted at the *PMM2* promoter with CRISPR-Cas9 and the c.-167G>T SNP aimed being integrated by providing a ssODN (homologous DNA repair template) allowing the cell's naturally occurring DNA repair mechanism to follow the HDR pathway after Cas9 cleavage. HepG2 cells are derived from a hepatocellular carcinoma of a 15-year-old male. HepG2 is an immortalised, non-tumorigenic cell line with epithelial morphology and a modal chromosome number of 55 (hyperdiploid) [265,266].

Two reasons explain why HepG2 cells were chosen for this purpose: 1) Based on the previously described bioinformatics analysis characterising the human *PMM2* promoter region more precisely, identifying a chromatin loop and showing the chromatin segmentation (Figure 35) in different commercially available cell lines, HepG2 cells stood out being the only cell type in which the chromatin seemed to be active throughout the looping area. 2) As HepG2 cells are a liver-derived cell line and as liver involvement in HIPKD has been described [209] HepG2 cells were deemed a suitable disease model for HIPKD.

In order to achieve scarless genome editing, the CRISPR-Cas9 system was used to generate the c.-167G>T *PMM2* promoter SNP in HepG2 cells. Different formats of the CRISPR-Cas9 system and ways to deliver it to cells are commercially available. Those can be DNA expression plasmids, *in vitro* mRNA transcripts or ribonucleoprotein particles (RNPs) where the recombinant Cas9 protein and the specific sgRNA build a complex [267]. For the purposes of this study, a DNA plasmid-based CRISPR-Cas9 expression system was used as a starting point and delivered to the cells together with the ssODN using a reagent-based transfection system. The advantage here was seen in the nature of the CRISPR construct chosen, having a GFP expression gene integrated as an additional feature enabling microscopical tracking of the transfection. The optimisation of the transfection conditions for HepG2 cells turned out to be quite elaborate and time-consuming, therefore fast-growing and known to be easy to transfect HEK293T cells (human embryonic kidney cells 293, ATCC® CRL-3216™) were recruited to initially test the efficiency of the two differently designed CRISPR-Cas9 expression plasmids (6.3.5) to find out which one was the best to use. HEK293T cells were only used for the purpose of assessing the overall efficiency of the two CRISPR constructs and were not taken further in the experimental setting, e.g. no single cell colonies were obtained from these cells. After the construct-assessment, CRISPR experiments were carried out using HepG2 cells exclusively.

Although sufficient CRISPR activity was measured repetitively in HepG2 cells, this plasmid-based approach did eventually not produce any homozygous HepG2 clone after performing several experiments and large-scale genotyping of single cell colonies. As a consequence, a system switch to CRISPR-RNPs was commenced, which were shuttled into HepG2 cells by electroporation and again in co-transfer with the ssODN. The results and outcomes of both attempts to generate a homozygous HIPKD cell model are outlined in this chapter.

6.2 The aims of this chapter

- Design of a CRISPR gRNA which targets the *PMM2* promoter with high efficiency and only +/- 10 bp up- or downstream to the nucleotide position of the desired SNP (c.-167G>T)
- Assessment of cutting efficiency of CRISPR gRNA using TIDE
- Co-transfection of CRISPR gRNA, Cas9 and ssODN for homology directed repair
- Development of stable HIPKD cell model using HepG2 harbouring the desired SNP in homozygous state

6.3 Results

6.3.1 Verification of *PMM2* promoter region of interest in HEK293T and HepG2 cells

In order to guarantee that the *PMM2* promoter region of interest in HepG2 and HEK293T cells harbours no SNP or other alterations to the wt sequence, gDNA was extracted and the *PMM2* amplicon (Primers: *PMM2* For3+Rev3,Table 8) sent for sequence verification. Figure 43 shows a representative example of the band resulting from amplification of gDNA from HepG2 cells on a 2% agarose gel.

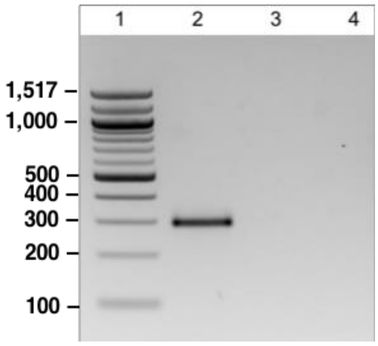
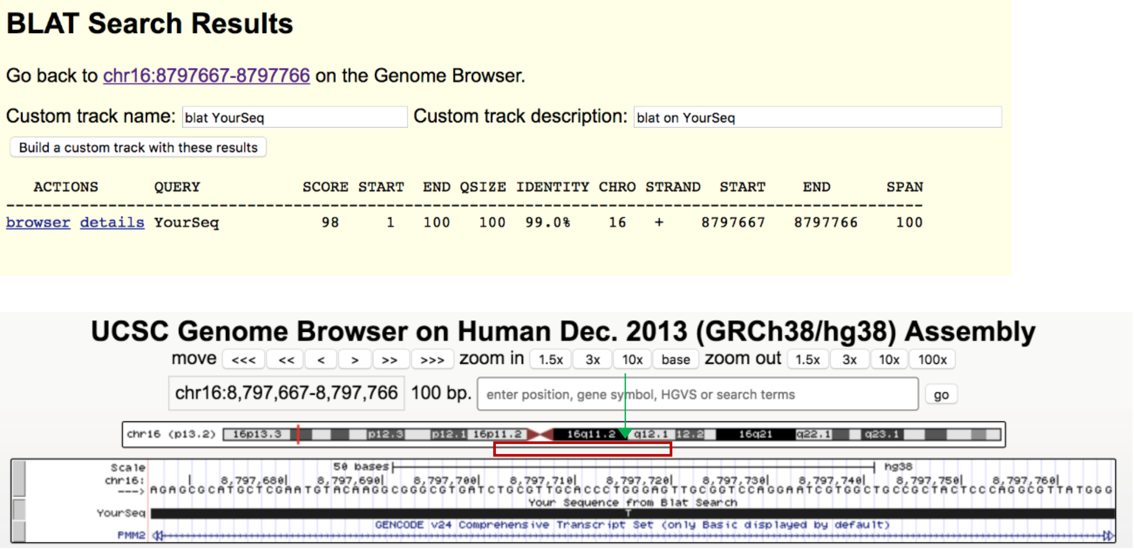


Figure 43: Representative gel image of PCR amplicon with primer pair *PMM2* For3/Rev3. gDNA extracted from wt HepG2 cells. Lane 1: 100 bp Ladder (Figure 75). Lane 2: PCR product *PMM2* For3/Rev3 (296 bp). Lane 3: Empty master mix, Blank control H_2O only.

To demonstrate what the DNA sequence of the *PMM2* region of interest in hg38 (Human genome version 38) is like, Figure 44 illustrates a BLAT (BLAST-like alignment tool) search result done within the UCSC genome browser (https://genome.ucsc.edu/). The section shows a 100 bp stretch of the amplicon with nucleotide -167 in the centre. Within the input sequence for the BLAT analysis the "G" at nt -167 was replaced by the desired SNP ("T") to simplify localisation of the CRISPR editing site.

Input sequence used for BLAT analysis:

AGAGCGCATGCTCGAATGTACAAGGCGGGCGTGATCTGCGTTGCACCCTTGGAG
TTGCGGTCCAGGAATCGTGGCTGCCGCTACTCCCAGGCGTTATGGG



BLAT Search Results

Go back to chr16:8797667-8797766 on the Genome Browser.

Custom track name: blat YourSeq   Custom track description: blat on YourSeq

```
   ACTIONS       QUERY             SCORE START  END QSIZE IDENTITY CHRO STRAND  START     END       SPAN
----------------------------------------------------------------------------------------------------
browser details YourSeq              98     1   100   100   99.0%    16    +   8797667   8797766    100
```

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

chr16:8,797,667-8,797,766 100 bp.

Figure 44: UCSC BLAT result (hg38) showing the genomic ROI within the *PMM2* promoter with nt position c.-167G in the centre.
Top: The 100 bp sequence matches the leading strand on chromosome 16 with 99% identity. Bottom: The first track shows the genomic coordinates and the base composition of the of the BLATed input sequence on chromosome 16. The red box highlights the transcription factor binding site for ZNF143 which includes nt position c.-167G marked by the green arrow. The second track shows the input sequence from BLAT search with the SNP (c.167G>T). The third track shows the GENECODE version 24 track indicating *PMM2*.

The *PMM2* amplicons from both, wt HepG2 and HEK293T gDNA, were sent for unidirectional Sanger sequencing using the *PMM2* For3 primer. Both sequencing reactions match the UCSC hg38 *PMM2* sequence shown above to 100%. Figure 45 shows a section of the sequenced *PMM2* promoter region in both cell lines.

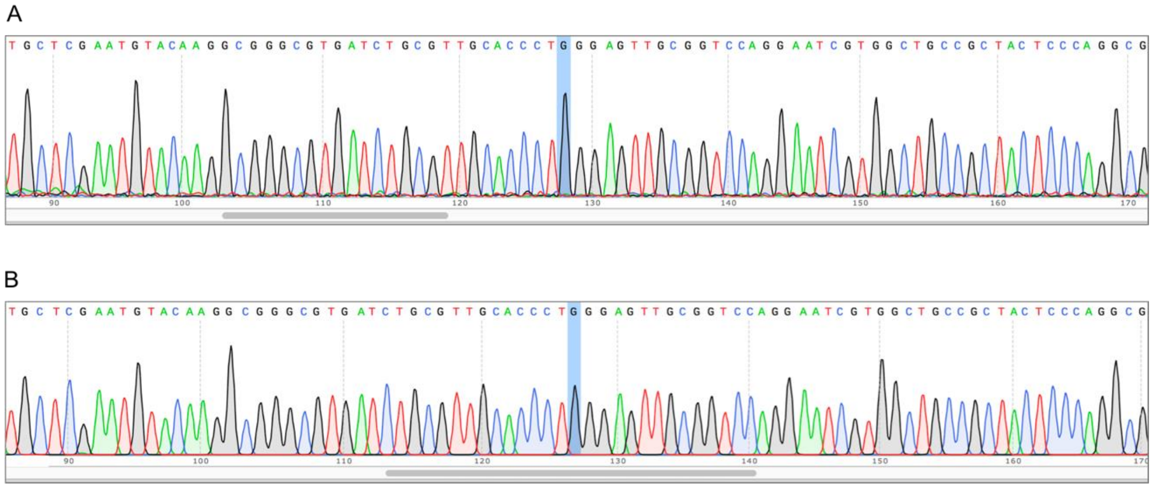


Figure 45: Sanger sequencing of *PMM2* For3/Rev3 amplicon with primer *PMM2* For3 in wt HEK293T and HepG2 cells.
The CRISPR site (c.-167G) is highlighted in light blue, flanked by about 40 nt either side. A: Electropherogram showing the wt sequence of the *PMM2* ROI in HEK293T cells. B: Electropherogram showing the wt sequence of the *PMM2* ROI in HepG2 cells.

6.3.2 Finding the right gRNA for a successful gene edit

For a CRISPR HDR gene edit it is crucial to choose gRNA target regions which are as close to the CRISPR site as possible. To meet this requirement, gRNA target sequences were chosen directly from the DNA sequence around nt position c.-167G. The sequence was pasted into SnapGene® and screened for the closest PAM motifs to the CRISPR site. Two gRNA target regions were chosen to start with. Both are flanking the CRISPR site where the SNP is planned to be introduced in ultimate proximity. One of them, hPMM2_1, is located on the leading strand, the other one, hPMM2_2, on the reverse strand of the *PMM2* promoter. Figure 46 shows the genomic region of the *PMM2* promoter region of interest generated with SnapGene®. Primer sequences, CRISPR site, gRNA sequences with corresponding PAM sites and the transcription start site within exon 1 are annotated.

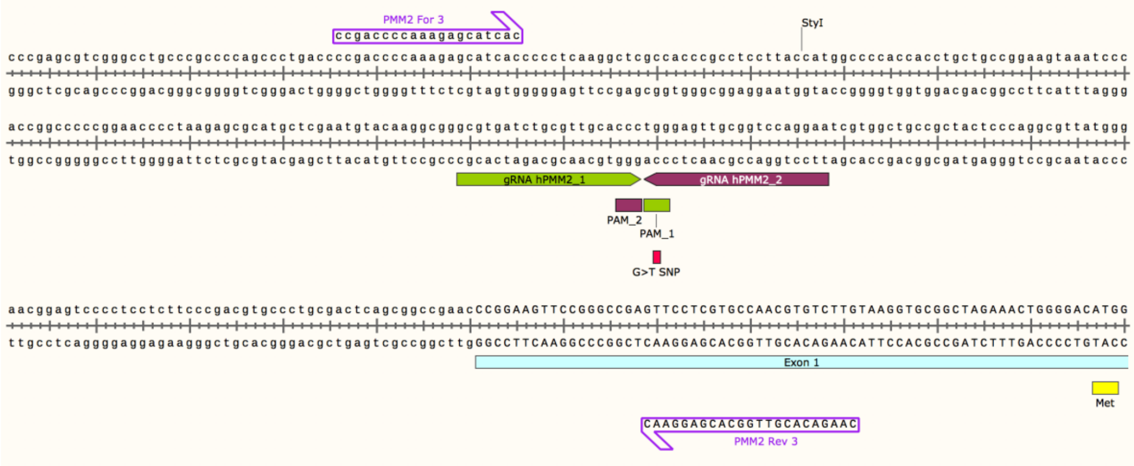


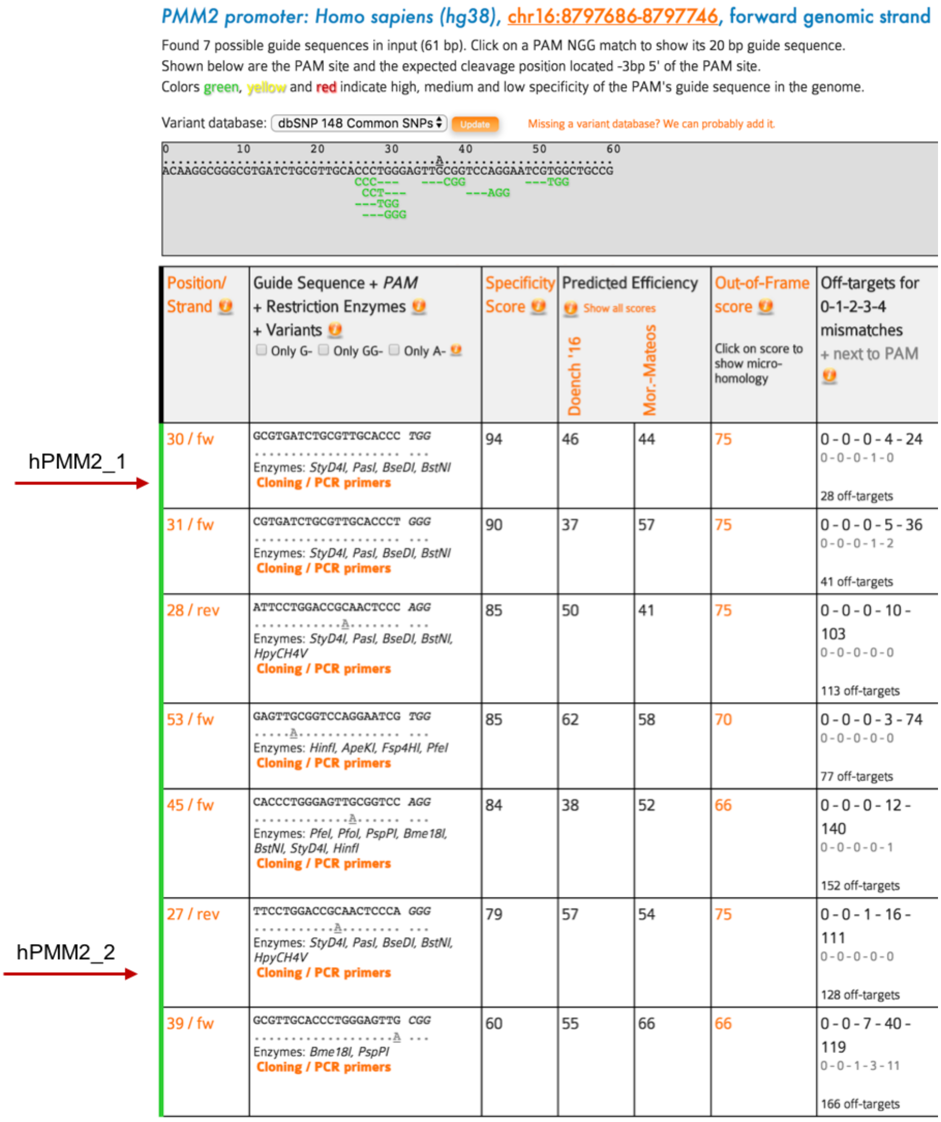
Figure 46: SnapGene® sequence showing the CRISPR region within the *PMM2* promoter. Purple arrows: *PMM2* For3 and Rev3 primer pair and their binding sites limiting a 296 bp DNA section. Green arrow: gRNA hPMM2_1 binding site and target sequence. Green box: PAM motif for gRNA hPMM2_1. Magenta arrow: gRNA hPMM2_2 binding site and target sequence. Magenta box: PAM motif for gRNA hPMM2_2. Red box: Nucleotide position c.-167G to be edited to c.-167G>T. Turquoise area: Beginning of exon 1 of *PMM2*. Yellow box: Start codon methionine of *PMM2* coding region. Label StyI: StyI restriction site.

The gRNA target region (please see below) was put in and analysed with two independent frequently used CRISPR design tools, CRISPOR and CHOPCHOP v2, to assess uniqueness and specificity of the two gRNAs of choice. For both analyses, human genome hg38 was selected and the PAM motif was set to NGG referring to the specific PAM site for the Cas9 nuclease originated from *Streptococcus pyogenes*.

The input sequence used to analyse the quality of the chosen gRNA targets was the following:

ACAAGGCGGGCGTGATCTGCGTTGCACCCTGGGAGTTGCGGTCCAGGAATCGTGGCTGCCG

CRISPOR lists and ranks resulting gRNA target sequences in a table (Figure 47) according to their specificity from high to low. The CRISPOR scoring system ranges from 0-100. In total seven gRNA targets were found based on the DNA sequence input resulting in hPMM2_1 being the best match on rank 1 and gRNA hPMM2_2 being ranked number six. However, both are annotated as being good gRNA targets indicated by the green mark on the left implying specificity scores of above 50. The score was 94 for hPMM2_1 and 79 for hPMM2_2. The efficiency scores predicting the cutting efficiency of the guides were higher for hPMM2_2, with 57 and 54 according to the algorithms from Doensch 2016 [268] and Moreno-Mateos [269], respectively. For hPMM2_1, these were 46 and 44. The chance of frameshift deletions is indicated by the out-of-frame score, which were the same for both gRNA targets with a score of 75. Nevertheless, since a homology directed repair is desired for the gene edit, this is not of much importance here. In terms of possible off-targets, guide hPMM2_1 clearly seems to be the better option with 28 off-targets opposed to 128 off targets with guide hPMM2_2 and this is also confirmed by the ranking. Out of all possible off-targets there is one with three mismatches which is within the seed region of the gRNA target of hPMM2_1. For hPMM2_2 none of the predicted off-targets falls within the seed region.



PMM2 promoter: Homo sapiens (hg38), chr16:8797686-8797746, forward genomic strand

Found 7 possible guide sequences in input (61 bp). Click on a PAM NGG match to show its 20 bp guide sequence.
Shown below are the PAM site and the expected cleavage position located -3bp 5' of the PAM site.
Colors green, yellow and red indicate high, medium and low specificity of the PAM's guide sequence in the genome.

Variant database: dbSNP 148 Common SNPs Missing a variant database? We can probably add it.

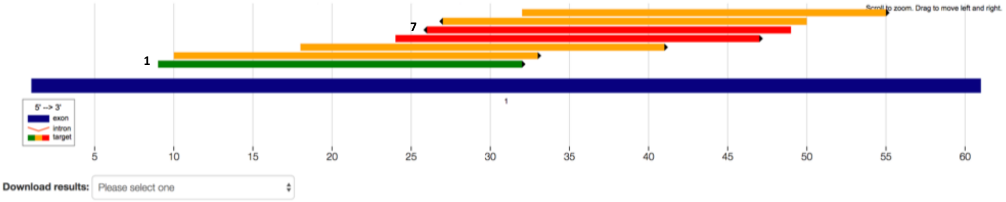
```
0         10        20        30        40        50        60
.....................................A.......................
ACAAGGCGGGCGTGATCTGCGTTGCACCCTGGGAGTTGCGGTCCAGGAATCGTGGCTGCCG
                          CCC---   ---CGG        ---TGG
                           CCT---       ---AGG
                          ---TGG
                           ---GGG
```

Position/ Strand	Guide Sequence + *PAM* + Restriction Enzymes + Variants ☐ Only G- ☐ Only GG- ☐ Only A-	Specificity Score	Predicted Efficiency Show all scores Doench '16	 Mor.-Mateos	Out-of-Frame score Click on score to show micro-homology	Off-targets for 0-1-2-3-4 mismatches + next to PAM
30 / fw	GCGTGATCTGCGTTGCACCC *TGG* Enzymes: *StyD4I, PasI, BseDI, BstNI* **Cloning / PCR primers**	94	46	44	75	0 - 0 - 0 - 4 - 24 0 - 0 - 0 - 1 - 0 28 off-targets
31 / fw	CGTGATCTGCGTTGCACCCT *GGG* Enzymes: *StyD4I, PasI, BseDI, BstNI* **Cloning / PCR primers**	90	37	57	75	0 - 0 - 0 - 5 - 36 0 - 0 - 0 - 1 - 2 41 off-targets
28 / rev	ATTCCTGGACCGCAACTCCC *AGG* ....A....... Enzymes: *StyD4I, PasI, BseDI, BstNI, HpyCH4V* **Cloning / PCR primers**	85	50	41	75	0 - 0 - 0 - 10 - 103 0 - 0 - 0 - 0 - 0 113 off-targets
53 / fw	GAGTTGCGGTCCAGGAATCG *TGG* .....A.............. Enzymes: *HinfI, ApeKI, Fsp4HI, PfeI* **Cloning / PCR primers**	85	62	58	70	0 - 0 - 0 - 3 - 74 0 - 0 - 0 - 0 - 0 77 off-targets
45 / fw	CACCCTGGGAGTTGCGGTCC *AGG* ....A...... Enzymes: *PfeI, PfoI, PspPI, Bme18I, BstNI, StyD4I, HinfI* **Cloning / PCR primers**	84	38	52	66	0 - 0 - 0 - 12 - 140 0 - 0 - 0 - 0 - 1 152 off-targets
27 / rev	TTCCTGGACCGCAACTCCCA *GGG* ....A........ Enzymes: *StyD4I, PasI, BseDI, BstNI, HpyCH4V* **Cloning / PCR primers**	79	57	54	75	0 - 0 - 1 - 16 - 111 0 - 0 - 0 - 0 - 0 128 off-targets
39 / fw	GCGTTGCACCCTGGGAGTTG *CGG* ....A Enzymes: *Bme18I, PspPI* **Cloning / PCR primers**	60	55	66	66	0 - 0 - 7 - 40 - 119 0 - 0 - 1 - 3 - 11 166 off-targets

Figure 47: CRISPOR result showing seven gRNAs to target the desired *PMM2* CRISPR site. The table is sorted by specificity score from high to low. Red arrows annotate the chosen gRNAs hPMM2_1 (Rank 1) and hPMM2_2 (Rank 6). Green bar on the left of each guide: Based on specificity score. Green indicates a good specificity score of above 50. Column 1: Position of the PAM on the input sequence and the strand. Column 2: gRNA target sequence with PAM, variants, restriction sites which overlap the three base pairs 5' of the PAM, instruction for cloning and PCR primer. Column 3: Specificity score. Column 4: Efficiency score predicting cutting efficiency. Column 5: Out-of-Frame score. Colum 6: Off-target mismatch counts to up to 4 mismatches and mismatch count per off-target falling into seed region. Scores range from 0-100.

Figure 48 shows the result of the CHOPCHOP analysis, also indicating guide hPMM2_1 as the best choice (green) out of in total seven possible guides for the region of interest, with no off-targets detected. The scoring system in CHOPCHOP ranges from 0-1. Differently to CRISPOR, guide hPMM2_2 was categorised as a bad guide ranked on last position (rank 7). In terms of its specificity it was marked red (red) as shown in Figure 48 A.

A



Ranking	Target sequence	Genomic location	Exon	Strand	GC (%)	Self-complementarity	Off-targets				Efficiency
							0	1	2	3	
1	GCGTGATCTGCGTTGCACCCTGG	sequence:10	1	+	65	0	1	0	0	0	0.48
2	GAGTTGCGGTCCAGGAATCGTGG	sequence:33	1	+	61	0	1	0	0	1	0.56
3	ATTCCTGGACCGCAACTCCCAGG	sequence:28	1	-	61	2	1	0	0	1	0.52
4	CGTGATCTGCGTTGCACCCTGGG	sequence:11	1	+	65	0	1	0	0	1	0.48
5	GCGTTGCACCCTGGGAGTTGCGG	sequence:19	1	+	70	0	1	0	1	0	0.51
6	CACCCTGGGAGTTGCGGTCCAGG	sequence:25	1	+	70	0	1	0	0	4	0.43
7	TTCCTGGACCGCAACTCCCAGGG	sequence:27	1	-	65	1	1	0	0	9	0.58

B

Gene specific part of sgRNA

GCGTGATCTGCGTTGCACCCTGG

Off-targets

Genomic location	Number of mismatches	Sequence (including mismatches)
chr16:8797695	0	GCGTGATCTGCGTTGCACCCTGG

C

Gene specific part of sgRNA

TTCCTGGACCGCAACTCCCAGGG

Off-targets

Genomic location	Number of mismatches	Sequence (including mismatches)
chr13:18250109	3	CCCTGGGAcTTGCtGTCCtGGAA
chr16:8797712	0	CCCTGGGAGTTGCGGTCCAGGAA
chr20:23981776	3	CCCTGGGAcTTGCtGTCCtGGAA
chr22:18349290	3	TTCCaGGACaGCAAgTCCCAGGG

Figure 48: CHOPCHOP result showing seven gRNAs to target the desired *PMM2* CRISPR site. A Top: Dynamic visualisation displaying all target site options with a colour code according to the scoring indicating specificity. Green: Best. Yellow: Okay. Red: Bad. Chosen gRNAs are numbered according to ranks in table (A Bottom). A Bottom: Column 1 shows the ranking of the gRNA listed. gRNA hPMM2_1 is on rank 1 and gRNA hPMM2_2 on Rank 7. Column 2: gRNA target sequence including PAM. Column 3*: Genomic location. Column 4*: Exon where target sequence is located. Column 5: Strand. Column 6: Self-complementarity within the gRNA and between gRNA and RNA backbone. Column 7: Off-target mismatch counts to up to 3 mismatches. Column 8: Efficiency score (0-1). B: Gene specific part for gRNA hPMM2_1 indicating genomic location, number of mismatches and sequence of possible off-targets including the mismatches. C: Gene specific part for gRNA hPMM2_2 indicating genomic location, number of mismatches and sequence of possible off-targets including the mismatches.

*Here neglectable as CHOPCHOP algorithm is not able to recognise where the input sequence originates from when pasted manually.

Although revealing a higher efficiency with a score of 0.58 opposed to 0.48 for hPMM2_1. The self-complementarity accounted zero for hPMM2_1, whereas this was accounted one for hPMM2_2. Finally, nine off-targets with 3 mismatch counts each were predicted for hPMM2_2 making it the least specific guide RNA target out of all guides listed (Figure 48 A and C). The GC content for both were within the range (40% - 70%) with 65% each. In conclusion, both CRISPR design tools predict hPMM2_1 being the better gRNA for the purposes of this study. Although CRISPOR indicated both as reasonable guides to use, CHOPCHOP clearly indicated that hPMM2_2 does not meet the requirements for a good guide.

6.3.3 From transfection/electroporation to the living cell line

The following workflow diagram shall provide an overview of the major steps involved when the HepG2 cell line was generated with CRISPR-Cas9, either plasmid- or RNP-based, starting from delivering the system to the cells with reagent-based transfection or electroporation until positive clones were available to be expanded.

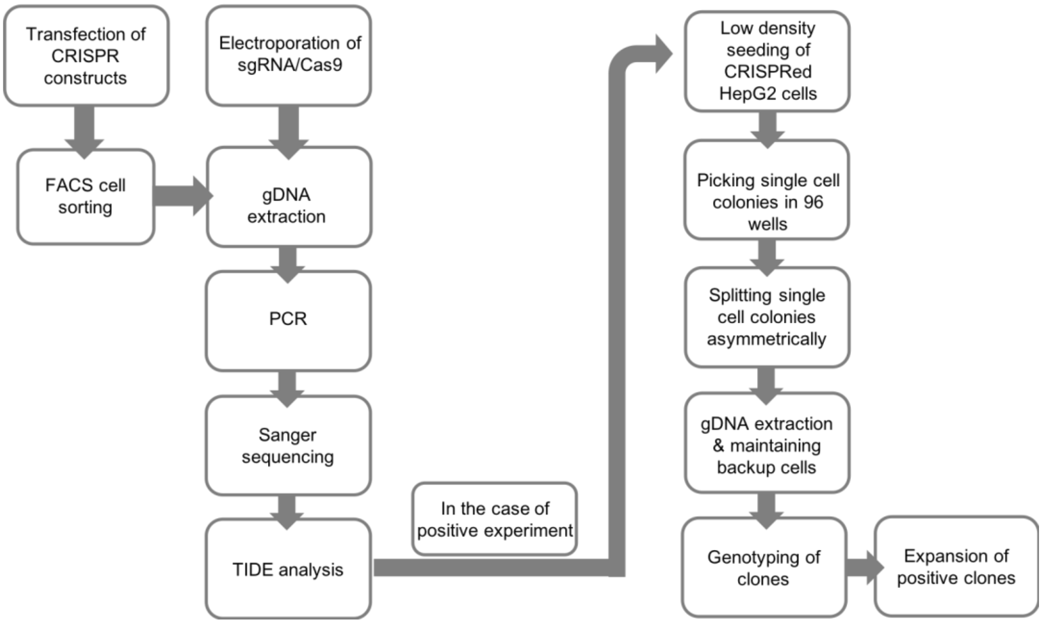


Figure 49: Workflow of a full CRISPR experiment.
The process starting from delivery of the CRISPR-Cas9 system to the cells with either transfection or electroporation to expansion of positive clones is depicted.

Please note, that all reagent-based transfections were performed as semi-reverse co-transfections whilst the cells were in suspension. A traditional transfection of HepG2 cells would be performed when the cells are adherent on the well surface. In the following paragraphs the term transfection or co-transfection, rather than semi-reverse transfections or semi-reverse co-transfections will be used for simplistic reasons.

6.3.4 Identification of intact CRISPR constructs

In total, three transformed *E. coli* DH5α clones per CRISPR construct were used for plasmid preparation. To verify whether the cloning was successful, both purified CRISPR constructs were tested in a double restriction digest with *BbsI* and *EcoRV* and analysed on a 1% agarose gel (Figure 50). Either one (10 kb) or two bands (3 and 7 kb) were expected, as successful ligation of the guide oligos removed the *BbsI* restriction site in PX458.

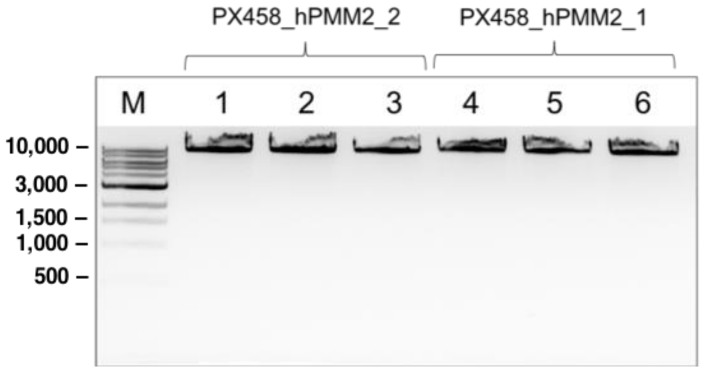


Figure 50: *BbsI* and *EcoRV* restriction digest to validate CRISPR constructs after cloning. Constructs: PX458_hPMM2_1 and PX458_hPMM2_2. Lane 1: 1 kb ladder (Figure 76). Lanes 2-4: Purified plasmid PX458_hPMM2_2 from three *E. coli* DH5α clones. Lanes 5-7: Purified plasmid PX458_hPMM2_1 from three *E. coli* DH5α clones. Sample loaded and run on a 1% agarose gel.

Restriction digest of purified plasmid all show a linearised band of about 10 kb, indicating successful cloning of the guide oligos into the target vector PX458. The smeary appearance of the bands could be due to nicks in the plasmids which are changing the supercoiled structure of the construct. DNA concentration of each CRISPR construct was assessed using the NanoDrop™ and those with the highest concentration were sent for sequence verification (U6 Forward primer) to confirm the results of the digest.

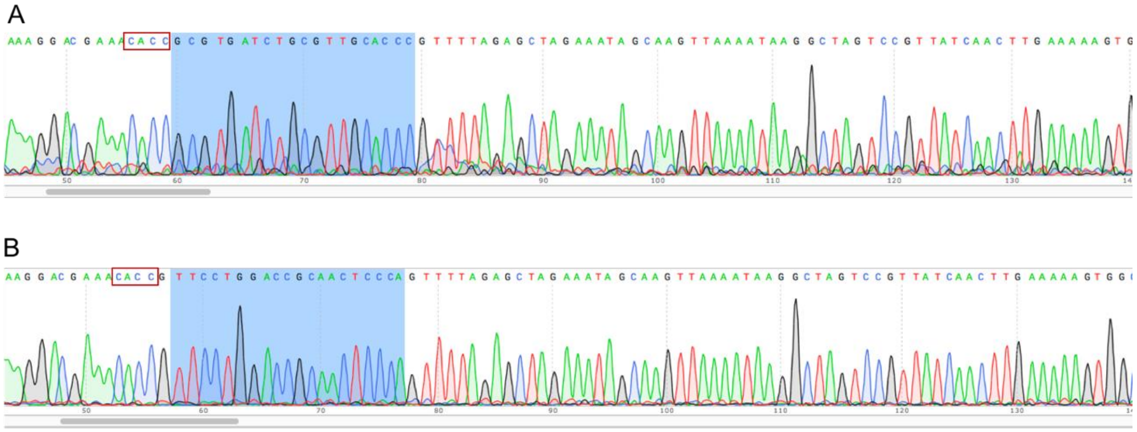


Figure 51: Sanger sequencing of CRISPR constructs.
A: PX458_hPMM2_1. B: PX458hPMM2_2. Highlighted in blue are the gRNA target sequences cloned into the CRISPR expression vector PX458. Red box: 5'-overhangs of guide oligos complementary to *BbsI* cut sites. Guanine added in front of the gRNA target sequence in B to initiate transcription.

In both cases the gRNA target sequence was inserted and the two CRISPR constructs ready to use for transfection.

6.3.5 Testing the efficiency of CRISPR-Cas9 constructs in HEK293T cells

The first step for setting up the CRISPR-HDR experiment was, 1) to find out which of the two designed CRISPR constructs works best eventually and 2) under which conditions. A first test to find out how efficient or inefficient the two designed CRISPR constructs were, was done with HEK293T cells, simply because this cell line is an easy to transfect cell line and the transfection protocol was already optimised, whilst transfection optimisation with HepG2 cells was still ongoing.

The sequence verified GFP-expressing CRISPR constructs (PX458_hPMM2_1 and PX458_hPMM2_2) were separately delivered to HEK293T cell using the FuGENE HD transfection reagent (2.1.3.1) to assess the cutting efficiency of the CRISPR-Cas9 system with the two chosen gRNAs. To be able to estimate the proportion of transfected cells, the same picture was taken twice with two different channels. One was simply illuminated with white light to obtain the brightfield image of the cells to be able to see the total number of cells within the selected area. The second one was shot through the green fluorescence channel of the cell imager to visualise the green glowing cells due to the expressed GFP from the CRISPR constructs once they have reached the nucleus.

The GFP fluorescence signal was analysed after 24 hours of transfection and pictures were taken as described (Figure 52).

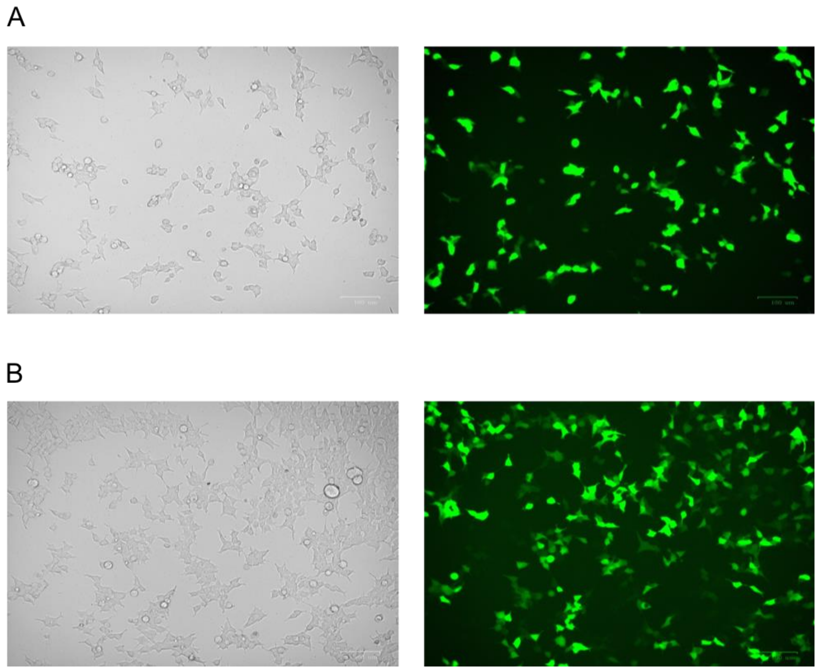
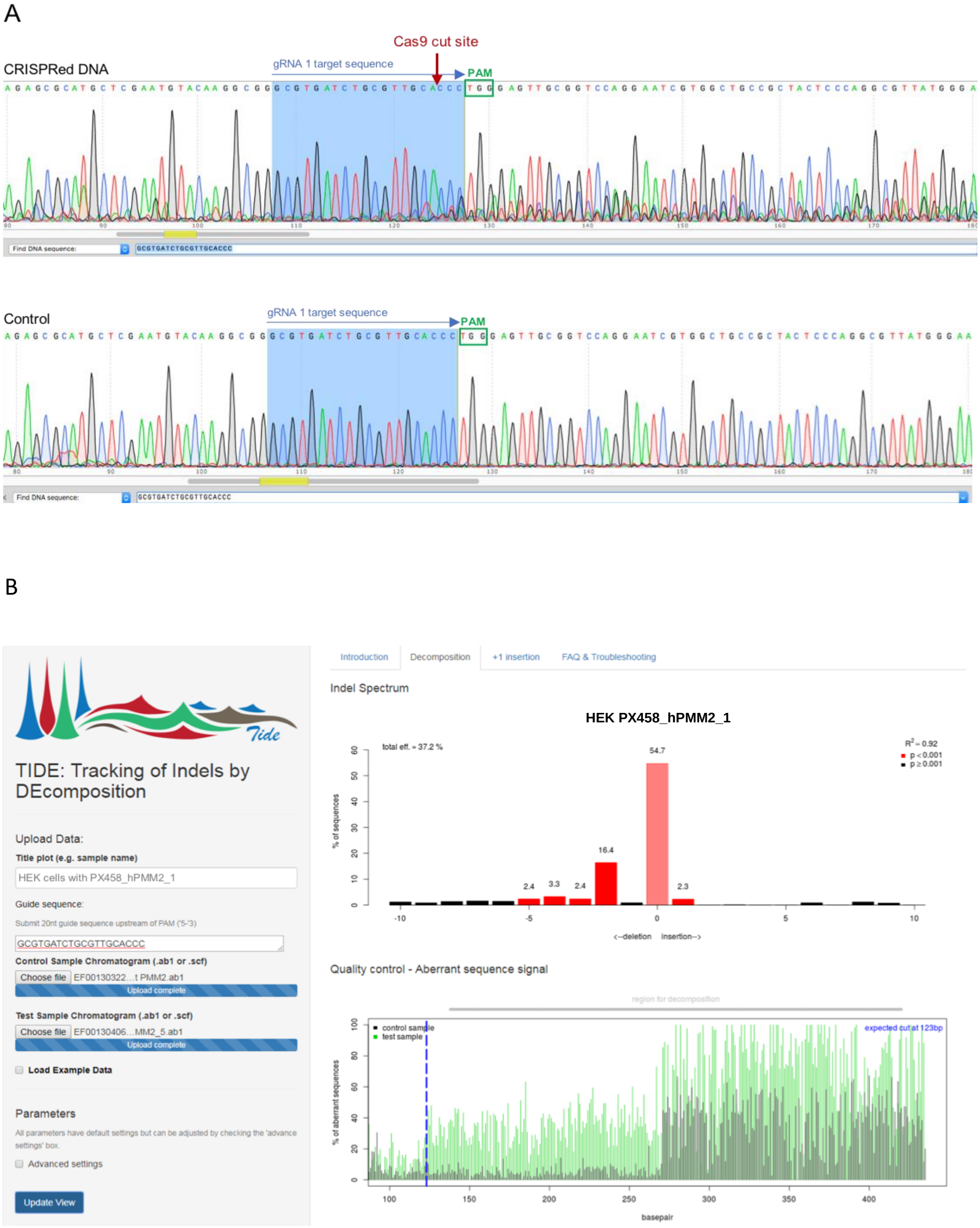


Figure 52: Semi-reverse transfection of HEK293T cells with 550 ng each of the GFP-expressing CRISPR constructs.
A left panel: Brightfield image of PX458_hPMM2_1 transfected HEK293T cells 24 hours after transfection. A right panel: Fluorescence GFP signal 24 hours post-transfection with PX458_hPMM2_1 in HEK293T cells. B left panel: Brightfield image of HEK293T cells shot 24 hours after transfection with PX458_hPMM2_2. B right panel: Fluorescence GFP signal 24 hours post-transfection with PX458_hPMM2_2 in HEK293T cells. Pictures were taken with the ZOE cell imager. Scale bar (bottom right corner): 100 nM.

In both cases, the transfection has worked, as both transfected cell population show a decent fluorescence signal. Comparing the brightfield images and the corresponding green fluorescence pictures it seems that the majority of HEK293T cells have been successfully transfected with the two CRISPR constructs.

To properly investigate this further, cells were harvested and lysed ON after 48 of transfection and gDNA was extracted the following day (2.4.1.2) and used to analyse the CRISPR site on the genomic level. The gDNA from both differently transfected cell populations was amplified as described previously (2.4.1.3) using the *PMM2* For3/Rev3 primer pair specific for the *PMM2* promoter region of interest. The two figures (Figure 53, Figure 54) below display the electropherograms from Sanger sequencing and the following TIDE analysis to quantify the cutting efficiency of the two CRISPR constructs used, each carrying one of the particular gRNAs (hPMM2_1 and hPMM2_2).

The results of transfected cells with PX458_hPMM2_1 are shown in Figure 53. In A, the sequencing traces of the CRISPRed cells compared to a clean wildtype control are shown, both depicting the region of interest within the *PMM2* promoter highlighting the CRISPR site (blue). Whether or not the gRNA target sequence can still be found within the sequence after gene editing the cells, provides a first hint to what extent the CRISPR-Cas9 system has changed the DNA region of interest. In the case of gRNA hPMM2_1, the target sequence was still fully detectable, indicating that the wt sequences are dominating the DNA pool obtained from the transfected cell population. However, after the Cas9 cut site the sequence is clearly shifted to the left causing a “messy” looking appearance of the traces. This indicates indel formation within the DNA pool and confirms that the designed CRISPR construct and particularly the gRNA included, have worked. To measure how efficient the system has worked was analysed in TIDE (Figure 53 B). Both sequencing traces shown in A were put in the software with the gRNA hPMM2_1 target sequence as a reference and differences were quantified. The top graph in B reflects the Indel spectrum after alignment of the two sequences in percent. Zero means no deletion or insertion detected, which accounted in this experiment for 54.7% of the edited sequence. To the left are the percentages of deletions of up to 10 nt and to the right are the percentages of insertions of up to 10 nt. The total cutting efficiency of this experiment was decent with 37.2%. The indel spectrum shows that mainly two nucleotide deletions (16.4%) have happened during CRISPR. Some deletions of three (2.4%), four (3.3%) and five nucleotides (2.4%) were detected, as well as some bigger deletions with five to ten nucleotides were identified being deleted. Only a few insertions could be indicated. Mainly insertions of only one nucleotide (2.3%) at a time were spotted and a few insertions with six, eight and nine nucleotides were found being inserted during the experiment. The bottom graph in B shows the aberrant sequence signal of the two input sequences. Black reflects the wt sequence and green the sample sequence. And as seen in the electropherogram right after the cut site, the sequences differ significantly from each other. The increasing peaks after about 275 bp can be ignored as they reflect noise since the full amplicon sequenced only has a size of 296 bp.



TIDE: Tracking of Indels by DEcomposition

Upload Data:

Title plot (e.g. sample name)
HEK cells with PX458_hPMM2_1

Guide sequence:
Submit 20nt guide sequence upstream of PAM ('5-'3)
GCGTGATCTGCGTTGCACCC

Control Sample Chromatogram (.ab1 or .scf)
EF00130322…t PMM2.ab1

Test Sample Chromatogram (.ab1 or .scf)
EF00130406…MM2_5.ab1

☐ **Load Example Data**

Parameters

All parameters have default settings but can be adjusted by checking the 'advance settings' box.

☐ Advanced settings

Figure 53: Sanger sequencing and TIDE analysis of CRISPR region of edited HEK293T cells. Pooled HEK293T gDNA edited with hPMM2_1 compared to a control (wt HEK293T pooled gDNA). A: Sanger sequencing. Highlighted in blue are the gRNA target sequences. Green boxes mark the PAM sites. Red arrow indicates the Cas9 cut sites. Both sequencing traces shown are in forward direction previously amplified with primer *PMM2* For3/Rev3 and sequenced with primer *PMM2* For3. A Top: Electropherogram of sequenced HEK293T gDNA amplicon after transfection with hPMM2_1. A bottom: wt electropherogram of *PMM2* promoter ROI. B: TIDE analysis comparing electropherograms from A. gRNA sequence hPMM2_1 was provided as a reference string. B Top: Indel spectrum and calculated cutting efficiency in percent. B Bottom: Aberrant sequence signal of the two input sequence traces and the expected cut site. Black: wt sequence. Green: Sample sequence. Blue: Expected cut site.

The electropherograms in Figure 54 A reflect the result after editing the cells with CRISPR construct PX458_hPMM2_2 compared to the clean control wt sequence. Apart from the first 5 bases, the gRNA target sequence cannot be detected anymore. This indicates significant changes of the pooled DNA sequence after CRISPR editing these cells. The sequence shift to the right after the Cas9 cut site clearly demonstrates indel formation within the DNA pool due to CRISPR and confirms that gRNA hPMM2_2 is a good candidate to target the PMM2 promoter region of interest.

Quantification of the CRISPR efficiency (Figure 54 B) when the cells were transfected with construct PX458_hPMM2_2, TIDE measured an overall cutting efficiency of 47.9%. Similarly, to guide hPMM2_1, mainly deletions of one nucleotide (21.7%) took place during CRISPR altering the input DNA from the HEK293T cell population. Overall, the edited sequence matched the control wt sequence to only 46.5%. Further deletions of two to ten nucleotides were measured with percentages ranging from 5.2% (two nucleotide deletions) to below 1.6% (three, four, five, six and ten nucleotide deletions). Deletions of seven, eight and nine nucleotides ranged from 1.6% to 3.6%. Only 3.3% of one nucleotide insertions were detected and some insertions of two, eight and ten nucleotides with very low percentages were found. The aberrant sequence signal clearly matched with what was already reflected by the two electropherograms, showing a significant difference in the sequencing traces after the Cas9 cut site indicating indel formation. This is demonstrated in the bottom graph by the high and low peaks in green and black, which reconstruct the CRISPRed DNA and wt sequences, respectively. As said previously, the noisy peaks after about 275 bp in the diagram are irrelevant and can be ignored.

TIDE: Tracking of Indels by DEcomposition

Upload Data:

Title plot (e.g. sample name)

HEK cells with PX458_hPMM2_2

Guide sequence:

Submit 20nt guide sequence upstream of PAM ('5-'3)

TTCCTGGACCGCAACTCCCA

Control Sample Chromatogram (.ab1 or .scf)

Choose file EF00130322…t PMM2.ab1

Upload complete

Test Sample Chromatogram (.ab1 or .scf)

Choose file EF00130309…MM2_A.ab1

Upload complete

☐ **Load Example Data**

Parameters

All parameters have default settings but can be adjusted by checking the 'advance settings' box.

☐ Advanced settings

Figure 54: Sanger sequencing and TIDE analysis of CRISPR region of edited HEK293T cells. Pooled HEK293T gDNA edited with PX458_hPMM2_2 compared to a control (wt HEK293T pooled gDNA). A: Sanger sequencing. Highlighted in blue are the gRNA target sequences if possible. Green boxes mark the PAM sites. Red arrow indicates the Cas9 cut sites. Both sequencing traces shown are in forward direction previously amplified with primer *PMM2* For3/Rev3 and sequenced with primer *PMM2* For3. A Top: Electropherogram of sequenced HEK293T gDNA amplicon after transfection with PX458_hPMM2_2. A bottom: wt electropherogram of *PMM2* promoter ROI. B: TIDE analysis comparing electropherograms from A. gRNA sequence hPMM2_2 was provided as a reference string. B Top: Indel spectrum and calculated cutting efficiency in percent. B Bottom: Aberrant sequence signal of the two input sequence traces and the expected cut site. Black: wt sequence. Green: Sample sequence. Blue: Expected cut site.

In conclusion, the CRISPR-Cas9 system seems to work better with gRNA hPMM2_2 in HEK293T cells, as the cutting efficiency after TIDE analysis was 10% higher, with 47.9% compared to the construct carrying gRNA hPMM2_1, with 37.2%. Hence, all further CRISPR transfections and co-transfections to finally generate the desired HepG2 cell line were done with CRISPR construct PX458_hPMM2_2.

6.3.6 A first attempt of targeting HepG2 cells with the CRISPR-Cas9 system

Before being able to perform a CRISPR-HDR experiment, the transfection conditions to deliver PX458_hPMM2_2 into HepG2 cells had to be optimised (For details see section 2.1.3.2). After testing different amounts of plasmid DNA (550ng, 880 ng, 1µg, data not shown) to get an optimal outcome, delivering 1 µg of the CRISPR construct PX458_hPMM2_2 resulted in the best transfection efficiency in HepG2 cells. This was in first place indicated by the number of green fluorescent cells as shown in Figure 55.

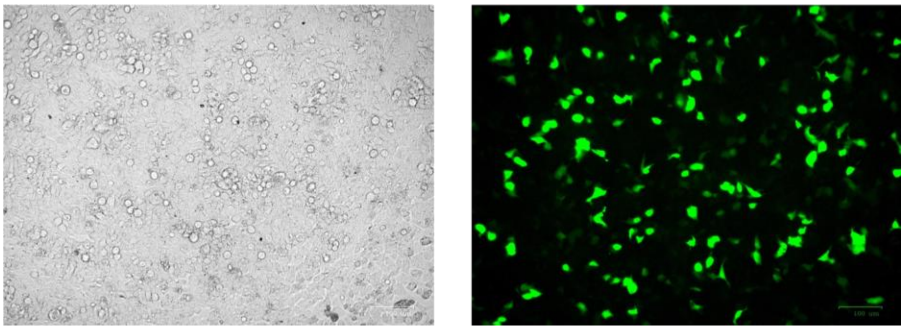


Figure 55: Semi-reverse transfection of HepG2 cells with 1 µg of the GFP-expressing CRISPR construct PX458_hPMM2_2.
Left panel: Brightfield image of PX458_hPMM2_2 transfected HepG2 cells 48 hours after transfection. Right panel: Fluorescence GFP signal 48 hours post-transfection with PX458_hPMM2_2 in HepG2 cells. All pictures were taken with the ZOE cell imager. Scale bar (bottom right corner): 100 nM.

By comparing the two images in Figure 55, it becomes obvious that the number of transfected cells was quite low in relation to the high cell density present in the well. This indicates an overall low transfection efficiency reached by the setting used. However, it was deemed sufficient to proceed and the cells were FACS-sorted after 72 hours post transfection. This helped to separate the green GFP-expressing cells from the non-transfected cells. The green fluorescent cells form all wells were collected in one. This guaranteed an enrichment of transfected cells to increase the chances of a better overall outcome of the experiment. The gDNA was extracted from a portion of these cells and the pooled gDNA was analysed by Sanger sequencing, after amplification of the region of interest (*PMM2* For3/Rev3 primer pair). Figure 56 shows the result of the sequencing reaction.

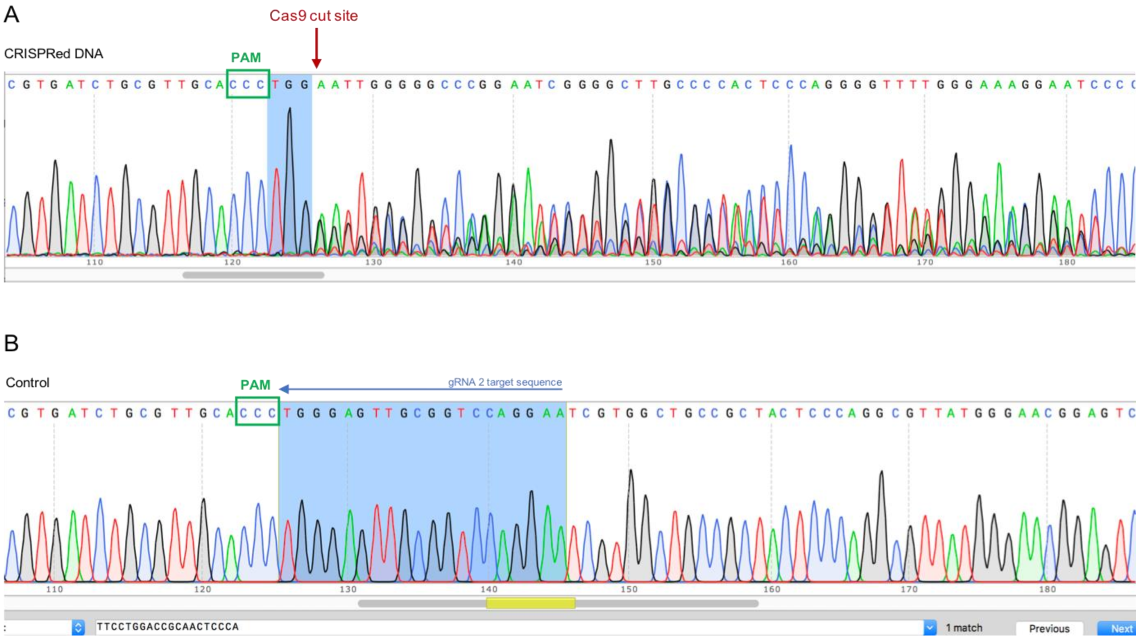
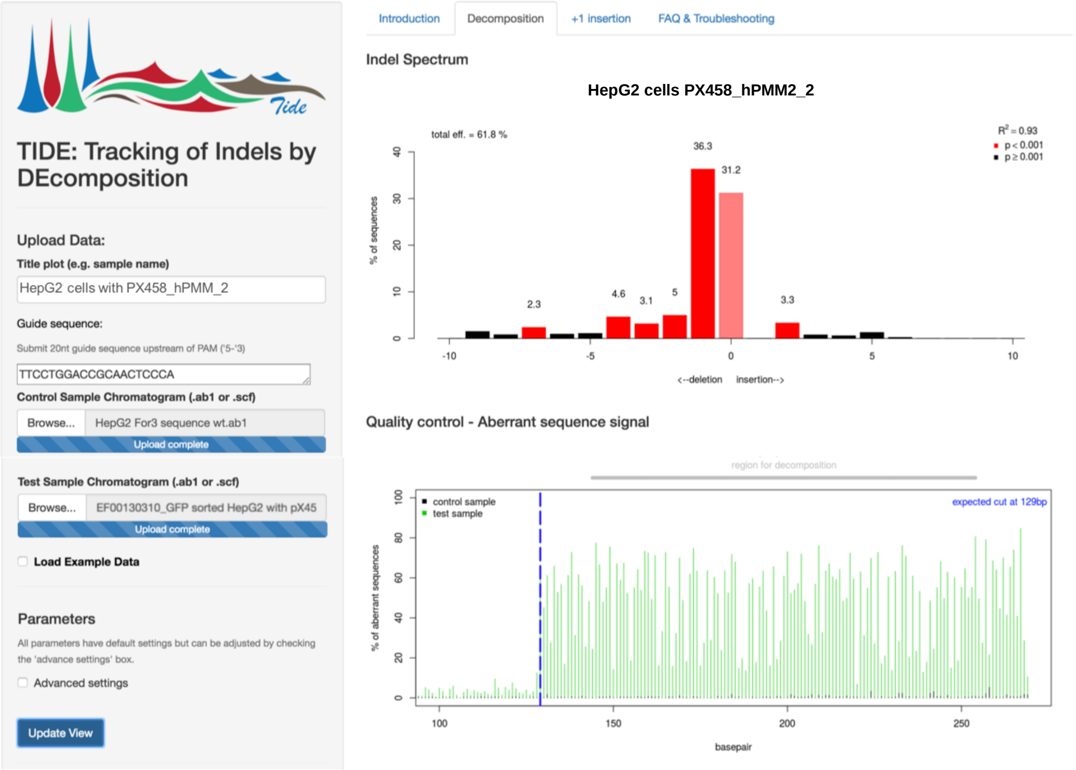


Figure 56: Electropherograms of pooled gDNA from transfected HepG2 cells with CRISPR construct PX458_hPMM2_2 (1 µg) compared to a wt control.
Both sequencing traces shown are in forward direction previously amplified with primer *PMM2* For3/Rev3 and sequenced with primer *PMM2* For3. Highlighted in blue are the gRNA target sequences if possible. Green box marks the PAM site. Red arrow indicates the Cas9 cut site. A: Electropherogram of sequenced HepG2 gDNA amplicon after transfection with PX458_hPMM2_2. B: Wt electropherogram of *PMM2* promoter ROI from HepG2 cells.

The electropherogram of the edited gDNA in Figure 56 A shows the desired “messy” appearing sequencing traces as expected after targeting the cells with the CRISPR-Cas9 system. Indels have formed right after the Cas9 cut site and have shifted the sequence to the right to an extent the gRNA target sequence cannot be identified anymore. This indicates, that the majority of the cells have been edited, which was confirmed performing a TIDE analysis (Figure 57). B shows the clean wt control sequence in comparison.



TIDE: Tracking of Indels by DEcomposition

Upload Data:

Title plot (e.g. sample name): HepG2 cells with PX458_hPMM_2

Guide sequence: Submit 20nt guide sequence upstream of PAM ('5-'3): TTCCTGGACCGCAACTCCCA

Control Sample Chromatogram (.ab1 or .scf): HepG2 For3 sequence wt.ab1 — Upload complete

Test Sample Chromatogram (.ab1 or .scf): EF00130310_GFP sorted HepG2 with pX45 — Upload complete

☐ Load Example Data

Parameters

All parameters have default settings but can be adjusted by checking the 'advance settings' box.

☐ Advanced settings

Figure 57: TIDE analysis comparing and quantifying the alterations of sequencing traces from PX458_hPMM2_2-edited HepG2 cells to a wt control.
sgRNA sequence hPMM2_2 was provided as a reference string. Top: Indel spectrum and calculated cutting efficiency in percent. Bottom: Aberrant sequence signal of the two input sequence traces and the expected cut site. Black: Wt sequence. Green: Sample sequence. Blue: Expected cut site.

TIDE measured an overall cutting efficiency of 61.8% within the pooled sample of FACS-sorted HepG2 cells when transfected with PX458_hPMM2_2. As previously seen and again shown by the indel spectrum, mainly deletions of one nucleotide (36.3%) were detected. In addition, deletions of five to nine nucleotides were found with lower percentages, ranging from 5% to below 2%. The total number of insertions was significantly lower with 3.3% and less of insertions with two to six nucleotides. The aberrant sequence signal nicely reflects the result of the Sanger sequencing (Figure 56), by reconstructing the significant alterations between the CRISPRed and the wt DNA samples right after the Cas9 cut site demonstrated by the high green peaks (CRISPR sample) opposed to the short black ones (wt) in the bottom graph.

The concentration of the “green” transfected HepG2 cells has worked quite well as demonstrated by the high cutting efficiency obtained using PX458_hPMM2_2. This experimental setting was then kept for further experiments optimising the co-delivery with the ssODN – the second step necessary to create the desired HepG2 cell line harbouring the c.-167G>T SNP causative for the HIPKD phenotype.

6.3.7 Optimisation of co-delivering CRISPR construct and ssODN

The first co-transfection was done using the same conditions as previously described and the ssODN was added. To get a first impression on how the cells handle the additional DNA input, a standard volume of 0.5 µL (2.1 µg) of the 100 µM ssODN were used. The result is illustrated in Figure 58 showing a portion of the transfected cells.

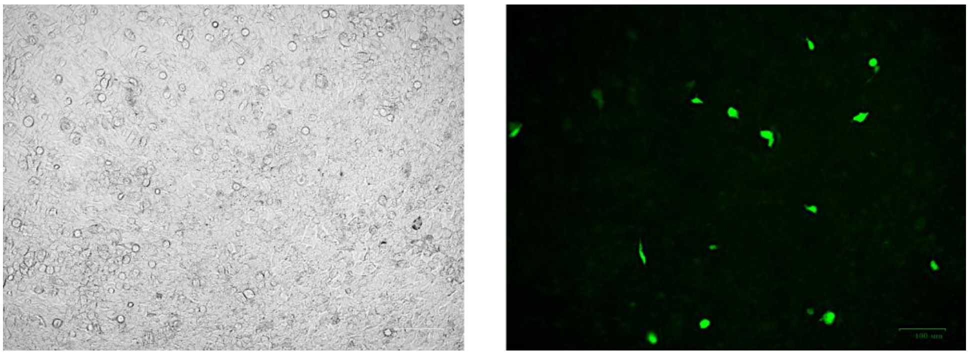


Figure 58: Optimisation of semi reverse co-transfection of HepG2 cells.
1 µg of the CRISPR construct PX458_hPMM2_2 and 2.1 µg of the ssODN were used. Left panel: Brightfield image of transfected HepG2 cells 72 hours after transfection. Right panel: Fluorescence GFP signal in HepG2 cells 72 hours post-transfection. Images were taken with the ZOE cell imager. Scale bar (bottom right corner): 100 nM.

Many cells survived the transfection procedure as demonstrated by the high cell density shown in the brightfield image (left). However, only a few cells showed GFP expression (right picture) indicating insufficient uptake of the CRISPR construct in the majority of cells. In a next experiment, different amounts of the ssODN DNA were delivered to the cells, to find the right balance between the already optimised amount set for the CRISPR construct (1 µg) and the ssODN. Figure 59 shows the transfected cells 72 hours after co-transfection.

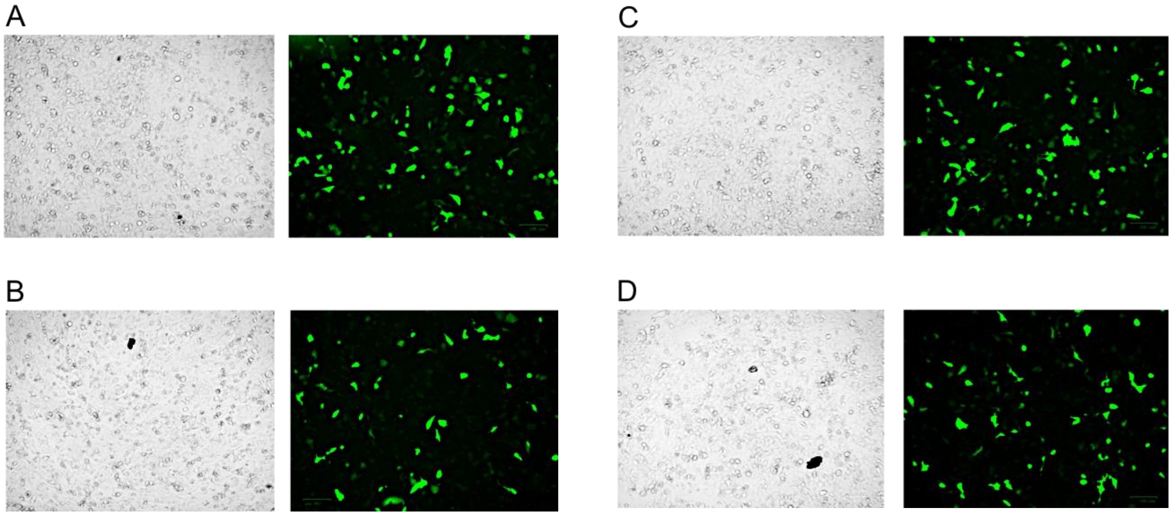


Figure 59: Optimisation of semi reverse co-transfection of HepG2 cells testing different amounts of ssODN.
CRISPR construct: PX458_hPMM2_2 (1 µg). All images were taken 72 hours post-transfection. A: Brightfield and GFP expression of HepG2 cells transfected with 210 ng of ssODN. B: Brightfield and GFP expression of HepG2 cells transfected with 340 ng of ssODN. C: Brightfield and GFP expression of HepG2 cells transfected with 425 ng of ssODN. D: Brightfield and GFP expression of HepG2 cells transfected with 850 ng of ssODN. All pictures were taken with the ZOE cell imager. Scale bar (bottom right corner): 100 nM.

Images A-B show the total cell density (brightfield, left images) and the density of GFP-expressing cells (right images) when co-transfected with 210 ng, 340 ng, 425 ng and 850 ng of the ssODN DNA, respectively. Compared to the result shown in Figure 58, an overall increase of the transfection efficiency was observed in all four cases but did not differ substantially between the different conditions. However, the number of GFP-expressing cells seemed to be slightly higher when co-transfected with either 210 ng (Figure 59 A) or 425 ng (Figure 59 C) of the ssODN compared to B (340 ng) and D (850 ng). In order to compensate for partial degradation, 425 ng instead of 210 ng of the ssODN were finally chosen for further experiments in order to insert the desired SNP in the *PMM2* promoter via the CRISPR HDR pathway.

6.3.8 Successful co-delivery of the CRISPR-Cas9 system and ssODN in HepG2 cells

The following experiment has been performed a few times. The results in Figure 60 show HepG2 cells after co-delivery of PX458_hPMM2_2 (1 µg) together with 425 ng of the ssODN, which finally caused a sufficient gene edit at the CRISPR target site (see also Figure 61 and Figure 62).

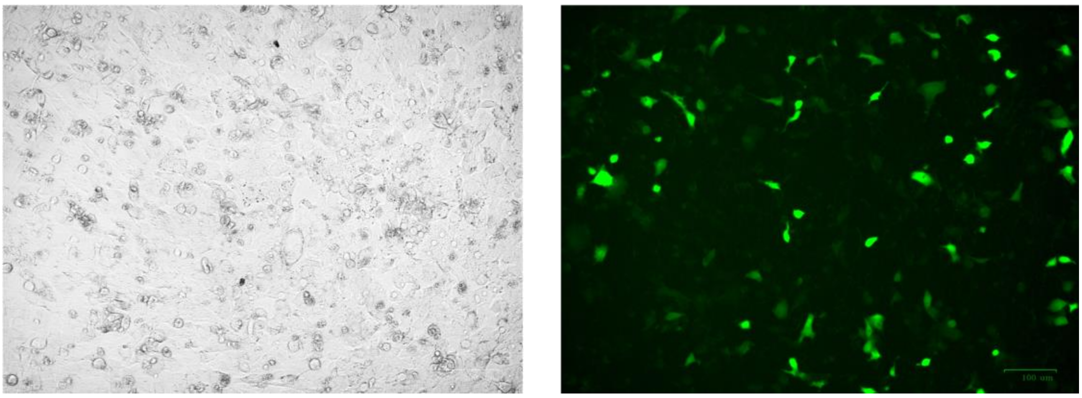


Figure 60: Successful CRISPR co-transfection of HepG2 cells.
Semi reverse co-transfection of CRISPR construct PX458_hPMM2_2 (1 µg) together with 425 ng of the HDR oligo donor template. Left panel: Brightfield image of transfected HepG2 cells 72 hours after transfection. Right panel: GFP expression from PX458_hPMM2_2 in HepG2 cells 72 hours post-transfection. Images were taken with the ZOE cell imager. Scale bar (bottom right corner): 100 nM.

In total six 24 wells with 250,000 cells each were co-transfected during this experiment. The outcome was similar to what was observed previously. The number of transfected GFP-expressing cells (Figure 60, right) was not very high but enough for being separated from the non-transfected ones in a FACS cell sorting session. All “green” cells were collected in the same tube to get the maximum cell concentration possible. Only a few thousand (~7,000) cells were collected, therefore the sorted cells were kept in culture for nine days for recovery and to obtain more cells to work with. After this period, the cells were harvested and divided in three parts: One portion was used for gDNA extraction, some cells were seeded in low density to obtain single cell colonies and the other portion was put in cryopreservation at -80°C to be stored as a backup. From the extracted gDNA, the *PMM2* promoter region of interest was amplified as described before and the amplicon sent for Sanger sequencing. The sequencing result is shown in Figure 61.

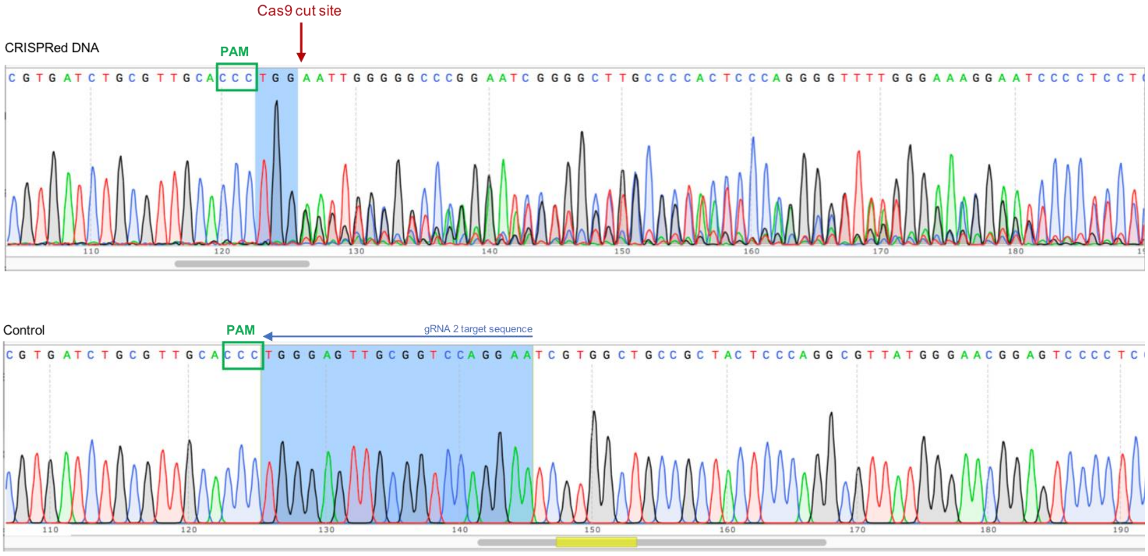


Figure 61: Electropherograms of successful CRISPR co-transfection of HepG2 cells.
Pooled gDNA samples from co-transfected HepG2 cells with CRISPR construct PX458_hPMM2_2 (1 µg) and 425 ng of the ssODN compared to a control (pooled gDNA wt HepG2 cells) were used. Both sequencing traces shown are in forward direction previously amplified with primer *PMM2* For3/Rev3 and sequenced with primer *PMM2* For3. Highlighted in blue are the gRNA target sequences if possible. Green box marks the PAM site. Red arrow indicates the Cas9 cut site. A Top: Electropherogram of sequenced HepG2 gDNA amplicon after transfection with PX458_hPMM2_2 and ssODN. A bottom: wt electropherogram of *PMM2* promoter ROI from HepG2 cells.

Compared to the clean wt sequence (Figure 61 Bottom), the CRISPRed cells clearly have been edited as demonstrated by the shifted sequencing traces becoming apparent after the Cas9 cut site (Figure 61 top) due to indel formation. Furthermore, the gRNA target sequence was not detectable anymore within the sample sequence (Top), indicating the number of edited cells exceeded the number of non-edited cells. To analyse this precisely and as done before, the electropherograms were both put in the TIDE analysis software to quantify alterations between the two sequences and to assess the cutting efficiency of PX458_hPMM2_2. TIDE is not able to detect and quantify HDR events in a pooled sample, therefore the result represents the overall performance of the CRISPR-Cas9 system during this experiment only. Figure 62 shows the TIDE analysis performed.

TIDE: Tracking of Indels by DEcomposition

Upload Data:

Title plot (e.g. sample name)

CRISPR HDR HepG2 PMM2

Guide sequence:

Submit 20nt guide sequence upstream of PAM ('5-'3)

TTCCTGGACCGCAACTCCCA

Control Sample Chromatogram (.ab1 or .scf)

Browse... HepG2 For3 sequence wt_June 2017.ab1

Upload complete

Test Sample Chromatogram (.ab1 or .scf)

Browse... EF00130440_CRISPR HDR transfection 26

Upload complete

☐ **Load Example Data**

Parameters

All parameters have default settings but can be adjusted by checking the 'advance settings' box.

☐ Advanced settings

Update View

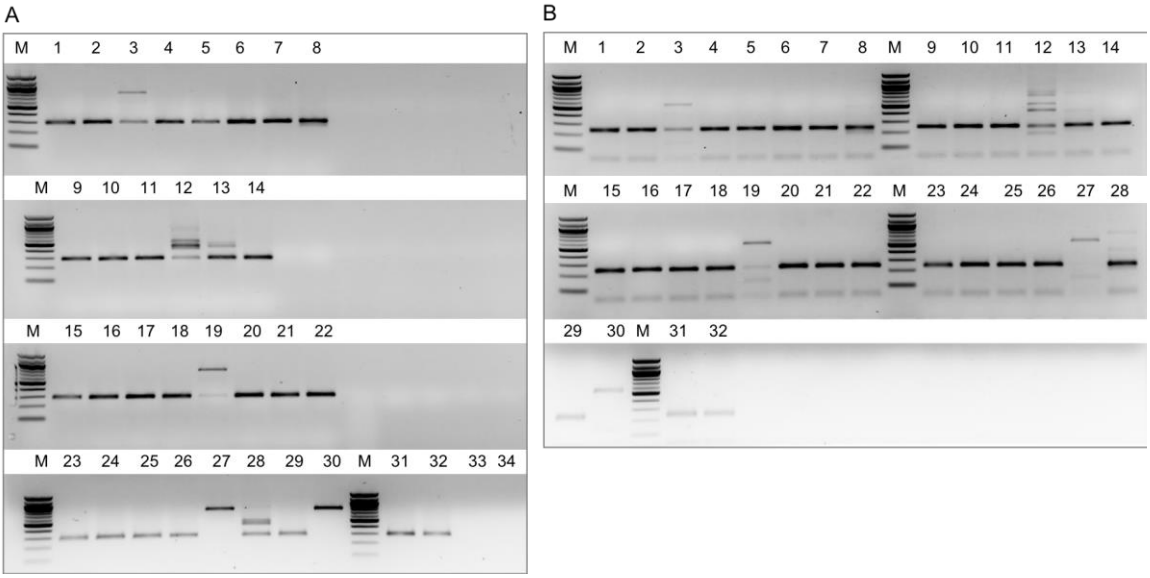
Figure 62: TIDE analysis comparing and quantifying the alterations of sequencing traces from HepG2 cells transfected with PX458_hPMM2_2 and the ssODN to a wt control.
sgRNA sequence hPMM2_2 was provided as a reference string. Top: Indel spectrum and calculated cutting efficiency in percent. Bottom: Aberrant sequence signal of the two input sequence traces and the expected cut site. Black: wt sequence. Green: Sample sequence. Blue: Expected cut site.

The result of the TIDE analysis coincided with a presumed strong performance of the CRISPR-Cas9 system as the total cutting efficiency within the pool of FACS-sorted cells turned out being quite high with 67%. The indel spectrum, represented by the top graph in Figure 62, again indicates deletions of one nucleotide as being the most prominent change introduced after DNA repair. But also, deletions of two to 10 nucleotides have been detected by the software with percentages ranging from 9.4% (2 nt deletion) to below 3.5% (four nt deletion) with seven nucleotide deletions being the least present. On the side of the insertions, those with one nucleotide being inserted were most dominant with 8.8%. However, some insertions of two, three, four, five and seven nucleotides with very low quantities were detected as well. As previously seen and expected, the aberrant sequence signal indicates significant changes after the Cas9 cut site indicated by the high green (sample sequence) and the low black peaks (control sequence) depicted in the bottom graph (Figure 62) and coincides with the altered sequence seen in the electropherogram of the CRISPRed cells.

In order to get an idea whether the HDR repair has worked in some of the edited cells, a test restriction digest (2.4.8) with the same *PMM2* amplicon used for Sanger sequencing was performed (data not shown). The desired SNP creates an additional restriction site for *StyI* (Alias Eco130I), therefore the band patterns of digested DNA with and without the SNP should differ from each other on an agarose gel, if a sufficient number of cells have been repaired according to the ssODN (2.4.4). However, this was not detected on the basis of this experiment. To investigate, whether HDR events have happened within this population of cells, the single cell colonies seeded in low density underwent genotyping.

6.3.9 Genotyping of CRISPRed single cell colonies

In total 200 single cell colonies were picked from the low-density plates (2.4.7) and further expanded separately on 96-well plates for genotyping. They were split asymmetrically (70% and 30%) and cells were lysed and gDNA extracted when the majority of the clones on the 70% plate reached a confluency of above 70%, usually no longer than 3-5 days after splitting (2.4.7). To screen the clones for the desired mutation, first the *PMM2* promoter region of interest was amplified as usual using the *PMM2* For3/Rev3 primer pair and on the basis of the gDNA obtained from each clone. The PCR was analysed on a 2% agarose gel and the amplicon further used for digestion with *StyI* to identify HDR-positive clones (2.4.8). Figure 63 shows a representative example of the analytical agarose gels after PCR and *StyI* digest of the first 31 clones.



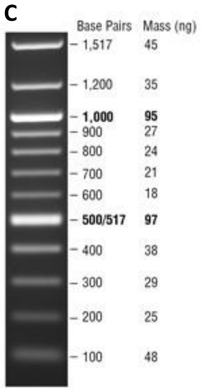


Figure 63: Genotyping of 31 single cell clones after co-transfection with PX458_hPMM2_2 and the homologous ssODN (representative for all 200 clones screened).
2% agarose gels showing the PCR amplicons before (A) and after *StyI* restriction digest (B). A and B: M: 100 bp DNA ladder. Lanes 1-8: Clones 1-8. Lanes 9-14: Clones 9-14. Lanes 15-22: Clones 15-22. Lanes 23-30: Clones 23-30. Lanes 31-32: Clone 31 and wt control. Lane 33 (A only): Master mix only (negative control). Lane 34 (A only): H_2O only (negative control). C: 100 bp DNA standard used for A and B (Figure 75).

The resulting DNA band after PCR was expected to run at 300 bp for both, wt and HDR-positive clones. Although this was the case for the majority of clones, in some of them (clones 3, 8, 12, 13, 14, 19, 27, 28, and 30) the DNA band pattern differed significantly, with additional bands appearing for example at 1,000 bp (e.g. clones 3 & 19) and/or 450 bp, 500 bp and 600 bp (e.g. clones 12, 13 and 28). It is possible, that in these cases a partial insertion of the CRISPR construct has happened explaining the larger DNA bands. Clone 8 stands out with only one additional DNA band just below the expected 300 bp PCR product, which could be due to a bigger indel formation. Although these are no candidates for a positive HDR-clone, they at least indicate CRISPR activity during this experiment once more.

The *StyI* restriction digest was meant to indicate whether the 300 bp PCR amplicons were simply expected to be wt clones or HDR-positive clones, with either two DNA bands at 246 bp and 50 bp or three bands of 144 bp, 102 bp and 50 bp, respectively (see section 2.4.8). Analysing the result after gel electrophoresis (Figure 63 B), none of the candidate clones (Lanes: 1-2; 4-7; 9-11; 14-18; 20-26; 29; 31) showed the DNA band pattern (144 bp, 102 bp and 50 bp) resulting from *StyI* digest when the desired c.-167G>T SNP was present. Comparable results were observed whilst screening the following 170 single cell colonies (data not shown).

Even though the chances of a positive HDR-clone faded after the *StyI* digest, clones 1, 2, 4, 5 and 10 were sent for Sanger sequencing to spot-check the genomic nature of the present clones form this experiment. Three out of the five sequenced clones indeed did show sequence alterations and therefore were clearly edited – that counts for 60% and agrees with the expected cutting efficiency of 67% quantified by TIDE (Figure 62). This confirms a strong activity of the delivered CRISPR-Cas9 system and in particular the gRNA. However, the electropherograms (data not shown) of edited clones only show shifted sequencing traces indicating indels but not the distinct repair according to the provided ssODN which was desired.

In conclusion, the CRISPR-Cas9 system on PX458 has worked sufficiently in HepG2 cells as several clones were identified being edited by the system. However, there was no evidence of clones which were repaired after the HDR repair pathway and therefore a change of strategy for creating the HepG2 HIPKD cell model was decided.

6.3.10 Successful co-delivery of Cas9 protein, sgRNA and ssODN in HepG2 cells – a CRISPR RNP approach

In order to increase the chances of the HDR repair in CRISPR-edited HepG2 cells to happen, further experiments were carried out using Cas9 protein and sgRNA (RNPs) together with the homologous ssODN. The three components were delivered to the cells by electroporation rather than reagent-based transfection. For this purpose, gRNA hPMM2_2 was to be synthesised in sgRNA format (see section 2.4.5.2).

After a few days of recovery, the electroporated cells were divided in three parts. The majority of cells was put in cryopreservation at -80°C to save a backup population in stock. Three times 5,000 cells were seeded in low density using 10 cm cell culture dishes to obtain single cell colonies for genotyping later on. These were kept in culture for 14 days at 37°C and 5% CO_2 and the medium was replaced every five days. The remainder of cells was used to extract gDNA in order to validate the overall outcome of this experiment and the pooled gDNA sample was used as a template for PCR amplification under familiar conditions (2.4.1.3). As a first quality control, the 300 bp amplicon was digested with the restriction endonuclease *StyI* to pre-check whether some HDR events have happened amongst the pool of CRISPR-targeted cells. The digested sample and an equally treated wt control were put on an 2% agarose gel for analysis (Figure 64).

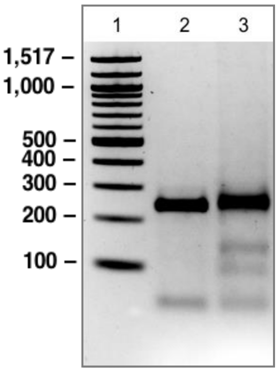


Figure 64: *StyI* restriction digest of pooled electroporated HepG2 cells with Cas9 protein, sgRNA hPMM2_2 and the homologous ssODN compared to a wt control.
Lane 1: 100 bp DNA ladder (Figure 75). Lane 2: Wt control. Lane 3: CRISPR-Cas9 edited sample. 2% agarose gel electrophoresis.

The digested wt control sample shows the expected DNA band pattern, with products of 246 bp and 50 bp in size (Figure 64, Lane 2). Differently, the restriction digest of gDNA from the experimental cells resulted in four DNA bands appearing after electrophoresis. Besides the two bands (246bp, 50 bp) from cutting at the wt site, an additional *StyI* site was generated and restriction resulted in two extra DNA bands, at 144 bp and 102 bp, as a consequence of HDR repair due to introduction of the desired HIPKD SNP.

To further confirm mutagenesis in HepG2 cells the same pooled gDNA sample was sent for Sanger sequencing (Figure 65).

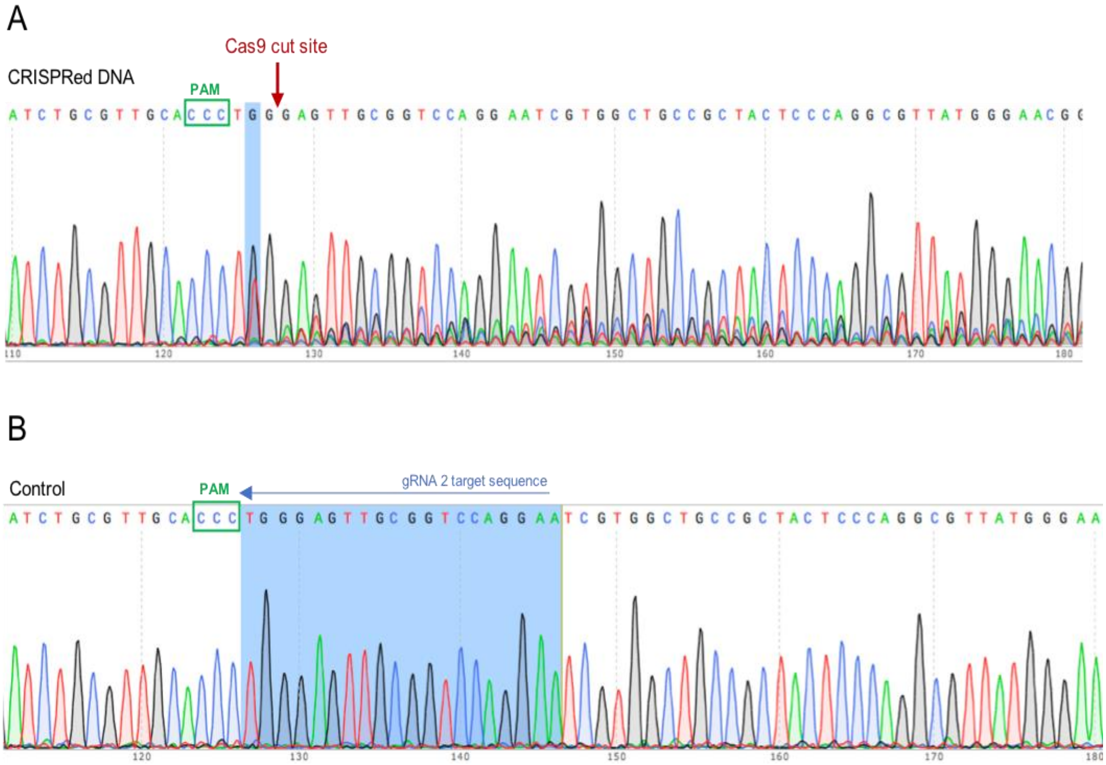
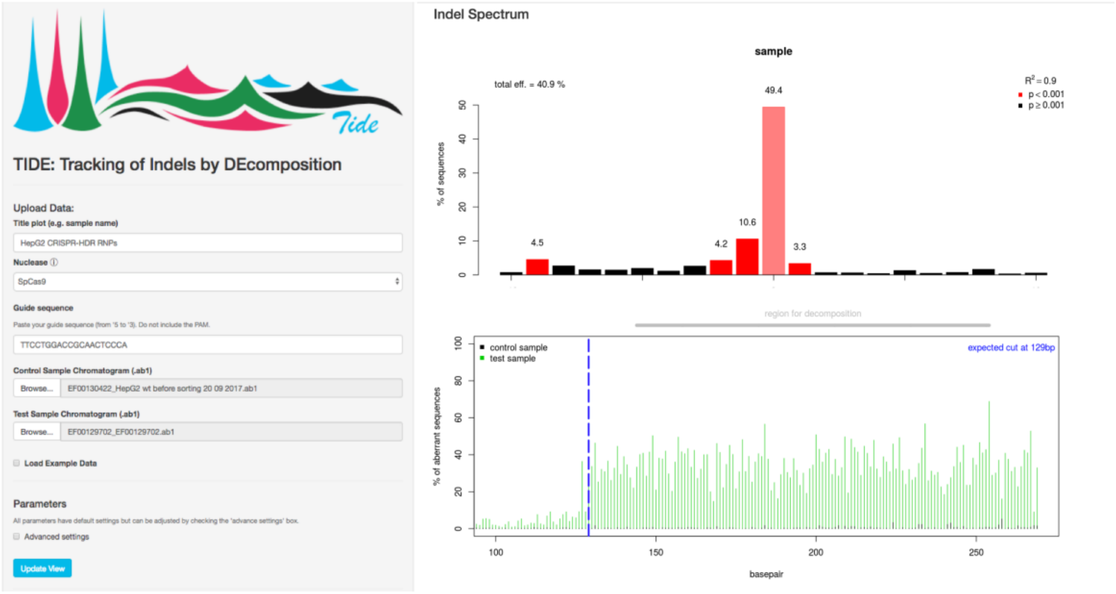


Figure 65: Electropherograms of pooled gDNA from electroporated HepG2 cells with Cas9 protein, sgRNA hPMM2_2 and the homologous ssODN.
A: CRISPRed DNA. B: wt. Both sequencing traces shown are in forward direction previously amplified with primer *PMM2* For3/Rev3 and sequenced with primer *PMM2* For3. The green box marks the PAM site. The red arrow indicates the Cas9 cut site. A: Electropherogram of sequenced HepG2 gDNA amplicon after electroporation. Highlighted in blue is the CRISPR site c.-167G. B: Wt electropherogram of the *PMM2* promoter ROI from HepG2 cells. Highlighted in blue is the sgRNA target site.

Compared to the clean wt sequence (Figure 65 B), a clear sequence shift as seen in the electropherogram showing the sequencing traces of the edited cells (Figure 65 A) indicates sufficient CRISPR activity and indel formation. The CRISPR site highlighted in blue shows a double peak for two nucleobases, guanine “G” and thymine “T”. Hence, both the wt and the mutant version, HepG2 cell pool as the desired SNP is a guanine to thymine substitution (c.-167G>T).

The overall efficiency and quality, the CRISPR-Cas9 induced mutagenesis has worked with, was analysed in TIDE as usual (Figure 66).



TIDE: Tracking of Indels by DEcomposition

Upload Data:

Title plot (e.g. sample name): HepG2 CRISPR-HDR RNPs

Nuclease: SpCas9

Guide sequence

Paste your guide sequence (from '5 to '3). Do not include the PAM.

TTCCTGGACCGCAACTCCCA

Control Sample Chromatogram (.ab1): EF00130422_HepG2 wt before sorting 20 09 2017.ab1

Test Sample Chromatogram (.ab1): EF00129702_EF00129702.ab1

☐ Load Example Data

Parameters

All parameters have default settings but can be adjusted by checking the 'advance settings' box.

☐ Advanced settings

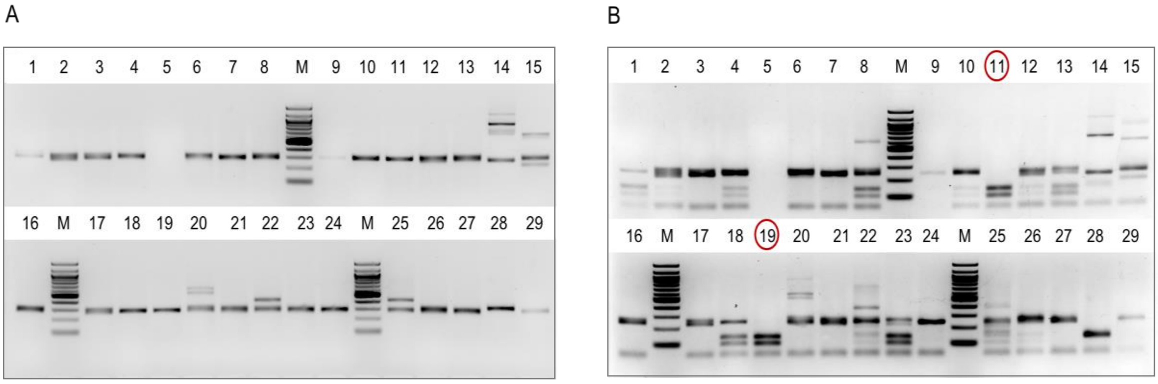
Figure 66: TIDE analysis comparing and quantifying the alterations of sequencing traces from HepG2 cells edited with Cas9 protein, sgRNA hPMM2_2 and the homologous ssODN. Comparison to wt control. sgRNA sequence hPMM2_2 was provided as a reference string. Top: Indel spectrum and calculated cutting efficiency in percent. Bottom: Aberrant sequence signal of the two input sequence traces and the expected cut site. Black: Wt sequence. Green: Sample sequence. Blue: Expected cut site.

The calculated cutting efficiency of the CRISPR-Cas9 system under guidance of sgRNA hPMM2_2 amounted to 40% in this experiment. As seen previously, the majority of changes were due to deletions of one nucleotide (10.6%), followed by deletions of two and nine nucleotides of above 4%. Deletions of sizes in between occurred with lower percentages, not defined numerically within the Indel spectrum (Figure 66, top graph). Some insertions of one nucleotide were detected with a frequency of 3.3% and others from two to ten nucleotides were acknowledged with lower percentages. About 50% of the compared sequences were identical with 0% changes. The reconstruction of the aberrant sequence signals in the bottom graph (Figure 66) matches with the appearance of the two electropherograms, showing a significant difference in the sequencing traces after the Cas9 cut site. This indicates indel formation and is demonstrated by the high and low peaks in green and black, representing the edited and wt DNA sequences, respectively.

6.3.11 Genotyping of single cell HepG2 colonies to obtain a homozygous HIPKD clone

In total 144 single cell HepG2 clones CRISPRed with RNPs, which all had gone through the entire editing process were included in the screening to identify the precise genotypes needed for further experiments. For a full set of desired clones, it was of interest to find at least and most importantly one positive HIPKD clone homozygous (c.-167G>T/c.-167G>T) for the desired SNP within the *PMM2* promoter and additionally at least one homozygous wt (c.-167G/c.-167G) clone. In order to find these clones, the gDNA of each clone was extracted, the region of interest amplified in a standard PCR (2.4.1.3) and the amplicons used as a template for restriction digest with *StyI*. As described previously (see section 6.3.9) to do this, single cell colonies were picked, kept separated in 96 well plates and split asymmetrically with different cell densities (70% and 30%). gDNA of each clone was extracted after a few days of growth (70% 96 well plate) and amplified with the *PMM2* For3/Rev3 primer pair. Each amplicon then was digested with *StyI*, to enable the identification of HDR events within the population of screened HepG2 clones (see Figure 26). The results obtained from 29 clones after gel electrophoresis of the samples are shown in Figure 67 which depicts a representative example for both, the PCR reaction and the *StyI* restriction digest.

The backup plate (30% 96 well plate) was kept in culture to keep every clone included in the screening alive.



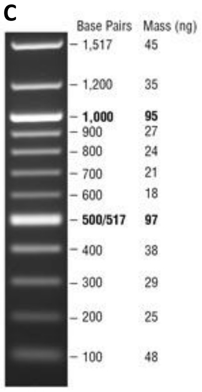


Figure 67: Genotyping PCR and *StyI* restriction digest of 29 HepG2 single cell colonies edited with Cas9 protein, sgRNA hPMM2_2 and the homologous ssODN.
Representative example for all 144 clones screened. 2% agarose gels showing the PCR amplicons before (A) and after *StyI* restriction digest (B). A and B: M: 100 bp DNA ladder (Figure 75). Lanes 1-8: Clones 1-8. Lanes 9-16: Clones 9-16. Lanes 17-24: Clones 17-24. Lanes 25-29: Clones 25-29. Red circles indicate clones with DNA band patterns particular for a homozygous HDR. C: 100 bp DNA standard used for A and B (Figure 75).

The majority of all clones included in the genotyping PCR resulted in the expected amplicon running at 300 bp as seen on the example in Figure 67 A. About 12% of all clones included in the screening showed additional bands of larger sizes between 400 and 1,200 bp likely as a result of large indel formations. This is reflected by clones 14, 16, 20, 22 and 25 (Figure 67 A). However, all 144 samples were incorporated in the *StyI* restriction digest, resulting in a variety of band patterns emerging as seen in Figure 67 B, thus indicating a mixed population of genotypes amongst all the single cell HepG2 clones screened (here represented by 29 clones). Table 14 lists the combinations of DNA band patterns and the frequency they appear with, within the population of all 144 edited HepG2 clones screened as a result of *StyI* restriction.

Table 14: Six different categories of DNA band patterns observed after *StyI* restriction digest. 144 single cell HepG2 clones electroporated with CRISPR-Cas9 RNPs and ssODN were screened. Column 2: Different DNA band patterns present. Column 3: DNA band sizes according to pattern. Column 4: Possible genotype(s). Column 5: Number of clones per category. Wt: c.-167G/ c.-167G. HDR: c.-167G>T/ c.-167G>T. Indel: Random insertions and deletions at the CRISPR target site.

	# DNA bands	**DNA band sizes (bp)**	**Interpretation**	**# of clones**
1	2	246, 50	wt/wt; wt/Indel; Indel/Indel	38
2	3	246, 230, 50	wt/Indel	9
3	3	144, 102, 50	HDR/HDR	8
4	4	246, 150, 100, 50	HDR/wt; HDR/Indel	39
5	5	246, 230, 150, 100, 50	HDR/Indel	10
6	multiple	50 – 1,200	Indel/Indel; wt/Indel; HDR/Indel	27
Total number of clones count:				131*

* Variation in total number of clones due to empty (no DNA) samples or results not clearly interpretable

For further experiments, the aim was to ideally identify three homozygous HDR clones and three clean wt clones to provide a control group which has gone through the same experimental process. In total, eight out of all these clones showed the HDR specific band pattern and therefore seemed to be homozygous for the c.-167G>T mutation, here represented by clones 11 and 19 (Figure 67 A, red circles). The majority of genotyped HepG2 clones showed DNA band patterns indicating indel formation after *StyI* digest. However, according to their DNA band pattern 38 clones (category 1, Table 14) theoretically remained wt and although the presence of small indels could not be excluded, these were considered to be potential candidates for the control group. In order to identify the desired clones all eight clones appearing homozygous for the c.-167G>T mutation and five seemingly clean wt clones were included in Sanger sequencing analysis for confirmation.

Figure 68 depicts the electropherograms of positive homozygous HDR clones and clean wt clones only.

Figure 68: Sanger sequencing of positive HepG2 homozygous HDR (c.-167G>T/c.167G>T) and wt clones obtained from CRISPR-Cas9 editing and HDR repair.

Nucleotide position c.-167 to be edited is highlighted in blue. Green boxes mark the PAM sites. Red arrow indicates the Cas9 cut sites. All sequencing traces shown are in forward direction previously amplified with primer *PMM2* For3/Rev3 and sequenced with primer *PMM2* For3. A: Electropherograms of three homozygous clones (Clone 1, Clone 2, Clone 3) showing the c.-167G>T SNP. B: Electropherograms of three clean wt clones (WT Clone 1, WT Clone 2, WT Clone 3) with no SNP included. C: Electropherogram of non-edited HepG2 wt cells. Blue arrow marks the gRNA target site.

In the case of the electropherograms in Figure 68 no indels were expected in these sequencing traces and therefore performing a TIDE analysis was not relevant here. However, Figure 69 depicts the TIDE analysis for homozygous Clone 3 exemplary for all homozygous clones for the purposes of illustration.


TIDE: Tracking of Indels by DEcomposition

Upload Data:

Title plot (e.g. sample name): Homozygous HIPKD clone 3

Nuclease: SpCas9

Guide sequence: TTCCTGGACCGCAACTCCCA

Paste your guide sequence (from '5 to '3). Do not include the PAM.

Control Sample Chromatogram (.ab1): EF00132769_EF00132769.ab1

Test Sample Chromatogram (.ab1): EF00132963_EF00132963.ab1

☐ Load Example Data

Figure 69: TIDE analysis comparing and quantifying the alterations of sequencing traces of positive HDR Clone 3 homozygous for the c.167G>T SNP to a wt control.
sgRNA sequence hPMM2_2 was provided as a reference string. Top: Indel spectrum and calculated cutting efficiency in percent. Bottom: Aberrant sequence signal of the two input sequence traces and the expected cut site. Black: wt sequence. Green: Sample sequence. Blue: Expected cut site.

TIDE quantifies sequence alterations by aligning the two input sequencing traces and then calculates the overall efficiency of CRISPR-Cas9 of a particular experiment. Therefore, it is not surprising that the cutting efficiency was determined 0% or nearly 0% in the case of the verified clones, as those compared sequences differ at one (homozygous HIPKD clones) or none (wt clones) nucleotide position only. In addition, as expected, no insertions or deletions were detected by the software as seen in the top graph (Figure 69). The aberrant sequence signal by contrast nicely confirms sequence equality of both, wt and mutant, apart from one large peak indicating the desired SNP in the test sample sequence (green) due to CRISPR editing and HDR repair, appearing just before the Cas9 cut site (blue) and matching the expectation thoroughly.

All six clones (3x HDR positive (c.-167G>T/c.167G>T), 3x wt) were expanded in cell culture and their genotypes reconfirmed with gDNA obtained from each clone afterwards (Data not shown). Unfortunately, wt clones 1 and 2 were not clean anymore after expansion as indels have appeared in the sequence (data not shown).

For further experiments, stocks of each clone, homozygous and wt, were prepared for cryopreservation at -80°C.

6.3.12 Testing positive homozygous HepG2 HIPKD clones for off-targets

Homozygous HIPKD clones were only screened for the most likely off-target effects as calculated by CRISPOR and CHOPCHOP, the tools used for gRNA design. Only one off-target with two mismatches was predicted by CRISPOR and chosen to be screened for. All the other ones had three or more mismatches and were not incorporated in the analysis as those off-target effects are very unlikely to happen. The off-target chosen to look at in the homozygous HIPKD clones is located within the first intron of the gene TTTY14 on chromosome Y. The sequence variations between gRNA hPMM2_2 and the off-target are shown below (Table 15).

Table 15: Difference between gRNA hPMM2_2 and off-target sequence.
Mismatches are marked in red.

	Gene	**Sequence**	**PAM**
gRNA hPMM2_2	*PMM2*	TTCCTGGACCGCAACTCCCA	GGG
Off-target ChrY	*TTTY14*	TTCCTGGCCCGTAACTCCCA	TAG

The Sanger sequences of all homozygous HIPKD clones after amplification with specific primers were compared to a wt control HepG2 sequence of the same region. No difference between the sample and control sequence was observed. Hence, gRNA hPMM2_2 did not induce the predicted potential off-target effect of on chromosome Y (data not shown).

6.4 Discussion

The CRISPR-Cas9 system has rapidly become the gold-standard method for genome editing projects over the past five years within the fields of biomedical and biotechnical sciences to create model organisms to study human disease. Moreover, it us used by the industry sector to improve crops and livestock breeding. It has surpassed the advantages of similar techniques, such as ZNFs (zinc finger nucleases) and TALENs (transcription activator-like effector nucleases) due to its ease of use, cost-effectiveness and its efficiency in nearly any cell type. ZNFs and TALENs are site-specific endonucleases fused to a DNA binding domain (Protein-DNA binding) which need elaborate engineering and cloning for every single locus to reach specificity. The CRISPR-Cas9 system by contrast, only needs the individual gRNA for targeted gene editing and functions after the RNA-DNA hybridisation principle – thus, providing a much simpler option in terms of design and synthesis [270,271].

Cong et al. and Mali et al. are the pioneers in demonstrating that CRISPR-Cas9 system can be used to successfully genome edit mammalian cells. Since then, the system has been extensively used in biomedical research and contributed to the generation of multiple model systems in eukaryotic cells and animals, such as zebrafish, mice and monkey [272-274]. Moreover, the first trials applying CRISPR in humans to correct disease-causing mutations are ongoing since 2016 [275,276].

For the reasons and advantages mentioned above, its preciseness and reliability, the CRISPR-Cas9 system was chosen for this study to generate a HIPKD cell model with HepG2 cells having the *PMM2* promoter mutation incorporated in homozygous state. However, and despite its simplicity and reliability opposed to other genome editing methods there are still some challenges involved, especially when genome engineering a cell line for the first time. First of all, it is essential to find and chose a strong gRNA with high specificity and minimum risk of producing off-target effects at the same time, as those can have severe effects on the genome. To simply knock-out a genes' function following the NHEJ repair, various regions within the coding region of the gene of interest are possibly suitable to be targeted with CRISPR [277]. Scientist have created libraries with highly efficient predesigned gRNAs commercially available which are specific for a broad range of genes with often more than one appropriate gRNA per gene. Aiming for HDR to introduce a change at a particular site within the DNA, the search for an optimal gRNA target sequence is far more limited. It is essential that the specific gRNA binds close to the site to edit. Ideally, within a distance of no more than 10 nt up- or downstream of the desired genome edit, as with increasing distance between gRNA target-site and site of modification the editing efficiency decreases significantly [278]. A successful CRISPR-HDR gene edit is therefore highly dependent on PAM motifs being in very close proximity to the target site. In the case of the HIPKD HepG2 cell line, two potential gRNA target sequences just adjacent of position *PMM2* c.-167G were chosen and first tested for their efficiency in HEK293T cells. Prior to transfection they were assessed for their specificity *in silico* with CHOPCHOP and CRISPOR. Yet, our experience demonstrates, that such *in silico* tools are still limited: according to both those CRISPR web tools, gRNA hPMM2_1 seemed the better candidate to use, reaching higher specificity scores compared to gRNA hPMM2_2. Moreover, in CHOPCHOP gRNA hPMM2_2 was not even categorised as being a reasonable gRNA to use. However, after transfections and subsequent TIDE analysis gRNA hPMM2_2 gave the better result (Figure 53, Figure 54) with a higher cutting efficiency and was therefore chosen to proceed with in the actual experiment with HepG2 cells to generate the desired HIPKD cell model.

Such difficulties with the different web-based search tools arise for example due to variation of parameters the developers include to predict unwanted off-target effects of the gRNAs observed and the number of mismatches they consider [274]. CHOPCHOP for instance developed its own unique method which only searches for potential off-targets with maximum three mismatches and limits potential off-targets to the 20 bp upstream of the PAM without for example taking possible DNA “bulges”(unpaired short DNA which can loop-out) [279] into account, which are also prone to cause off-target effects [279]. A general restriction is also, that many tools rely solely on prediction of sequence homology between the gRNA and the targeted genome, often resulting in inaccurate rankings due to the lack of actual experimental data. The first version of CHOPCHOP is such an example [280]. But as new experimental methods are on the rise, generating unbiased data, this helps to improve predictions and to identify “true” off-targets by providing more insights into the molecular mechanisms of the CRISPR-Cas9 system and its affinity to the targeted DNA. CRISPOR has been developed more recently and aims for collecting all such experimentally confirmed off-targets known so far to provide a more elaborate service and to improve the assessment of off-target effects when designing a CRISPR experiment. In addition, it includes off-targets with up to four mismatches and provides also the numbers of off-targets which are not within the seed region [215]. The seed region is the first 10 bp of the 3’ gRNA target sequence just adjacent to the PAM, to which the gRNA begins to anneal to its target. Homology between the gRNA and DNA seed sequences is crucial because only if they match properly the gRNA will continue to anneal completely to mediate Cas9 cleavage. The gRNA first scans potential target sites to ensure specificity, a safety mechanism which prevents cutting of the host genome, considering that the CRISPR system is a defence mechanism in prokaryotes. In research, this capability of the gRNA helps to minimise unwanted off-target effects. In the case of mismatches within the seed region the system likely abolishes cleavage and therefore the risk of causing off-targets is significantly reduced. It is the mismatches outside the seed region, distal of the PAM that are responsible for the risk of off-targets, as the gRNA is more likely to tolerate those [281]. That explains why tools like CRISPOR distinguish between potential off-targets within and outside the seed region to estimate a gRNAs specificity. Overall, CRISPOR seems to be a reliable tool to trust and therefore the poor ranking of gRNA hPMM2_2 in CHOPCHOP was not pivotal. Although CRISPOR identified 128 potential off-targets (opposed to 9 in CHOPCHOP) for gRNA hPMM2_2, none of them included mismatches outside the seed region and therefore those off-targets were unlikely to happen (Figure 47). This was confirmed by looking into one of the potential off-targets to be introduced by gRNA hPMM2_2 in the homozygous HIPKD HepG2 clones (data not shown). CRISPOR predicted one off-target effect with two mismatches on chromosome Y, this was the most likely one to happen as the other off-targets predicted had three or more mismatches. The rule of thumb here is, the more mismatches within the sequence the less likely the off-target effect to happen [282]. PCR and subsequent Sanger sequencing of the off-target region on chromosome Y showed that no off-target had been introduced.

Additional issues getting the CRISPR experiments with HepG2 cells to work were related to the transfection itself, the delivery method and the format in which the system was supplied. The efficiency of transfection heavily depends on the cell line of choice as some cell lines are easy and others hard-to-transfect for various reasons. HepG2 is known to be a difficult-to-transfect cell line. This is also demonstrated by the lower amount of GFP expression in HepG2 cells compared to HEK293T cells (Figure 52, Figure 55). The first attempts to transfect HepG2 cells overall failed (data not shown) and a lot of optimisation in terms of culture conditions, subculturing and handling prior to transfection was needed. HepG2 cells are hard-to-transfect mainly due to their disposition to grow in clumps. This significantly reduces the accessibility of the cell membrane, affecting the transfection efficiency. Overall, several external factors have their influences on the success or failure of transfection and need to be addressed [283]: It is essential that the cells to transfect are healthy and actively dividing. Different seeding densities were also tested as this also varies across cell types. Changing the procedure from traditional transfection (adherent cells) to the semi-reverse transfection (cells in suspension) made a big difference and improved the uptake of the CRISPR system. In order to advance the overall health state of the HepG2 cell population used for this study, the frequency of subculturing was increased, and passage numbers were kept below 15 passages. Furthermore, the composition of the culture medium was improved by adding additional growth factors to promote cell growth (personal communication). To prevent HepG2 cells from clumping a lot of effort was put into the separation of the cells, which has contributed to increase transfection efficiency.

Chemical, reagent-based transfection was used initially to deliver the CRISPR-Cas9 system on a plasmid, as well as the HDR repair template to HepG2 cells. After some optimisation this protocol gave a good overall cutting efficiency after FACS-sorting the

GFP-expressing (the transfected) cells. However, after genotyping about 200 single cell colonies (Figure 63) and although it induced sufficient mutagenesis as demonstrated by clones being repaired through the NHEJ pathway, this approach did not provide any evidence for clones being repaired through the HDR pathway. Hence, we initially did not witness the inclusion of the desired SNP within the *PMM2* promoter. It is recognised that there is a significant difference in terms of efficiency between NHEJ and HDR, with HDR being by far the rarer event, which shows its impact here [274,284] One major reason surely is that an HDR is restricted to late S- and G2-phase, when DNA-replication is completed during the cell cycle and a repair template is required. The NHEJ can take place during G1, S- and G2-phase, hence competes with the HDR pathway [285,286]. Plasmid-based delivery of the system holds some limitations disadvantageous for CRISPR. Part of the plasmid for example can be integrated in the host genome randomly at the on- and also off-target sites, with the latter being difficult to recognise [287]. Such events were observed in the clones screened with large additional bands appearing after PCR of the CRISPR target region (Figure 63). Another obstacle of using CRISPR plasmid DNA is its translocation into the nucleus through the nuclear pore complex (in non-dividing cells) to allow transcription of the system. As CRISPR expression plasmids are quite big, this is limiting their ability to move smoothly through the cytoplasm with the potential to 1) not reaching the nucleus at all and 2) getting degraded by cytoplasmic nucleases, thus reducing CRISPR efficacy. The half-life of plasmid DNA within the cytoplasm amounts to 50-90 minutes [288]. Moreover, aiming for an HDR event by delivering a ssODN (DNA repair template) in parallel, part of it might got degraded already before Cas9 protein and sgRNA were synthesised and active, as HepG2 cells were quite slow with only a few “green” glowing cells being observed after 24 hours post transfection. Finding the desired homozygous HIPKD clone seemed very unlikely continuing with the plasmid-based approach. For this reason, the experimental set-up was changed to using CRISPR ribonucleoproteins (Cas9 protein and sgRNA complexes, RNPs). Those RNPs were delivered to the cells via electroporation as it significantly enhances the frequency of CRISPR compared to reagent-based delivery, demonstrated by different studies. This approach was successfully applied resulting in three homozygous HIPKD clones (Figure 67, Figure 68). The big advantage of using CRISPR RNPs is that once within the cytoplasm, the system is ready and active to immediately cleave the targeted DNA. Simultaneously, RNPs get degraded rapidly within the cytoplasm, therefore contributing to reduced off-target mutations compared to CRISPR plasmids [274,289,290]. In conclusion, this CRISPR experiment provides a nice example demonstrating the power of CRISPR RNPs compared to a traditional plasmid-based approach. CRISPR RNPs are more efficient as they are active once within the cell and are also more reliable due to reduced off-target effects compared to CRISPR plasmids and the risk of large plasmid insertions at on- and off-target sites is also prevented.

Chapter 7: General Discussion

The goal of this study was to gain insights into the disease mechanisms of HIPKD. In the beginning, the patients presented with an ARPKD-like polycystic phenotype with additional appearance of hyperinsulinism. They were initially tested for mutations in the key genes known to be responsible for those rare disorders under the assumption that they may be affected by two separate disorders. However, the causative mutation could not be identified. Thus, the combination of these two phenotypes without the involvement of any known mutations plus the subsequently noted presence of liver cysts in some patients strongly suggested that the patients presented a novel disorder, limited to kidney, pancreas and liver which had not been described before. This motivated us to investigate the genetic cause of this unknown rare disease. The identification of the *PMM2* promoter variant (c.-167G>T) implicated the exact same gene that triggers CDG-1a, a multisystem disorder which is caused by defective protein N-glycosylation due to loss-of-function mutations in *PMM2*. This prompted the question of what mechanisms underlie the difference between the restricted HIPKD phenotype and CDG-1a.

In vitro and *in silico* studies helped us to identify some important key findings suggesting a dysfunctional *PMM2* gene regulation as the major point of the HIPKD disease mechanism, which is restricted to the three affected organs due to presumably altered tissue-specific transcription (hypothesis). This study shall contribute to a better understanding of the underlying disease mechanisms of HIPKD and to work towards the proof of this hypothesis.

7.1 The *PMM2* promoter mutation and gene pleiotropy

As outlined before, mutations in the *PMM2* gene are related to the severe multisystem disorder CDG-1a [180]. However, careful clinical observations excluded this diagnosis in our patients. This was in particular demonstrated by 1) a normal glycosylation pattern of transferrin which is altered in CDG-1a and a typical indicator for the disease and 2) the restricted HIPKD phenotype. While renal cysts and hyperinsulinism have been described as part of the spectrum of manifestations of CDG-1a, consistent with impaired glycosylation as the common underlying basis, the classification of HIPKD as a mild form of CDG-1a was excluded as neurological features, an invariable characteristic in CDG-1a patients, were absent in our HIPKD patients. In CDG-1a the brain is most sensitive to the loss of *PMM2* function [291], whereas there is no evidence that the brain is affected in HIPKD. The non-coding *PMM2* mutation somehow shifts the organ-sensitivity related to

PMM2 dysfunction and limits the disease manifestations distinctively to kidney, pancreas and liver, while *PMM2* activity remains apparently sufficient in the rest of the body. The impact of the promoter mutation results in an organ-specific impairment and thereby distinguishes HIPKD from CDG-1a. Hence, our findings clearly indicate *PMM2* pleiotropy – the ability of one gene to cause different phenotypes – which has not been reported before. Such gene pleiotropies are common phenomena of central evolutionary origin when complex organisms have formed from a restricted group of genes [292]. Although known for more than 100 years now, mechanisms of pleiotropies and how they develop are poorly understood. The effects of gene pleiotropies are quite versatile as some result in only two different disorders whilst others cause multiple different abnormalities. Different mutations in *TRPV4* are the genetic cause for at least nine syndromes with phenotypes ranging from neuromuscular disorders to skeletal dysplasia [293]. Another example are allelic variants in IL-10, which form a complex group of different autoimmune and inflammatory disorders, such as asthma, psoriasis, tuberculosis or Bowel disease which are either the result of enhanced IL-10 function or its deficiency due to genetic alterations (dinucleotide repeats, SNPs) [294].

The classic example within the renal field are mutations in the *OCRL1* gene, either causing the systemic Lowe syndrome which presents with oculocerebrorenal features or Dent disease. Dent disease is exclusively limited to the kidneys and specifically the proximal renal tubules. It is categorised as one cause of renal Fanconi syndrome and presents characteristic features such as low-molecular weight proteinuria, glycosuria, and phosphaturia. It was first linked to mutations in the *CLCN5* gene before *OCRL1* was discovered as a second disease gene [222,295]. As a potential disease mechanism, another enzyme with similar function as the one encoded by *ocrl1* might compensate for *ocrl1* disruption as it has been observed in mice. If this or a similar mechanism applies in human, variation in terms of the expression of the compensating enzyme amongst different tissues could explain the restricted Dent phenotype seen in some patients with particular *OCRL1* mutations, which differ from Lowe syndrome [296].

Studies characterising pleiotropic genes, SNPs and proteins helped to identify their contribution and impact on human disease [297]: About 12% of disease-causing proteins are considered of pleiotropic nature and typically involve rare rather than common genetic variation, as it is the case also for HIPKD. Interestingly, the majority of pleiotropic SNPs are coding mutations within exonic areas of the affected genes – in contrast to HIPKD. Thus, HIPKD is a rare example of genetic pleiotropy with involvement of a non-coding variant causing human disease.

The mechanisms of pleiotropic effects are not well understood but are likely highly versatile depending on the molecular and metabolic involvement of the encoded protein.

For HIPKD we expect the *PMM2* promoter mutation to cause a problem on the gene regulatory level that results in dysfunctional *PMM2* expression with an impact in the three affected organs, kidney, liver and pancreas only, thus implying a tissue-specific dysregulation of the gene. Tissue-specific enhancer-promoter interactions indeed have been described to promote pleiotropic effects during development with both enhancer and gene promoters enclosed within distinct highly conserved chromatin structures called topologically associating domains (TADs) or chromatin loops which are also described as “regulatory landscapes”: for example, the regulators of the vertebrate gene *Hoxd13*, which is needed for the development of digits or external genitals, are organised in such a TAD. Deletions of parts of this TAD have reduced the transcription rates of *Hoxd13* in digits compared to genital tubercles, thus validating 1) TADs as a gene regulatory unit for tissue-specific transcription and also 2) reduced gene expression due to altered transcriptional regulation [298]. Similar principles for the disease mechanism are assumed to be the case in the HIPKD cell types.

Recently, two interesting cases which describe pleiotropic diseases with parallels to HIPKD were reported. Gabbay et al. describe *PTF1A* pleiotropy: Homozygous truncating coding mutations in the *PTF1A* gene are known to cause pancreatic and cerebellar agenesis. However, they identified novel compound heterozygous mutations in the same gene, a truncating one on one allele and the other one in a pancreatic enhancer element. Together, these lead to the manifestation of pancreatic agenesis only [299]. The second study reported how a novel homozygous SNP creates a neo-ARE (a new AU-rich element) at the 3’-UTR of the *SLC4A4* gene which significantly reduces the expression of the gene and results in an isolated ocular phenotype whereas coding mutations in SLC4A4 cause a systemic phenotype which manifests in the eyes and the kidney and often induce also intellectual and growth abnormalities. They suggest that the neo-ARE interacts with a mRNA decay protein which destabilises the *SLC4A4* mRNA tissue-specific in the eye only due to the involvement of tissue-specific transcription factors of yet unknown identity [300].

In both cases pleiotropic effects 1) are based on non-coding mutations with 2) distinct disease phenotypes, due to 3) tissue-specific alteration of the regulation of the affected gene. It will be interesting to follow future investigations on the disease mechanisms of these two cases.

7.2 HIPKD and tissue-specific transcription

Various factors have their influence on the complex development of mammals. As outlined above, one of them is gene pleiotropy but also protein modification and diverse splicing patterns contribute to differentiation during development and later on. However, a major part influencing such processes relies on the genome and genomic regulation itself [301]. Although the genetic material within the nucleus of every single cell of an organism is virtually identical, the spatiotemporal gene expression amongst different tissues is highly versatile to allow the development of unique functions in each tissue. Such tissue-specific transcription is crucial for the development of multicellular organisms when stem cells actually differentiate into particular tissues and organs [302]. Transcriptional processes are highly organised and controlled by different regulatory elements (e.g. enhancer, repressors, activators, insulators, silencer, enhancing RNAs) which can be bound by transcription factors, epigenetic modifications (DNA methylation, histone modification) and also the chromatin landscape of particular regulatory TADs mentioned earlier [303]. It is thus not surprising that abnormalities of transcriptional regulation that change gene expression are often associated with human disease, as it is the case for HIPKD, for example, in cancer [304]. The involvement of such non-coding SNPs in human disease are well recognised since the early 2000's [305], however the association of gene regulation and disease-phenotypes and their mechanisms, in particular on the cell-specific level, raises many questions still to answer [306]. Thus, our study of HIPKD and our ambition to decipher its pathomechanism will shed light onto the molecular actions of non-coding disease-causing variants.

More than 1 million enhancer elements exist which are responsible to regulate about 20,000 protein-coding genes only. On average, that accounts for at least four enhancer elements which interact with a single gene in each cell type of the human body [307,308]. Hence, genetic variation and complex morphology is not predominantly driven by the variety of genes and their promoters but by their transcriptional control instead [309]. About 98% of the entire genome is occupied by non-coding DNA and around 40% of it are predicted to be of regulatory nature [310]. So-called gene deserts can stretch over distances in the megabase range and indeed long-range promoter interactions with distal elements are not unusual, forming transcriptional networks. Chromatin conformation studies have identified more than 1,000 of such long-range interactions independently in three different human cell lines [308]. Due to their high abundance, such regulatory regions are prone to disease-causing SNPs which can alter the expression level of the allele on which a transcription factor binding site is disrupted. This is easily measurable when such mutations appear in heterozygous state and was demonstrated for example in B-cells [311] and is also supported by our data with a significantly reduced expression of the allele controlled by the mutant promoter in the compound heterozygous HIPKD promoter variant. Hence, our HIPKD case provides an example of a SNP which co-localises with a regulatory element which is disease-associated in a tissue-specific fashion. However, it is not a long time ago when the non-coding part of the human DNA, which occupies a remarkable 98% of the entire human genome, was majorly ignored by scientists and declared as “Junk” DNA [312]. Nowadays, we are aware of its importance and the completion of the human genome project in 2003 eventually has allowed scientists to vastly explore the entire human DNA and generate huge sets of data for analysis, made processable thanks to modern genetics [313]. Since then SNPs have been identified to be the most common unit of genetic variation. The study of millions of such SNPs [314] and their integration in genome-wide association studies (GWAS) has helped to rapidly identify disease-associated variants and have shown that 93% of such variants indeed lie within the non-coding formerly called “Junk” DNA regions.

Maurano et al. performed elaborate GWAS-SNP analyses on disease-causing variants in the non-coding regulatory DNA [306]: In this respect, they found that multiple disease-causing variants predominantly enrich in transcription factor binding sequences and exert rather small effects – at least that is the case for common variants. Such variants are defined as small or weak effect variants which modulate the chromatin conformation in distinct areas rather than influencing entire regulatory networks. In addition, they have demonstrated the selective enrichment of disease-associated variants in particular cell types and tissues related to common diseases, for example the T-cell subtype Th17 in Crohn’s disease. Moreover, on the basis of DNAse I mapping to explore DNAse I hypersensitivity sites (DHSs; marker for regulatory DNA regions), they noted a distinct cell-specific localisation of multiple independent GWAS SNPs in disease-relevant cell types or tissues for many common diseases, suggesting a tissue-specific regulatory effect of such variants. They also pick up on what we today call TADs and how these are modified by transcription factor binding which results in “alterations of chromatin structure” as it was already reviewed and described in 1988 by Gross and Garrard, at the time termed “nucleosomal substructures” [315]. In this context, they also point out that the “weak” variants can change the “local chromatin architecture”. Although these observations are related to common variants the concept however is in favour of our hypothesis of an altered tissue-specific loop formation in HIPKD caused by the *PMM2* promoter variant in the ZNF143 binding site.

A recent study on how tissue-specific gene regulation works was performed by Sonawane et al., profiling genotype-tissue expression in 38 tissues [316]: They concluded that tissue-specific transcription is not so much dependent on individual tissue-specific transcription factors but rather on entire transcription networks formed of network edges (connections between transcription factors and target genes), network nodes (genes) and regulatory nodes (transcription factors). Knowing that only 2,000 genes encode for transcription factors it is unlikely that the multiple tissue-specific processes and functions rely solely on this small number of transcriptional active proteins. Instead, their data support the existence of global core networks shared by common and tissue-specific biological processes which are adjusted and tailored particularly to those unique processes by building “tissue-specific regulatory paths”. They postulate that it is the target-pattern which makes transcription factors exert their tissue-specific role rather than their differential expression. In conclusion, both tissue and non-tissue-specific transcription factors seem to mediate specific biological processes according to their data. This should be kept in mind whilst further investigating the HIPKD pathomechanism.

7.3 HIPKD and chromatin loop formation

Only a few years ago a more substantial analysis on the higher order of the organisation of the 3D genome within the nucleus was performed based on the Hi-C methodology [317]: These data sets have helped to identify TADs as frequent and fundamental chromatin structures which form large (megabase-sized) interactive, gene regulatory regions that allow for distinct interactions between gene promoters and their regulatory elements which otherwise were too far away from each other within the linear genome [318]. These TADs are highly conserved structures suggesting their early development during evolution [317]. Typically, these TAD and sub-domains are limited by CTCF motifs (CCCTC) which are bound by CTCF zinc finger proteins. CTCFs colocalise with TAD-stabilising cohesin ring structures [319] and are architectural proteins which separate chromatin regions and hence chromatin interactions. They function as insulators within gene regulation but also support distinct regulatory interactions within the TAD they are anchoring – they build regulatory hubs within the genome in which regulatory elements are enriched to facilitate promoter-enhancer interaction across long distances rather than just blocking unwanted interactions [224]. TADs can be further subdivided into smaller chromatin loops and sub-TADs which often form tissue-specific regulatory units. This was seen in a mouse model, where novel enhancer-promoter interactions formed after disturbing CTCF binding within such a sub-TAD [225]. Dysregulation of gene expression due to derangement of CTCF binding indeed has been observed in many different cell types and also as a cause for human disease [320,321].

This indicates the importance of TAD formation for gene regulation, in addition to enhancer elements and transcription factors and also reflects the complexity of the 3D chromosomal organisation [322].

We have identified that *PMM2* lies within a structural chromatin loop 220 kb in size which is limited by CTCF motifs in the HIPKD-relevant tissues as revealed by our *in silico* analysis for HepG2 cells (Figure 35). The interaction of CTCF with other architectural proteins has been well observed to form a variety of sub-loops or sub-TADs (also called “functional loops”), which facilitate tissue-specific interactions [323].

One such key architectural protein is ZNF143, which, as our results suggest, plays a key role in the HIPKD pathomechanism, as the *PMM2* promoter SNP is located right in the ZNF143 DNA binding site (Figure 31). Interestingly, Bailey et al. have identified ZNF143 as a novel chromatin-looping factor which is majorly involved in discrete chromatin formation to facilitate interactions between promoters and distal regulatory elements [226]: They have shown that ZNF143 directly binds to gene promoters to stimulate chromatin interactions. This agrees with our EMSA data which demonstrated that ZNF143 binds to the intact *PMM2* promoter [209]. Although it is ubiquitously expressed [324], this suggested that ZNF143 may also play a role in tissue-specific chromatin conformation. Indeed, this was confirmed by studying chromatin interactions of the same locus in different cell lines in a 3C chromosome conformation capture analysis. Moreover, and in favour of our hypothesis of altered chromatin loop formation, modifications within ZNF143 binding sites through SNPs have indeed been shown to significantly change its DNA-binding affinity with consequent altered chromatin interaction patterns [226]. This suggests the potential formation of disease-associated chromatin interactions when conformational control through ZNF143 is impaired. This, we suspect, is the case in HIPKD and is underpinned by our EMSA data indicating substantial impairment of ZNF143 binding to the defective *PMM2* promoter. Thus, transformation of the structural chromatin loop into a functional sub-loop by ZNF143 binding would enable normal *PMM2* regulation, whereas this would be prevented in the presence of the promoter mutation. Additional data on ZNF143 interaction with other factors also support this model of *PMM2* regulation [228]: ZNF143 highly colocalises with CTCF and cohesin. All three factors are expressed globally and are highly enriched at TAD boundary regions. Although the exact molecular mechanisms of the ZNF143/CTCF/cohesin interplay are not yet fully understood these findings do implicate their mutual function on 3D chromatin conformation and associated specific gene regulation [228].

Back to our HIPKD hypothesis, such a relationship between ZNF143, CTCF and cohesin would support the necessity of ZNF143 “pulling” the loop down through binding to the intact *PMM2* promoter and CTCF/cohesion. This would facilitate proximity of the *PMM2* promoter with a set of distinct regulatory elements at the anchors of the loop to enable specific spatiotemporal regulation of the gene.

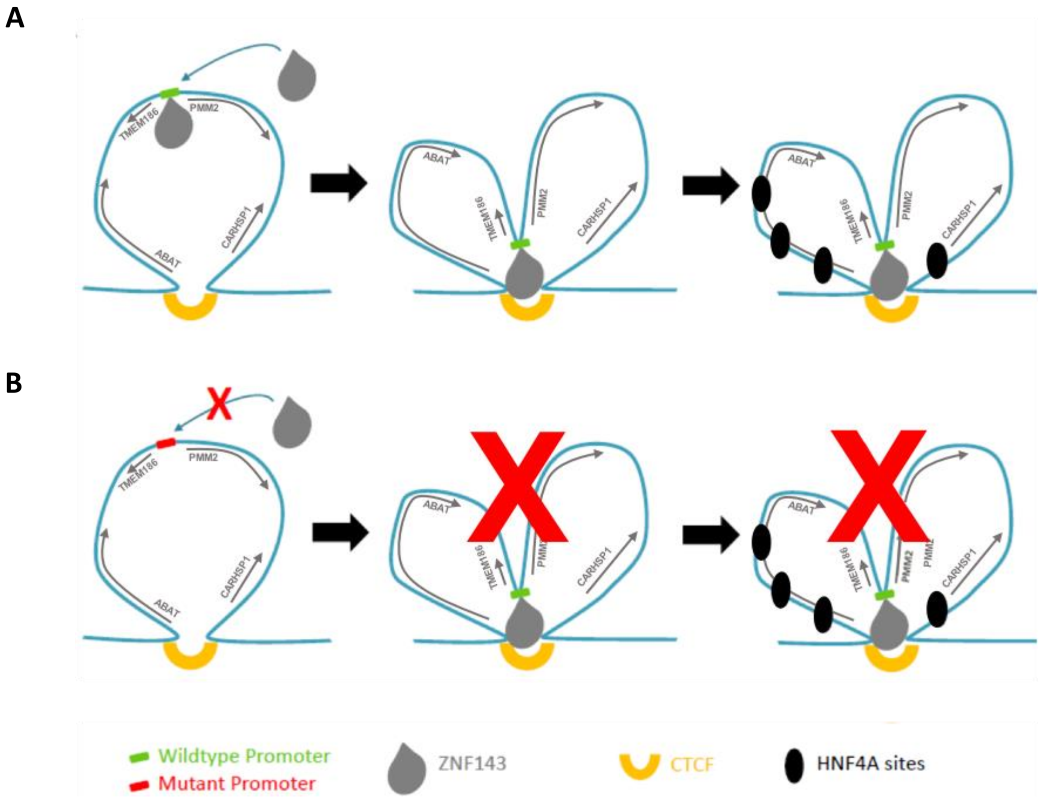


Figure 70: Tissue-specific chromatin loop formation suspected in healthy state (A) and HIPKD (B). Grey arrows mark the genes considered within the loop including *PMM2*.

As mentioned earlier, one such regulator facilitating tissue-specific *PMM2* expression could be the transcription factor HNF4A. HNF4A is an attractive candidate due to its predominant expression in kidney, liver and pancreas – the three organs affected in HIPKD [325]. As clusters of HNF4A sites have been found within the loop *in silico* this could explain the tissue-specific *PMM2* expression, which is impaired in HIPKD patients and responsible for the restricted phenotype.

The concept of specific promoter-enhancer interactions and related individual chromatin conformations as a key mechanism of tissue-specific gene expression was already proposed in 2000 and reviewed by Kleinjan et al. [326,327]. The model of such “functional gene expression modules” is shown in Figure 71.

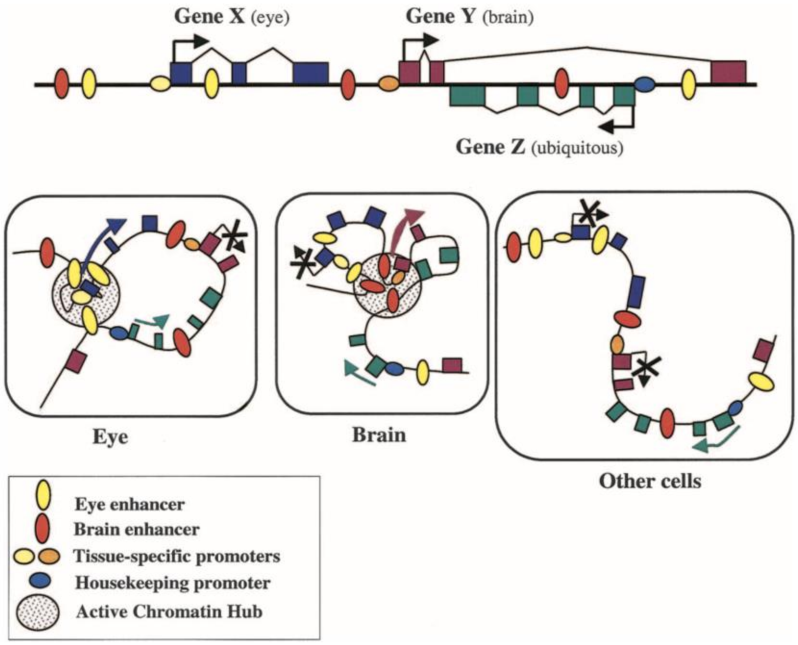


Figure 71: Putative model for the individual tissue-related regulation of different genes located in the same DNA region after Dillon and Sabattini 2000.
Three genes are located in the same DNA region. Gene X (eye) and gene Y (brain) are tissue-specific, gene Z (housekeeping gene) is ubiquitously expressed. The transcription of the genes X and Y depends on chromatin formation and different active enhancer elements interacting with gene promoters in eye and brain that are bound by different transcription factors. The housekeeping gene is transcribed in all tissues. Figure taken from Kleinjan 2005.

A question remains whether the expression of the other three genes located in the structural loop, *TMEM186*, *ABAT* and *CARHSP1*, might also be altered and thus potentially could also have an effect on HIPKD. This would be easiest to consider for *TMEM186,* which is regulated by the same promoter. Yet, the *ABAT* gene is also interesting, as it encodes the gamma-aminobutyrate transaminase which converts GABA (gamma-aminobutyric acid), a neurotransmitter into succinic semialdehyde. Interestingly, GABA enhances insulin secretion under low glucose conditions (as in HIPKD) in murine INS-1 cells [328]. Hence, one could speculate that altered loop formation in HIPKD could reduce *ABAT* expression, which in turn would increase GABA levels and thus enhances insulin secretion. However, an insulin-related phenotype has not been reported in patients with GABA-transaminase deficiency, a severe autosomal recessive disorder caused by biallelic mutations in the *ABAT* gene which damage the brain [329]. Unfortunately, the gene expression studies on all genes in the loop did not provide us with convincing results we trust due to poor input material and thus requires repetition in the future. This was outlined in chapter-specific discussion analysing the gene expression studies (4.4). However, the results obtained should not be ignored completely, especially for *CARHSP1* – the only gene for which the expression was significantly reduced in the compound heterozygous patient (Figure 39). With its two RNA binding domains, CARHSP1 is able to regulate crucial processes within protein biosynthesis, such as the regulation of ribosomal translation, mRNA degradation and the rate of transcription termination. Studies on TNF-α have demonstrated 1) CARHSP1 binding to the 3'UTR of TNF-α as well as 2) reduced protein production and mRNA levels of TNF-α when *CARHSP1* was knocked down [231]. This example reflects the negative effect impaired *CARHSP1* expression can have on the expression of other genes, which theoretically could also apply in the HIPKD case. However, insights regarding the distinct function of *CARHSP1* are limited as most research has been conducted on the signalling pathways which phosphorylate and dephosphorylate *CARHSP1* [330]. Thus, extensive research on this particular involvement of *CARHSP1* on transcriptional and translational processes is needed in general and also specifically concerning a major role within the HIPKD disease mechanism.

Furthermore, altered or reduced expression of the additional looping genes could potentially be caused by the predicted 3D conformational change of the chromatin loop in the HIPKD patient. As it was discussed in section 7.3, the 3D architecture of the genome indeed has an important influence on transcriptional regulation and could affect the expression of all genes in the loop. Imagining such a scenario, certain regulatory elements might not be able to interact with each other in the case of HIPKD to properly control gene expression in the three affected organs. So, one could theoretically argue that the phenotype seen in the patients may not or at least not solely be a result of impaired protein N-glycosylation due to reduced *PMM2* expression but also dysregulation of additional genes, e. g. *CARHSP1*. However, a substantial contribution of these genes is unlikely, considering that most patients are compound heterozygous for the promoter mutation and a coding *PMM2* mutation. Thus, expression of the other genes would only be changed from one allele and if this would be sufficient to cause disease, then the heterozygous carriers of the promoter mutation, without a coding mutation in trans (e.g. a parent) would be expected to also exhibit a phenotype. Hence, if the *PMM2* promoter mutation initiated reduced expression of *CARHSP1* in our assay (Figure 39), triggered by an altered chromatin loop formation and this results in the HIPKD phenotype, the heterozygous carriers of the c.-167G>T SNP must express the phenotype too – but they don't exhibit any medical restrictions. Therefore, in the compound heterozygous patients, *PMM2* is the only gene with clearly impaired function: a) by decreased expression from the allele with the promoter mutation and b) by the loss-of-function mutation on the other allele. Organ-specific *PMM2* dysfunction thus is still considered to be the critical mechanism causing HIPKD.

To finalise, additional interesting findings on how the shape of the genome contributes to gene regulation with its conformation were made quite recently: Studies on the molecular dynamics within the nucleus have demonstrated that most of the time, transcription factors actually remain in search mode before they find and bind to their respective site within the regulatory hubs. This diffusion along the chromatin fibres is apparently highly dependent on the conformation of the genome and is prone to changes in dependence of alterations in the genomic architecture [331]. This supports the idea that tissue-specific gene regulation can be impaired due to the rearrangement of the regulating chromatin induced by disease-causing SNPs, simply preventing transcription factors (like ZNF143) to find their targets.

The HIPKD cell model generated with CRIPSR-Cas9 during this PhD (see Chapter 6:) will help to further investigate the impact of dysregulated *PMM2* expression and hopefully enable investigation of the hypothesis of altered tissue-specific loop formation in HIPKD as a result of the *PMM2* promoter SNP as the key mechanism of this rare disease.

7.4 HIPKD and glycosylation

Knowing *PMM2* is involved early during posttranslational protein N-glycosylation [185], we clearly link the occurrence of HI and cystogenesis to defective N-glycosylation in the HIPKD patients. As almost half of all naturally occurring proteins are N-glycosylated [194], there are potentially multiple proteins which have their function in the physiology of the nephron and insulin secretion pathway that are affected by *PMM2* dysfunction. Hence, when improperly N-glycosylated many of those proteins are presumably involved in cyst formation and the excess secretion of insulin. Indeed, HI and cystic kidneys have been associated with deficient N-glycosylation within the broad range of multisystemic manifestation in CDG-1a. Renal involvement as one of multiple features within CDG was first described in detail in 1993 when five patients were examined, partially presenting multiple renal microcyst but also distinct larger cysts as observed by ultrasound and histology [332]. Additional cases describing renal involvement in CDG-1a have followed and kidney features have been added to the long list of organ manifestations seen in CDG [183].

HI is related to many types of developmental syndromes and its association with defects in N-glycosylation has been linked to three subtypes of CDG [333]. Three cases with persistent HI either in combination with additional typical features of CDG-1a or in one of the cases also in isolated format have been reported in 2009 [334]. The clinical phenotypes of all three patients during infancy was dominated by HI. However, all three individuals showed the serum transferrin pattern typical within CDG-1a caused by deficient *PMM2* function. Of note, in the patient with isolated HI, only a heterozygous mutation was identified, hence *PMM2* activity was reduced, but not absent. whether *PMM2* deficiency in this patient was truly the cause of HI is unclear and thus this case does not contradict the obligate neurological involvement in CDG-1a.

Regardless, the awareness of HI as a leading symptom of CDG-1a, prompted the inclusion for testing of CDG in all cases with HI to reduce the risk of misdiagnoses. HI as a symptom has also been noted in two other types of CDG namely CDG-1b [335]. and CDG-1d [336,337]. The molecular mechanisms of HI within CDG are not fully understood but are likely related to the membrane-bound K_{ATP} channels in β-cells as both major components of the channel, SUR1 (*ABCC8* gene) and Kir6.2 (*KCNJ11* gene) are N-glycosylated [333] and mutations in the encoding genes are responsible for congenital HI [338]. Indeed, mutagenesis studies on SUR1 have shown that when SUR1-N-glycosylation sites are mutated this impairs SUR1 trafficking to the β-cell membrane and retains the protein at the ER where it is assembled. This suggests direct involvement of defective SUR1 N-glycosylation in CDG due to reduced cell surface expression which causes HI [339]. In contradiction to these findings are the observations that many CDG patients and also the HIPKD patients do respond to diazoxide treatment to lower the amounts of insulin secretion which suggests that K_{ATP} channels are intact in those cases [336]. Other proteins heavily involved in glucose homeostasis and insulin secretion are the glucose transporters GLUT1 and GLUT2, both membrane-bound proteins which are N-glycosylated. However, solid scientific evidence for this theory to date is missing but depletion of N-glycosylation sites of GLUT1 resulted in a significant increased uptake of glucose in rodents which theoretically could promote enhanced insulin release [340]. Further genes and their protein products which evidently have significance in HI are not N-glycosylated and therefore unlikely to be targets of defective *PMM2* function. However, it should be kept in mind that around 40-50% of HI cases are of unknown genetic cause [338].

Several other studies have linked renal cystogenesis and defective N-glycosylation of key proteins: Rare cases of *GANAB* mutations for example which cause ADPKD and also autosomal dominant polycystic liver (ADPLD) have been described in 2016 [341]: Mutations in *GANAB* affect the encoded protein glucosidase IIα subunit (GIIα) which is crucial for maturation and correct intracellular localisation of PC-1 and PC-2 to the cell surface and cilia of the proximal renal tubular epithelial cells. If GIIα is intact, it trims immature proteins in the ER by cleaving off glucose residues to transform them into their mature format. Missense mutations in *GANAB* negatively affect the maturation of

PC-1 and reduced amounts of mature PC-1 have been observed in *GANAB*$^{+/-}$ cells. Thus, incorrect glycosylation patterns of PC-1 prevent it to exert its full function within the nephron which often leads to cyst formation. Dysfunctional PC-1 is a major cause for ADPKD and cystogenesis as outlined in the introduction of this thesis.

Aberrant PC-1 glycosylation as a cause for kidney cysts in mice has also been observed and was linked to mutations in the *Aqp11* gene. Furthermore, PC-2 levels, its localisation and stability are highly dependent on its N-glycosylation status. Thus, inadequate N-glycosylation of PC-2 is likely an additional reason for cystogenesis in ADPKD [223].

In conclusion, key proteins involved in cystic kidney disease, such as PC-1 (*PKD1*), PC-2 (*PKD2*) and also fibrocystin (*PKHD1*) are N-glycosylated and therefore are all potential candidates to be incorrectly glycosylated in HIPKD. In addition, the genetic causes for glomerular cyst are quite heterogeneous as discussed earlier, expanding the list of proteins which could be affected, uromodulin for example is one of them. Furthermore, considering that many signalling pathways and regulatory processes involved in renal development are affected in PKD and also the different pathways regulating insulin secretion, this further extends the number of proteins which might be potential targets of glycosylation defects in HIPKD.

A novel devastating subtype of CDG was also reported in 2016 [342]: Truncating mutations in the *ALG9* gene triggered a severe phenotype similar to the one seen in the Gillessen-Kaesbach-Nishimura syndrome presenting with multiple malformations, skeletal dysplasia and renal cysts. *ALG9* encodes the alpha-1,2-mannosyltransferase – an enzyme which is involved in the lipid-linked precursor oligosaccharide assembly during the N-glycosylation pathway.

The key proteins causative for polycystic liver disease, Glucosidase IIβ subunit (GIIβ, *PRKCSH*) and translocation protein SEC63 (*SEC63*), are involved in the N-glycosylation pathway in the ER. Thus, mutations in the respective genes result in dysfunctional SEC63 and GIIβ which affects the correct glycosylation of proteins important for liver development [343].

Evidence for the glycosylation defect we postulate for HIPKD has been observed in a mouse β-cell line: We have shown that insulin secretion seemed to be enhanced when mouse β-cell proteins were deglycosylated [209]. However, we could not confirm this effect with deglycosylated human pancreatic islets because variations in insulin measurements were too high to draw distinct conclusions. Thus, studies on defective insulin secretion based on the deficient *PMM2* expression in HIPKD will be subject for future investigations (8.3).

7.5 HIPKD cell line – improvement and alternatives

As outlined in section 6.4, the use of HepG2 cells involved some difficulties whilst performing the CRISPR-Cas9 experiments. Apart from those technical issues, we were aware that HepG2 is surely not an ideal cell line for the purpose of this study as it is a cancer cell line knowing to be aneuploid (55 chromosomes on average) and genomically instable. Indeed, there is evidence that almost the entire chromosome 16 underwent a copy number gain in HepG2 and therefore the targeted locus likely is present more than twice in HepG2 [344]. Luckily, this would not hinder a sufficient performance of a chromatin conformation study, as long as CRISPR-Cas9 introduces the c.-167G>T SNP in homozygous state (personal communication). However, as we were restricted to work either with a human kidney, β-cell or at least a liver cell line, HepG2 was the best choice to start with as we were not aware of a reliable human β-cell line. While there are human kidney cell lines, the difficulty is that we do not know which kidney cells are affected by the promoter mutation: there are numerous different cell types in the kidney, ranging from highly specialised glomerular and tubular epithelial cells to mesenchymal cells and we currently do not know, which of these are affected. Without this information, choosing a specific cell line is impossible. Obviously, a similar problem arises for liver cells, as the liver also contains multiple different cell types. But, in support for the HepG2 cells was also the result of the *in-silico* analysis (Figure 35) reflecting HepG2 to be the only cell line in which the chromatin seems to be active throughout the predicted loop on chromosome 16. Unfortunately, no data from a human kidney or β-cell line were available to include in this chromatin segmentation analysis, therefore HepG2 was the best candidate cell line to use at this stage.

Other options to consider are the use of human iPS cells or to isolate the particular cell type (e. g. glomerulus) needed from the primary renal cell culture of the control kidney used in this study. Both versions could then be CRISPRed to generate the desired disease model. This would also allow to overcome the issues related to the gene expression studies outlined above when using end-stage kidney material form the compound heterozygous patient (4.4). However, the disadvantage of primary cells is their limited lifespan and ability to proliferate before they reach senescence [345]. A solution could be to immortalise the isolated cells from primary culture to keep the cell cycle active and cells dividing. As primary cells are difficult to transfect, an immortalisation is most efficient through viral infection to introduce certain genetic changes to guarantee for example consistent telomerase activity. But genetically modifying a genome always bears the risk of changing the phenotype of the targeted cells whilst going through the immortalisation process, which could impact on the experimental outcome [346].

Another time-saving alternative instead of editing human iPS cells or for example glomerular cells from primary culture is the use of stem cells obtained directly from one of the patients. This would save the labour of gene editing completely. These stem cells then could be programmed into iPS cells and afterwards differentiated into β-cells for insulin secretion studies and kidney cells for gene expression analysis and a capture C analysis to study whether altered chromatin conformation of the *PMM2* loop is the case in HIPKD. However, this approach will not solve the issue of limited cell proliferation once the cells are differentiated. This in particular is a problem to sufficiently conduct a capture C analysis which requires incredibly large cell populations (at least 10 million cells per sample per experiment) to generate solid and reproducible interaction maps with high resolution [347]. Thus, generating 3C libraries (the first crucial step in capture C) from rare tissues and cell types is nearly impossible. However, recent advances from the pioneers in the field of chromatin conformation capture produced high quality interaction profiles from only 10,000 – 20,000 cells [348]. This is of immense use for the scientific community to explore how structural interactions influence gene activation und could be an option to study the potential chromatin conformation change suggested in HIPKD (Figure 36).

The use of CRISPR RNPs and electroporation has tremendously improved the outcome of the tedious process producing this HIPKD cell model which failed using CRISPR-Cas9 expression plasmids and reagent-based transfection (Chapter 6:). What initially failed for 6 months worked out in the end within 1.5 months (electroporation to identification of homozygous single cell HIPKD clone) after changing the format of the system. Clearly, this indicates CRISPR RNPs as the better choice to perform a successful CRISPR knock-in through the HDR repair pathway, which is known to be a very inefficient process. Scientist are continuously working on improving the CRISPR technology to overcome such restrictions and are in particular keen on solving the problem of the yet inefficient CRISPR knock-in process. Big advances have been made at least at the stage of directed single base pair substitutions. “CRISPR base editors” catapult modern gene editing to its next level [349,350]: A catalytically inactive so called “dead” Cas9 (dCas9) fused with nucleobase deaminases is capable to convert A·T to G·C and G·C to A·T base pairs allowing programmable base editing without creating a double strand break and the use of an HDR donor template. The efficiency of those adenine (e.g. ABE7.10) and cytidine (e.g. BE4) base editors is reported to be an impressive 50% and reveals high product purity in human cells. In addition, the system significantly reduces the formation of indels to 0.1%. This approach demonstrates a powerful tool to precisely correct disease-causing point mutations and can be used to efficiently generate disease models for monogenic diseases, such as HIPKD. To date, the system is tied to some limitations as it is only able to perform base pair transitions (change of purine base to other purine base or pyrimidine to other pyrimidine base) but not transversions (purine changed to pyrimidine and vice versa). Therefore, only the four base pair transitions C-to-T, G-to-A, A-to-G, and T-to-C are yet possible to achieve with the system. Hence, in the case of HIPKD, the base editors available would only introduce a G-to-A instead of the desired G-to-T SNP at position -167 within the *PMM2* promoter due to the purin→purine/pyrimidine→pyrimidine restriction. Therefore, such base editors are not advantageous to use in the particular HIPKD case. However, vast enhancements on the base pair editing system are emerging and scientists work hard to improve efficiency and precision of the system and foremost to reduce or ideally eliminate the risk of unwanted off-target single base substitutions [351]. Surely, there is more exciting development within the field of cutting-edge genome editing technology to come in the near future which will help us to conquer current restrictions to the CRISPR technology.

Chapter 8: Future work

8.1 Gene expression studies with HIPKD cell line

The expression levels of all four genes (*PMM2*, *TMEM186*, *CARHSP1*, *ABAT*) within the predicted chromatin loop plus the expression of two control genes (*PPIA*, *CDH1*) were analysed in a dPCR setting on the basis of primary renal cells obtained from one of the HIPKD patients. However, the interpretation of the results appeared to be difficult and the analysis needs repetition. All six gene expression assays used in the digital PCR setting should be repeated using the new HIPKD cell model which was generated in this study and the results compared to a control. This will exclude the issues related to the quality of the cell material used in the previous setting (4.4).

8.2 Verification of the glycosylation status of proteins in HIPKD

A simple test to assess the overall glycosylation status in HIPKD could be performed with quantitative proteomic for example within an IEF assay. Isolated proteins taken from HIPKD and wild type control cells could be used for such an assay. If N-glycosylation is significantly affected in HIPKD this should be visible due to a different migration pattern of this sample within the electric field.

To focus on the glycosylation status of particular proteins, for example the SUR1 and Kir6.2 subunits of the K_{ATP} channel, potential differences could be observed with SDS-PAGE and Western blotting using specific antibodies. A shift in the migration pattern between mutant and control sample should be noticeable if the N-glycosylation pattern clearly differs.

8.3 Continue studying the impact of defective glycosylation on insulin secretion in HIPKD

The functional consequences of a potentially altered protein glycosylation in HIPKD is still questionable and needs further investigation. Although the use of human pancreatic islets seems to be a promising approach to measure insulin secretion, the experimental settings would require significant improvement in order to generate reliable data to overcome the errors which are related to the work with pancreatic islets (5.4). To prevent those errors a reliable approach is to generate a HIPKD β-cell line. This would allow insulin measurement without having to manipulate the cells with the deglycosylation enzymes to mimic a glycosylation effect. Such a cell line can be generated after the same principle as it was successfully applied to HepG2 cell using the CRISPR-Cas9 system. As we did not produce reliable data with commercially available human β-cell lines in previous studies, a good alternative to use would human inducible pluripotent stem cells (iPSC). The HIPKD promoter mutation could be introduced in those iPSC which then could be differentiated into β-cells [352,345]. Changes in insulin secretion could then be assessed in a glucose-induced insulin secretion assay exposing the HIPKD β-cells to low and high glucose concentrations and compare the result to a control group.

It could also be considered to go one step further and try to re-aggregate engineered β-cells into pseudo islets so that they would be closer to their natural environment. Studies on re-aggregated islets have shown that this has improved β-cells performance [354,355].

8.4 Validation of the HIPKD cell model

Prior to proceeding work with the new CRISPR engineered HIPKD cell model, its suitability should be verified. A ChIP-qPCR assay could help to assess ZNF143 binding to its binding site within the *PMM2* promoter in the HIPKD cell line compared to a wild type control [356]. This would quantify ZNF143 interactions with the *PMM2* promoter and would indicate significant differences in the mutant if there are some.

8.5 3D chromatin conformation in HIPKD

Tissue-specific altered chromatin loop formation due to the impaired binding of ZNF143 based on the *PMM2* promoter SNP was suggested to be the disease mechanism in HIPKD. This hypothesis needs to be validated and a Capture C analysis could investigate this further to eventually shed light onto the 3D architecture within HIPKD. This method has been successfully applied to analyse interactions between gene promoters and regulatory elements in detail at any given locus to understand the 3D folding-characteristics of the human genome. Its high throughput capacities allow to look at multiple specific interactions within only one experiment to 1) clarify the principles by which gene expression is controlled which is highly depended on inter- and intragenic interactions and 2) to enable scientists to investigate the functional effects of SNPs located in regulatory elements which are related to human diseases [347].

The new HIPKD cell model can be used in such an analysis to investigate the interactions of the *PMM2* promoter with regulatory elements, especially HNF4A, and differences compared to a wild type control can be assessed. In addition, this experiment should also be done with cells which originate from a tissue which is not affected by HIPKD to validate the tissue-specific loop formation enclosing *PMM2* expected in kidney, liver and pancreas.

8.6 Conclusion

This study provides evidence that the *PMM2* promoter mutation clearly reduces the expression of the *PMM2* gene at the allelic level in HIPKD patients. This finding contributes to the overall assumption that disrupted ZNF143 binding results in dysfunctional *PMM2* expression in HIPKD which is thought to be associated with impaired tissue-specific *PMM2* gene regulation in particular in the three affected organs. To be able to investigate this further, a HIPKD cell model was successfully generated with CRISPR-Cas9 and HepG2 cells which have the c.-167G>T SNP stably introduced in homozygous state. Preliminary data were obtained addressing a glycosylation defect, which causes raised insulin secretion in HIPKD. A trend supporting this expectation of increased insulin secretion in HIPKD has been observed but needs approval in the future.

Overall, the findings of this study and the discovery of HIPKD extent the genetic cause for HI and PKD and also provide an additional example of how non-coding mutations have their impact in human disease. In addition, the study reflects the importance of including non-coding variants in the genetics of human disease. Furthermore, it creates awareness on how appropriate gene regulation and expression substantially relies on the higher organisation of the human genome in the 3-dimensional space.

References

1 O'Brien, L. L. & McMahon, A. P. Induction and patterning of the metanephric nephron. *Semin Cell Dev Biol* **36**, 31-38, doi:10.1016/j.semcdb.2014.08.014 (2014).

2 Bertram, J. F., Douglas-Denton, R. N., Diouf, B., Hughson, M. D. & Hoy, W. E. Human nephron number: implications for health and disease. *Pediatr Nephrol* **26**, 1529-1533, doi:10.1007/s00467-011-1843-8 (2011).

3 Reiser, J. & Altintas, M. M. Podocytes. *F1000Res* **5**, doi:10.12688/f1000research.7255.1 (2016).

4 Deshmukh, S. R. The renal system explained. *Nottingham University Press* (2009).

5 Quaggin, S. E. & Kreidberg, J. A. Development of the renal glomerulus: good neighbors and good fences. *Development* **135**, 609-620, doi:10.1242/dev.001081 (2008).

6 Zhuo, J. L. & Li, X. C. Proximal nephron. *Compr Physiol* **3**, 1079-1123, doi:10.1002/cphy.c110061 (2013).

7 Boron, W. F. Medical physiology. *Elsevier* (2009).

8 Jaffe, A., Zurynski, Y., Beville, L. & Elliott, E. Call for a national plan for rare diseases. *J Paediatr Child Health* **46**, 2-4, doi:10.1111/j.1440-1754.2009.01608.x (2010).

9 Graham, E. C. Current trends in biobanking for rare diseases: a review. *Journal of Biorepository Science for Applied Medicine* (2014).

10 Schieppati, A., Henter, J. I., Daina, E. & Aperia, A. Why rare diseases are an important medical and social issue. *Lancet* **371**, 2039-2041, doi:10.1016/S0140-6736(08)60872-7 (2008).

11 Devuyst, O. *et al.* Rare inherited kidney diseases: challenges, opportunities, and perspectives. *Lancet* **383**, 1844-1859, doi:10.1016/S0140-6736(14)60659-0 (2014).

12 Soliman, N. A. Orphan kidney diseases. *Nephron Clin Pract* **120**, c194-199, doi:10.1159/000339785 (2012).

13 Wastfelt, M., Fadeel, B. & Henter, J. I. A journey of hope: lessons learned from studies on rare diseases and orphan drugs. *J Intern Med* **260**, 1-10, doi:10.1111/j.1365-2796.2006.01666.x (2006).

14 Gericke, C. A., Riesberg, A. & Busse, R. Ethical issues in funding orphan drug research and development. *J Med Ethics* **31**, 164-168, doi:10.1136/jme.2003.007138 (2005).

15 Bockenhauer, D. *et al.* Epilepsy, ataxia, sensorineural deafness, tubulopathy, and KCNJ10 mutations. *N Engl J Med* **360**, 1960-1970, doi:10.1056/NEJMoa0810276 (2009).

16 Aymè, S. Report on Rare Disease Research, Its Determinants in Europe and the Way Forward. *Paris: INSERM* (2011).

17 Loftus, H. & Ong, A. C. Cystic kidney diseases: many ways to form a cyst. *Pediatr Nephrol* **28**, 33-49, doi:10.1007/s00467-012-2221-x (2013).

18 Igarashi, P. & Somlo, S. Genetics and pathogenesis of polycystic kidney disease. *J Am Soc Nephrol* **13**, 2384-2398 (2002).

19 Song, X. *et al.* Systems biology of autosomal dominant polycystic kidney disease (ADPKD): computational identification of gene expression pathways and integrated regulatory networks. *Hum Mol Genet* **18**, 2328-2343, doi:10.1093/hmg/ddp165 (2009).

20 Martinez, J. R. & Grantham, J. J. Polycystic kidney disease: etiology, pathogenesis, and treatment. *Dis Mon* **41**, 693-765 (1995).

21 Pei, Y. *et al.* Bilineal disease and trans-heterozygotes in autosomal dominant polycystic kidney disease. *Am J Hum Genet* **68**, 355-363 (2001).

22 Harris, P. C. Autosomal dominant polycystic kidney disease: clues to pathogenesis. *Hum Mol Genet* **8**, 1861-1866 (1999).

23 Cornec-Le Gall, E., Audrezet, M. P., Le Meur, Y., Chen, J. M. & Ferec, C. Genetics and pathogenesis of autosomal dominant polycystic kidney disease: 20 years on. *Hum Mutat* **35**, 1393-1406, doi:10.1002/humu.22708 (2014).

24 Hateboer, N. *et al.* Comparison of phenotypes of polycystic kidney disease

types 1 and 2. European PKD1-PKD2 Study Group. *Lancet* **353**, 103-107 (1999).

25 Mochizuki, T. *et al.* PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* **272**, 1339-1342 (1996).

26 Hildebrandt, F. & Otto, E. Cilia and centrosomes: a unifying pathogenic concept for cystic kidney disease? *Nat Rev Genet* **6**, 928-940, doi:10.1038/nrg1727 (2005).

27 Streets, A. J. *et al.* Functional analysis of PKD1 transgenic lines reveals a direct role for polycystin-1 in mediating cell-cell adhesion. *J Am Soc Nephrol* **14**, 1804-1815 (2003).

28 Ibraghimov-Beskrovnaya, O. *et al.* Polycystin: in vitro synthesis, in vivo tissue expression, and subcellular localization identifies a large membrane-associated protein. *Proc Natl Acad Sci U S A* **94**, 6397-6402 (1997).

29 Paul, B. M. & Vanden Heuvel, G. B. Kidney: polycystic kidney disease. *Wiley Interdiscip Rev Dev Biol* **3**, 465-487, doi:10.1002/wdev.152 (2014).

30 Delmas, P. *et al.* Gating of the polycystin ion channel signaling complex in neurons and kidney cells. *FASEB J* **18**, 740-742, doi:10.1096/fj.03-0319fje (2004).

31 Yamaguchi, T., Hempson, S. J., Reif, G. A., Hedge, A. M. & Wallace, D. P. Calcium restores a normal proliferation phenotype in human polycystic kidney disease epithelial cells. *J Am Soc Nephrol* **17**, 178-187, doi:10.1681/ASN.2005060645 (2006).

32 Lin, F. *et al.* Kidney-specific inactivation of the KIF3A subunit of kinesin-II inhibits renal ciliogenesis and produces polycystic kidney disease. *Proc Natl Acad Sci U S A* **100**, 5286-5291, doi:10.1073/pnas.0836980100 (2003).

33 Harris, P. C. & Torres, V. E. Polycystic kidney disease. *Annu Rev Med* **60**, 321-337, doi:10.1146/annurev.med.60.101707.125712 (2009).

34 Zhou, J. Polycystins and primary cilia: primers for cell cycle progression. *Annu Rev Physiol* **71**, 83-113, doi:10.1146/annurev.physiol.70.113006.100621 (2009).

35 Nauli, S. M. *et al.* Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nat Genet* **33**, 129-137, doi:10.1038/ng1076 (2003).

36 Delmas, P. *et al.* Constitutive activation of G-proteins by polycystin-1 is antagonized by polycystin-2. *J Biol Chem* **277**, 11276-11283, doi:10.1074/jbc.M110483200 (2002).

37 Grantham, J. J. Polycystic kidney disease: from the bedside to the gene and back. *Curr Opin Nephrol Hypertens* **10**, 533-542 (2001).

38 Parnell, S. C. *et al.* Polycystin-1 activation of c-Jun N-terminal kinase and AP-1 is mediated by heterotrimeric G proteins. *J Biol Chem* **277**, 19566-19572, doi:10.1074/jbc.M201875200 (2002).

39 Horsley, V. & Pavlath, G. K. NFAT: ubiquitous regulator of cell differentiation and adaptation. *J Cell Biol* **156**, 771-774, doi:10.1083/jcb.200111073 (2002).

40 Puri, S. *et al.* Polycystin-1 activates the calcineurin/NFAT (nuclear factor of activated T-cells) signaling pathway. *J Biol Chem* **279**, 55455-55464, doi:10.1074/jbc.M402905200 (2004).

41 Li, X. *et al.* Polycystin-1 and polycystin-2 regulate the cell cycle through the helix-loop-helix inhibitor Id2. *Nat Cell Biol* **7**, 1202-1212, doi:10.1038/ncb1326 (2005).

42 Weimbs, T., Olsan, E. E. & Talbot, J. J. Regulation of STATs by polycystin-1 and their role in polycystic kidney disease. *JAKSTAT* **2**, e23650, doi:10.4161/jkst.23650 (2013).

43 Dere, R., Wilson, P. D., Sandford, R. N. & Walker, C. L. Carboxy terminal tail of polycystin-1 regulates localization of TSC2 to repress mTOR. *PLoS One* **5**, e9239, doi:10.1371/journal.pone.0009239 (2010).

44 Lal, M. *et al.* Polycystin-1 C-terminal tail associates with beta-catenin and inhibits canonical Wnt signaling. *Hum Mol Genet* **17**, 3105-3117, doi:10.1093/hmg/ddn208 (2008).

45 Patel, V. *et al.* Acute kidney injury and aberrant planar cell polarity induce cyst formation in mice lacking renal cilia. *Hum Mol Genet* **17**, 1578-1590, doi:10.1093/hmg/ddn045 (2008).

46 Fischer, E. *et al.* Defective planar cell polarity in polycystic kidney disease. *Nat Genet* **38**, 21-23, doi:10.1038/ng1701 (2006).

47 Simons, M. *et al.* Inversin, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signaling pathways. *Nat Genet* **37**, 537-543, doi:10.1038/ng1552 (2005).

48 Ibraghimov-Beskrovnaya, O. & Bukanov, N. Polycystic kidney diseases: from molecular discoveries to targeted

therapeutic strategies. *Cell Mol Life Sci* **65**, 605-619, doi:10.1007/s00018-007-7362-x (2008).

49 Chapin, H. C. & Caplan, M. J. The cell biology of polycystic kidney disease. *J Cell Biol* **191**, 701-710, doi:10.1083/jcb.201006173 (2010).

50 Yamaguchi, T. *et al.* Calcium restriction allows cAMP activation of the B-Raf/ERK pathway, switching cells to a cAMP-dependent growth-stimulated phenotype. *J Biol Chem* **279**, 40419-40430, doi:10.1074/jbc.M405079200 (2004).

51 Torres, V. E. Apoptosis in cystogenesis: hands on or hands off? *Kidney Int* **55**, 334-335, doi:10.1046/j.1523-1755.1999.00262.x (1999).

52 Pei, Y. A "two-hit" model of cystogenesis in autosomal dominant polycystic kidney disease? *Trends Mol Med* **7**, 151-156 (2001).

53 Reeders, S. T. Multilocus polycystic disease. *Nat Genet* **1**, 235-237, doi:10.1038/ng0792-235 (1992).

54 Brasier, J. L. & Henske, E. P. Loss of the polycystic kidney disease (PKD1) region of chromosome 16p13 in renal cyst cells supports a loss-of-function model for cyst pathogenesis. *J Clin Invest* **99**, 194-199, doi:10.1172/JCI119147 (1997).

55 Qian, F., Watnick, T. J., Onuchic, L. F. & Germino, G. G. The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell* **87**, 979-987 (1996).

56 Leonhard, W. N. *et al.* Scattered Deletion of PKD1 in Kidneys Causes a Cystic Snowball Effect and Recapitulates Polycystic Kidney Disease. *J Am Soc Nephrol* **26**, 1322-1333, doi:10.1681/ASN.2013080864 (2015).

57 Lantinga-van Leeuwen, I. S. *et al.* Lowering of Pkd1 expression is sufficient to cause polycystic kidney disease. *Hum Mol Genet* **13**, 3069-3077, doi:10.1093/hmg/ddh336 (2004).

58 Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-226, doi:10.1016/j.cell.2008.09.050 (2008).

59 Hopp, K. *et al.* Functional polycystin-1 dosage governs autosomal dominant polycystic kidney disease severity. *J Clin Invest* **122**, 4257-4273, doi:10.1172/JCI64313 (2012).

60 Gunay-Aygun, M. Liver and kidney disease in ciliopathies. *Am J Med Genet C Semin Med Genet* **151C**, 296-306, doi:10.1002/ajmg.c.30225 (2009).

61 Denamur, E. *et al.* Genotype-phenotype correlations in fetuses and neonates with autosomal recessive polycystic kidney disease. *Kidney Int* **77**, 350-358, doi:10.1038/ki.2009.440 (2010).

62 Desmet, V. J. Congenital diseases of intrahepatic bile ducts: variations on the theme "ductal plate malformation". *Hepatology* **16**, 1069-1083 (1992).

63 Ward, C. J. *et al.* The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. *Nat Genet* **30**, 259-269, doi:10.1038/ng833 (2002).

64 Rossetti, S. *et al.* A complete mutation screen of PKHD1 in autosomal-recessive polycystic kidney disease (ARPKD) pedigrees. *Kidney Int* **64**, 391-403, doi:10.1046/j.1523-1755.2003.00111.x (2003).

65 Menezes, L. F. *et al.* Polyductin, the PKHD1 gene product, comprises isoforms expressed in plasma membrane, primary cilium, and cytoplasm. *Kidney Int* **66**, 1345-1355, doi:10.1111/j.1523-1755.2004.00844.x (2004).

66 Zhang, J. *et al.* Polycystic kidney disease protein fibrocystin localizes to the mitotic spindle and regulates spindle bipolarity. *Hum Mol Genet* **19**, 3306-3319, doi:10.1093/hmg/ddq233 (2010).

67 Sun, L., Wang, S., Hu, C. & Zhang, X. Down-regulation of PKHD1 induces cell apoptosis through PI3K and NF-kappaB pathways. *Exp Cell Res* **317**, 932-940, doi:10.1016/j.yexcr.2011.01.025 (2011).

68 Wang, S. *et al.* Fibrocystin/polyductin, found in the same protein complex with polycystin-2, regulates calcium responses in kidney epithelia. *Mol Cell Biol* **27**, 3241-3252, doi:10.1128/MCB.00072-07 (2007).

69 Hiesberger, T. *et al.* Proteolytic cleavage and nuclear translocation of fibrocystin is regulated by intracellular Ca2+ and activation of protein kinase C. *J Biol Chem* **281**, 34357-34364, doi:10.1074/jbc.M606740200 (2006).

70 Bernstein, J. & Landing, B. H. Glomerulocystic kidney diseases. *Prog Clin Biol Res* **305**, 27-43 (1989).

71 Lennerz, J. K., Spence, D. C., Iskandar, S. S., Dehner, L. P. & Liapis, H. Glomerulocystic kidney: one hundred-year perspective. *Arch Pathol Lab Med* **134**, 583-605, doi:10.1043/1543-2165-134.4.583 (2010).

72 Bissler, J. J., Siroky, B. J. & Yin, H. Glomerulocystic kidney disease. *Pediatr Nephrol* **25**, 2049-2056; quiz 2056-2049, doi:10.1007/s00467-009-1416-2 (2010).

73 Hiesberger, T. *et al.* Mutation of hepatocyte nuclear factor-1beta inhibits Pkhd1 gene expression and produces renal cysts in mice. *J Clin Invest* **113**, 814-825, doi:10.1172/JCI20083 (2004).

74 Roder, P. V., Wu, B., Liu, Y. & Han, W. Pancreatic regulation of glucose homeostasis. *Exp Mol Med* **48**, e219, doi:10.1038/emm.2016.6 (2016).

75 Caicedo, A. in *Seminars in cell & developmental biology.* 11-21 (Elsevier).

76 Ionescu-Tirgoviste, C. *et al.* A 3D map of the islet routes throughout the healthy human pancreas. *Sci Rep* **5**, 14634, doi:10.1038/srep14634 (2015).

77 Chandra, R. & Liddle, R. A. Neural and hormonal regulation of pancreatic secretion. *Curr Opin Gastroenterol* **25**, 441-446, doi:10.1097/MOG.0b013e32832e9c41 (2009).

78 Jansson, L. & Hellerstrom, C. Glucose-induced changes in pancreatic islet blood flow mediated by central nervous system. *Am J Physiol* **251**, E644-647, doi:10.1152/ajpendo.1986.251.6.E644 (1986).

79 Langerhans, P. *Beitrage zur mikroskopischen Anatomie der Bauchspeichel Drüse.*, Friedrich-Wilhelms-Universität zu Berlin, (1869).

80 Pisania, A. *et al.* Quantitative analysis of cell composition and purity of human pancreatic islet preparations. *Lab Invest* **90**, 1661-1675, doi:10.1038/labinvest.2010.124 (2010).

81 Cabrera, O. *et al.* The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci U S A* **103**, 2334-2339, doi:10.1073/pnas.0510790103 (2006).

82 Bosco, D. *et al.* Unique arrangement of alpha- and beta-cells in human islets of Langerhans. *Diabetes* **59**, 1202-1210, doi:10.2337/db09-1177 (2010).

83 Wojtusciszyn, A., Armanet, M., Morel, P., Berney, T. & Bosco, D. Insulin secretion from human beta cells is heterogeneous and dependent on cell-to-cell contacts. *Diabetologia* **51**, 1843-1852, doi:10.1007/s00125-008-1103-z (2008).

84 Shapiro, A. M. *et al.* Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Engl J Med* **343**, 230-238, doi:10.1056/NEJM200007273430401 (2000).

85 Unger, R. H. Glucagon physiology and pathophysiology. *N Engl J Med* **285**, 443-449, doi:10.1056/NEJM197108192850806 (1971).

86 Freychet, L. *et al.* Effect of intranasal glucagon on blood glucose levels in healthy subjects and hypoglycaemic patients with insulin-dependent diabetes. *Lancet* **1**, 1364-1366 (1988).

87 Komatsu, M., Takei, M., Ishii, H. & Sato, Y. Glucose-stimulated insulin secretion: A newer perspective. *J Diabetes Investig* **4**, 511-516, doi:10.1111/jdi.12094 (2013).

88 Khan, A. H. & Pessin, J. E. Insulin regulation of glucose uptake: a complex interplay of intracellular signalling pathways. *Diabetologia* **45**, 1475-1483, doi:10.1007/s00125-002-0974-7 (2002).

89 Srivastava, A. K. & Pandey, S. K. Potential mechanism(s) involved in the regulation of glycogen synthesis by insulin. *Mol Cell Biochem* **182**, 135-141 (1998).

90 Dimitriadis, G. D. *et al.* Effects of hyperthyroidism on the sensitivity of glycolysis and glycogen synthesis to insulin in the soleus muscle of the rat. *Biochem J* **253**, 87-92 (1988).

91 Shepherd, P. R. & Kahn, B. B. Glucose transporters and insulin action--implications for insulin resistance and diabetes mellitus. *N Engl J Med* **341**, 248-257, doi:10.1056/NEJM199907223410406 (1999).

92 Larner, J. in *Biochemistry of the Glycosidic Linkage: An Integrated View* 597-619 (Elsevier, 1972).

93 Newsholme, E. A. & Dimitriadis, G. Integration of biochemical and physiologic effects of insulin on glucose metabolism. *Exp Clin Endocrinol Diabetes* **109 Suppl 2**, S122-134, doi:10.1055/s-2001-18575 (2001).

94 Manchester, K. Insulin and protein synthesis. *Biochemical Actions of Hormones* **1**, 267-320 (1970).

95 Haeusler, R. A., McGraw, T. E. & Accili, D. Biochemical and cellular properties of insulin receptor signalling. *Nat Rev Mol Cell Biol* **19**, 31-44, doi:10.1038/nrm.2017.89 (2018).

96 Mackenzie, R. W. & Elliott, B. T. Akt/PKB activation and insulin signaling: a novel insulin signaling pathway in the treatment

of type 2 diabetes. *Diabetes, metabolic syndrome and obesity: targets and therapy* **7**, 55 (2014).

97 Cynober, L. A. Plasma amino acid levels with a note on membrane transport: characteristics, regulation, and metabolic significance. *Nutrition* **18**, 761-766 (2002).

98 Carlson, M. G., Snead, W. L. & Campbell, P. J. Regulation of free fatty acid metabolism by glucagon. *J Clin Endocrinol Metab* **77**, 11-15, doi:10.1210/jcem.77.1.8100827 (1993).

99 Vons, C. *et al.* Regulation of fatty-acid metabolism by pancreatic hormones in cultured human hepatocytes. *Hepatology* **13**, 1126-1130 (1991).

100 Parmley, W. W., Glick, G. & Sonnenblick, E. H. Cardiovascular effects of glucagon in man. *N Engl J Med* **279**, 12-17, doi:10.1056/NEJM196807042790103 (1968).

101 Farah, A. E. Glucagon and the circulation. *Pharmacol Rev* **35**, 181-217 (1983).

102 Banarer, S., McGregor, V. P. & Cryer, P. E. Intraislet hyperinsulinemia prevents the glucagon response to hypoglycemia despite an intact autonomic response. *Diabetes* **51**, 958-965 (2002).

103 Taborsky, G. J., Jr., Ahren, B. & Havel, P. J. Autonomic mediation of glucagon secretion during hypoglycemia: implications for impaired alpha-cell responses in type 1 diabetes. *Diabetes* **47**, 995-1005 (1998).

104 Hope, K. M. *et al.* Regulation of alpha-cell function by the beta-cell in isolated human and rat islets deprived of glucose: the "switch-off" hypothesis. *Diabetes* **53**, 1488-1495 (2004).

105 Briant, L., Salehi, A., Vergari, E., Zhang, Q. & Rorsman, P. Glucagon secretion from pancreatic alpha-cells. *Ups J Med Sci* **121**, 113-119, doi:10.3109/03009734.2016.1156789 (2016).

106 Jiang, G. & Zhang, B. B. Glucagon and regulation of glucose metabolism. *Am J Physiol Endocrinol Metab* **284**, E671-678, doi:10.1152/ajpendo.00492.2002 (2003).

107 Adeva-Andany, M. M., Perez-Felpete, N., Fernandez-Fernandez, C., Donapetry-Garcia, C. & Pazos-Garcia, C. Liver glucose metabolism in humans. *Biosci Rep* **36**, doi:10.1042/BSR20160385 (2016).

108 Lamberts, S. W., van der Lely, A. J., de Herder, W. W. & Hofland, L. J. Octreotide. *N Engl J Med* **334**, 246-254, doi:10.1056/NEJM199601253340408 (1996).

109 Huang, X. Q. Somatostatin: Likely the most widely effective gastrointestinal hormone in the human body. *World J Gastroenterol* **3**, 201-204, doi:10.3748/wjg.v3.i4.201 (1997).

110 Patel, Y. C. General Aspects of the Biology and Function of Somatostatin. In: Somatostatin. *Basic and Clinical Aspects of Neuroscience* **Vol. 4**, 1-16 (1992).

111 Reichlin, S. Somatostatin. *N Engl J Med* **309**, 1495-1501, doi:10.1056/NEJM198312153092406 (1983).

112 Taborsky, G. J., Jr., Smith, P. H. & Porte, D., Jr. Differential effects of somatostatin analogues on alpha- and beta-cells of the pancreas. *Am J Physiol* **236**, E123-128, doi:10.1152/ajpendo.1979.236.2.E123 (1979).

113 Hauge-Evans, A. C. *et al.* Somatostatin secreted by islet delta-cells fulfills multiple roles as a paracrine regulator of islet function. *Diabetes* **58**, 403-411, doi:10.2337/db08-0792 (2009).

114 van der Meulen, T. *et al.* Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. *Nat Med* **21**, 769-776, doi:10.1038/nm.3872 (2015).

115 Kailey, B. *et al.* SSTR2 is the functionally dominant somatostatin receptor in human pancreatic beta- and alpha-cells. *Am J Physiol Endocrinol Metab* **303**, E1107-1116, doi:10.1152/ajpendo.00207.2012 (2012).

116 DiGruccio, M. R. *et al.* Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. *Mol Metab* **5**, 449-458, doi:10.1016/j.molmet.2016.04.007 (2016).

117 Watts, M., Ha, J., Kimchi, O. & Sherman, A. Paracrine regulation of glucagon secretion: the beta/alpha/delta model. *Am J Physiol Endocrinol Metab* **310**, E597-E611, doi:10.1152/ajpendo.00415.2015 (2016).

118 Kojima, M. *et al.* Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **402**, 656-660, doi:10.1038/45230 (1999).

119 Wierup, N., Svensson, H., Mulder, H. & Sundler, F. The ghrelin cell: a novel developmentally regulated islet cell in the

human pancreas. *Regul Pept* **107**, 63-69 (2002).

120 Tong, J. *et al.* Ghrelin suppresses glucose-stimulated insulin secretion and deteriorates glucose tolerance in healthy humans. *Diabetes* **59**, 2145-2151, doi:10.2337/db10-0504 (2010).

121 Stanley, S., Wynne, K. & Bloom, S. Gastrointestinal satiety signals III. Glucagon-like peptide 1, oxyntomodulin, peptide YY, and pancreatic polypeptide. *Am J Physiol Gastrointest Liver Physiol* **286**, G693-697, doi:10.1152/ajpgi.00536.2003 (2004).

122 Katsuura, G., Asakawa, A. & Inui, A. Roles of pancreatic polypeptide in regulation of food intake. *Peptides* **23**, 323-329 (2002).

123 Sam, A. H. *et al.* Circulating pancreatic polypeptide concentrations predict visceral and liver fat content. *J Clin Endocrinol Metab* **100**, 1048-1052, doi:10.1210/jc.2014-3450 (2015).

124 Williams, G., Harrold, J. A. & Cutler, D. J. The hypothalamus and the regulation of energy homeostasis: lifting the lid on a black box. *Proc Nutr Soc* **59**, 385-396 (2000).

125 Kalra, S. P. *et al.* Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocr Rev* **20**, 68-100, doi:10.1210/edrv.20.1.0357 (1999).

126 Schwartz, T. W. *et al.* Vagal, cholinergic regulation of pancreatic polypeptide secretion. *J Clin Invest* **61**, 781-789, doi:10.1172/JCI108992 (1978).

127 Schwartz, T. W. Pancreatic polypeptide: a hormone under vagal control. *Gastroenterology* **85**, 1411-1425 (1983).

128 Rodriguez-Diaz, R. *et al.* Innervation patterns of autonomic axons in the human endocrine pancreas. *Cell Metab* **14**, 45-54, doi:10.1016/j.cmet.2011.05.008 (2011).

129 Antcliff, J. F., Haider, S., Proks, P., Sansom, M. S. & Ashcroft, F. M. Functional analysis of a structural model of the ATP-binding site of the KATP channel Kir6.2 subunit. *EMBO J* **24**, 229-239, doi:10.1038/sj.emboj.7600487 (2005).

130 Aittoniemi, J. *et al.* Review. SUR1: a unique ATP-binding cassette protein that functions as an ion channel regulator. *Philos Trans R Soc Lond B Biol Sci* **364**, 257-267, doi:10.1098/rstb.2008.0142 (2009).

131 Haider, S., Antcliff, J. F., Proks, P., Sansom, M. S. & Ashcroft, F. M. Focus on Kir6.2: a key component of the ATP-sensitive potassium channel. *J Mol Cell Cardiol* **38**, 927-936, doi:10.1016/j.yjmcc.2005.01.007 (2005).

132 Mikhailov, M. V. *et al.* 3-D structural and functional characterization of the purified KATP channel complex Kir6.2-SUR1. *EMBO J* **24**, 4166-4175, doi:10.1038/sj.emboj.7600877 (2005).

133 Smith, A. J., Taneja, T. K., Mankouri, J. & Sivaprasadarao, A. Molecular cell biology of KATP channels: implications for neonatal diabetes. *Expert Rev Mol Med* **9**, 1-17, doi:10.1017/S1462399407000403 (2007).

134 Fournet, J. C. & Junien, C. Genetics of congenital hyperinsulinism. *Endocr Pathol* **15**, 233-240 (2004).

135 Henquin, J. C. Triggering and amplifying pathways of regulation of insulin secretion by glucose. *Diabetes* **49**, 1751-1760 (2000).

136 Seino, S., Iwanaga, T., Nagashima, K. & Miki, T. Diverse roles of K(ATP) channels learned from Kir6.2 genetically engineered mice. *Diabetes* **49**, 311-318 (2000).

137 Castiello, F. R., Heileman, K. & Tabrizian, M. Microfluidic perfusion systems for secretion fingerprint analysis of pancreatic islets: applications, challenges and opportunities. *Lab Chip* **16**, 409-431, doi:10.1039/c5lc01046b (2016).

138 Henquin, J. C., Dufrane, D., Gmyr, V., Kerr-Conte, J. & Nenquin, M. Pharmacological approach to understanding the control of insulin secretion in human islets. *Diabetes Obes Metab* **19**, 1061-1070, doi:10.1111/dom.12887 (2017).

139 Rorsman, P. & Renstrom, E. Insulin granule dynamics in pancreatic beta cells. *Diabetologia* **46**, 1029-1045, doi:10.1007/s00125-003-1153-1 (2003).

140 Meloni, A. R., DeYoung, M. B., Lowe, C. & Parkes, D. G. GLP-1 receptor activated insulin secretion from pancreatic beta-cells: mechanism and glucose dependence. *Diabetes Obes Metab* **15**, 15-27, doi:10.1111/j.1463-1326.2012.01663.x (2013).

141 Li, J. *et al.* DOC2 isoforms play dual roles in insulin secretion and insulin-stimulated glucose uptake. *Diabetologia* **57**, 2173-2182, doi:10.1007/s00125-014-3312-y (2014).

142 Rorsman, P. *et al.* The Cell Physiology of Biphasic Insulin Secretion. *News Physiol Sci* **15**, 72-77 (2000).

143 Sudhof, T. C. & Rothman, J. E. Membrane fusion: grappling with SNARE and SM proteins. *Science* **323**, 474-477, doi:10.1126/science.1161748 (2009).

144 Hong, W. SNAREs and traffic. *Biochim Biophys Acta* **1744**, 493-517 (2005).

145 Wang, Z. & Thurmond, D. C. Mechanisms of biphasic insulin-granule exocytosis - roles of the cytoskeleton, small GTPases and SNARE proteins. *J Cell Sci* **122**, 893-903, doi:10.1242/jcs.034355 (2009).

146 Heaslip, A. T. *et al.* Cytoskeletal dependence of insulin granule movement dynamics in INS-1 beta-cells in response to glucose. *PLoS One* **9**, e109082, doi:10.1371/journal.pone.0109082 (2014).

147 Creutzfeldt, W. The incretin concept today. *Diabetologia* **16**, 75-85 (1979).

148 Theodorakis, M. J. *et al.* Human duodenal enteroendocrine cells: source of both incretin peptides, GLP-1 and GIP. *Am J Physiol Endocrinol Metab* **290**, E550-559, doi:10.1152/ajpendo.00326.2004 (2006).

149 Wu, T., Rayner, C. K., Jones, K. & Horowitz, M. Dietary effects on incretin hormone secretion. *Vitam Horm* **84**, 81-110, doi:10.1016/B978-0-12-381517-0.00003-5 (2010).

150 Marathe, C. S., Rayner, C. K., Jones, K. L. & Horowitz, M. Glucagon-like peptides 1 and 2 in health and disease: a review. *Peptides* **44**, 75-86, doi:10.1016/j.peptides.2013.01.014 (2013).

151 Orgaard, A. & Holst, J. J. The role of somatostatin in GLP-1-induced inhibition of glucagon secretion in mice. *Diabetologia* **60**, 1731-1739, doi:10.1007/s00125-017-4315-2 (2017).

152 Davidson, M. H. Cardiovascular effects of glucagonlike peptide-1 agonists. *Am J Cardiol* **108**, 33B-41B, doi:10.1016/j.amjcard.2011.03.046 (2011).

153 Perry, T. *et al.* Glucagon-like peptide-1 decreases endogenous amyloid-beta peptide (Abeta) levels and protects hippocampal neurons from death induced by Abeta and iron. *J Neurosci Res* **72**, 603-612, doi:10.1002/jnr.10611 (2003).

154 Shigeto, M., Cha, C. Y., Rorsman, P. & Kaku, K. A role of PLC/PKC-dependent pathway in GLP-1-stimulated insulin secretion. *J Mol Med (Berl)* **95**, 361-368, doi:10.1007/s00109-017-1508-6 (2017).

155 Shigeto, M. *et al.* GLP-1 stimulates insulin secretion by PKC-dependent TRPM4 and TRPM5 activation. *J Clin Invest* **125**, 4714-4728, doi:10.1172/JCI81975 (2015).

156 Shigeto, M., Katsura, M., Matsuda, M., Ohkuma, S. & Kaku, K. Low, but physiological, concentration of GLP-1 stimulates insulin secretion independent of the cAMP-dependent protein kinase pathway. *J Pharmacol Sci* **108**, 274-279 (2008).

157 Sumara, G. *et al.* Regulation of PKD by the MAPK p38delta in insulin secretion and glucose homeostasis. *Cell* **136**, 235-248, doi:10.1016/j.cell.2008.11.018 (2009).

158 Aynsley-Green, A. *et al.* Practical management of hyperinsulinism in infancy. *Arch Dis Child Fetal Neonatal Ed* **82**, F98-F107 (2000).

159 Mehta, A. & Hussain, K. Transient hyperinsulinism associated with macrosomia, hypertrophic obstructive cardiomyopathy, hepatomegaly, and nephromegaly. *Arch Dis Child* **88**, 822-824 (2003).

160 Mohamed, Z., Arya, V. B. & Hussain, K. Hyperinsulinaemic hypoglycaemia:genetic mechanisms, diagnosis and management. *J Clin Res Pediatr Endocrinol* **4**, 169-181, doi:10.4274/jcrpe.821 (2012).

161 Glaser, B., Thornton, P., Otonkoski, T. & Junien, C. Genetics of neonatal hyperinsulinism. *Arch Dis Child Fetal Neonatal Ed* **82**, F79-86 (2000).

162 Flanagan, S. E., Kapoor, R. R. & Hussain, K. Genetics of congenital hyperinsulinemic hypoglycemia. *Semin Pediatr Surg* **20**, 13-17, doi:10.1053/j.sempedsurg.2010.10.004 (2011).

163 Ashcroft, F. M. ATP-sensitive potassium channelopathies: focus on insulin secretion. *J Clin Invest* **115**, 2047-2058, doi:10.1172/JCI25495 (2005).

164 Gribble, F. M., Tucker, S. J., Haug, T. & Ashcroft, F. M. MgATP activates the beta cell KATP channel by interaction with its SUR1 subunit. *Proc Natl Acad Sci U S A* **95**, 7185-7190 (1998).

165 Zerangue, N., Schwappach, B., Jan, Y. N. & Jan, L. Y. A new ER trafficking signal regulates the subunit stoichiometry of plasma membrane K(ATP) channels. *Neuron* **22**, 537-548 (1999).

166 Argmann, C. & Auwerx, J. Insulin secretion: SIRT4 gets in on the act. *Cell* **126**, 837-839, doi:10.1016/j.cell.2006.08.031 (2006).

167 Matschinsky, F. M. Regulation of pancreatic beta-cell glucokinase: from basics to therapeutics. *Diabetes* **51 Suppl 3**, S394-404 (2002).

168 Otonkoski, T. *et al.* Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic beta cells. *Am J Hum Genet* **81**, 467-474, doi:10.1086/520960 (2007).

169 Flanagan, S. E. *et al.* Diazoxide-responsive hyperinsulinemic hypoglycemia caused by HNF4A gene mutations. *Eur J Endocrinol* **162**, 987-992, doi:10.1530/EJE-09-0861 (2010).

170 Clayton, P. T. *et al.* Hyperinsulinism in short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency reveals the importance of beta-oxidation in insulin secretion. *J Clin Invest* **108**, 457-465, doi:10.1172/JCI11294 (2001).

171 Jaeken, J. & Matthijs, G. Congenital disorders of glycosylation: a rapidly expanding disease family. *Annu Rev Genomics Hum Genet* **8**, 261-278, doi:10.1146/annurev.genom.8.080706.092327 (2007).

172 Jaeken, J. & Carchon, H. Congenital disorders of glycosylation: the rapidly growing tip of the iceberg. *Curr Opin Neurol* **14**, 811-815 (2001).

173 Jaeken, J. Congenital disorders of glycosylation. *Ann N Y Acad Sci* **1214**, 190-198, doi:10.1111/j.1749-6632.2010.05840.x (2010).

174 Jaeken, J. Familial psychomotor retardation with markedly fluctuating serum prolactin, FSH and GH levels, partial TBG-deficiency, increased serum arylsulphatase A and increased CSF protein: A new syndrome? *Pediatric Research* **14:179** (1980).

175 Grunewald, S., Matthijs, G. & Jaeken, J. Congenital disorders of glycosylation: a review. *Pediatr Res* **52**, 618-624, doi:10.1203/00006450-200211000-00003 (2002).

176 Sparks, S. E. & Krasnewich, D. M. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (1993).

177 Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57-63, doi:10.1038/nature10423 (2011).

178 Neumann, L. M., von Moers, A., Kunze, J., Blankenstein, O. & Marquardt, T. Congenital disorder of glycosylation type 1a in a macrosomic 16-month-old boy with an atypical phenotype and homozygosity of the N216I mutation. *Eur J Pediatr* **162**, 710-713, doi:10.1007/s00431-003-1278-8 (2003).

179 Yuste-Checa, P. *et al.* The Effects of PMM2-CDG-Causing Mutations on the Folding, Activity, and Stability of the PMM2 Protein. *Hum Mutat* **36**, 851-860, doi:10.1002/humu.22817 (2015).

180 Jaeken, J. *et al.* Sialic acid-deficient serum and cerebrospinal fluid transferrin in a newly recognized genetic syndrome. *Clin Chim Acta* **144**, 245-247 (1984).

181 Grunewald, S. The clinical spectrum of phosphomannomutase 2 deficiency (CDG-Ia). *Biochim Biophys Acta* **1792**, 827-834, doi:10.1016/j.bbadis.2009.01.003 (2009).

182 Hagberg, B. A., Blennow, G., Kristiansson, B. & Stibler, H. Carbohydrate-deficient glycoprotein syndromes: peculiar group of new disorders. *Pediatr Neurol* **9**, 255-262 (1993).

183 de Lonlay, P. *et al.* A broad spectrum of clinical presentations in congenital disorders of glycosylation I: a series of 26 cases. *J Med Genet* **38**, 14-19 (2001).

184 Schachter, H. Congenital disorders involving defective N-glycosylation of proteins. *Cell Mol Life Sci* **58**, 1085-1104, doi:10.1007/PL00000923 (2001).

185 Matthijs, G. *et al.* Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). *Nat Genet* **16**, 88-92, doi:10.1038/ng0597-88 (1997).

186 Schollen, E. *et al.* Comparative analysis of the phosphomannomutase genes PMM1, PMM2 and PMM2psi: the sequence variation in the processed pseudogene is a reflection of the mutations found in the functional gene. *Hum Mol Genet* **7**, 157-164 (1998).

187 Van Schaftingen, E. & Jaeken, J. Phosphomannomutase deficiency is a cause of carbohydrate-deficient glycoprotein syndrome type I. *FEBS Lett* **377**, 318-320, doi:10.1016/0014-5793(95)01357-1 (1995).

188 Carchon, H., Van Schaftingen, E., Matthijs, G. & Jaeken, J. Carbohydrate-deficient glycoprotein syndrome type IA (phosphomannomutase-deficiency). *Biochim Biophys Acta* **1455**, 155-165 (1999).

189 Matthijs, G., Schollen, E., Heykants, L. & Grunewald, S. Phosphomannomutase deficiency: the molecular basis of the

classical Jaeken syndrome (CDGS type Ia). *Mol Genet Metab* **68**, 220-226, doi:10.1006/mgme.1999.2914 (1999).

190 Schollen, E. *et al.* Characterization of two unusual truncating PMM2 mutations in two CDG-Ia patients. *Mol Genet Metab* **90**, 408-413, doi:10.1016/j.ymgme.2007.01.003 (2007).

191 Haeuptle, M. A. & Hennet, T. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum Mutat* **30**, 1628-1641, doi:10.1002/humu.21126 (2009).

192 Perez, B. *et al.* The molecular landscape of phosphomannose mutase deficiency in iberian peninsula: identification of 15 population-specific mutations. *JIMD Rep* **1**, 117-123, doi:10.1007/8904_2011_26 (2011).

193 Jensen, O. N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* **8**, 33-41, doi:10.1016/j.cbpa.2003.12.009 (2004).

194 Fletcher, C. M., Coyne, M. J. & Comstock, L. E. Theoretical and experimental characterization of the scope of protein O-glycosylation in Bacteroides fragilis. *J Biol Chem* **286**, 3219-3226, doi:10.1074/jbc.M110.194506 (2011).

195 Freeze, H. H. Understanding human glycosylation disorders: biochemistry leads the charge. *J Biol Chem* **288**, 6936-6945, doi:10.1074/jbc.R112.429274 (2013).

196 Marquardt, T. & Denecke, J. Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies. *Eur J Pediatr* **162**, 359-379, doi:10.1007/s00431-002-1136-0 (2003).

197 Taylor, M. Introduction to glycobiology. *Oxford University Press* (2011).

198 Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172, doi:10.1186/1471-2105-8-172 (2007).

199 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

200 Marraffini, L. A. in *Streptococcus pyogenes : Basic Biology to Clinical Manifestations* (eds J. J. Ferretti, D. L. Stevens, & V. A. Fischetti) (2016).

201 van Erp, P. B., Bloomer, G., Wilkinson, R. & Wiedenheft, B. The history and market impact of CRISPR RNA-guided nucleases. *Curr Opin Virol* **12**, 85-90, doi:10.1016/j.coviro.2015.03.011 (2015).

202 Bassing, C. H. & Alt, F. W. The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst)* **3**, 781-796, doi:10.1016/j.dnarep.2004.06.001 (2004).

203 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

204 Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* **8**, 753-755, doi:10.1038/nmeth.1653 (2011).

205 Song, F. & Stieger, K. Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Mol Ther Nucleic Acids* **7**, 53-60, doi:10.1016/j.omtn.2017.02.006 (2017).

206 Saleh-Gohari, N. & Helleday, T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res* **32**, 3683-3688, doi:10.1093/nar/gkh703 (2004).

207 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355, doi:10.1038/nbt.2842 (2014).

208 Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344, doi:10.1038/nbt.3481 (2016).

209 Cabezas, O. R. *et al.* Polycystic Kidney Disease with Hyperinsulinemic Hypoglycemia Caused by a Promoter Mutation in Phosphomannomutase 2. *J Am Soc Nephrol* **28**, 2529-2539, doi:10.1681/ASN.2016121312 (2017).

210 Krebs, H. A. H., K. Untersuchungen uber die Harnstoffbildung im Tierkörper. *Hoppe-Seylers Z. Physiol. Chem.* , 33-66 (1932).

211 Gey, G. O. Maintenance of human normal cells in continuous culture: preliminary report. In Cultivation of mesoblastic tumors and normal cells and notes on methods of cultivation. *American Journal of Cancer* **27**, 45-76 (1936).

212 Nicholls, S. B., Chu, J., Abbruzzese, G., Tremblay, K. D. & Hardy, J. A. Mechanism of a genetically encoded dark-to-bright reporter for caspase activity. *J Biol Chem* **286**, 24977-24986, doi:10.1074/jbc.M111.221648 (2011).

213 Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotechnol* **35**, 453-462, doi:10.1038/nbt.3805 (2017).

214 Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168, doi:10.1093/nar/gku936 (2014).

215 Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148, doi:10.1186/s13059-016-1012-2 (2016).

216 Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44**, W272-276, doi:10.1093/nar/gkw398 (2016).

217 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).

218 Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553, doi:10.1371/journal.pone.0005553 (2009).

219 Anno, Y. N. *et al.* Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription. *Nucleic Acids Res* **39**, 3116-3127, doi:10.1093/nar/gkq1301 (2011).

220 Crichton, R. R. & Charloteaux-Wauters, M. Iron transport and storage. *Eur J Biochem* **164**, 485-506 (1987).

221 Guillard, M. *et al.* Transferrin mutations at the glycosylation site complicate diagnosis of congenital disorders of glycosylation type I. *J Inherit Metab Dis* **34**, 901-906, doi:10.1007/s10545-011-9311-y (2011).

222 Hoopes, R. R., Jr. *et al.* Dent Disease with mutations in OCRL1. *Am J Hum Genet* **76**, 260-267, doi:10.1086/427887 (2005).

223 Hofherr, A., Wagner, C., Fedeles, S., Somlo, S. & Kottgen, M. N-glycosylation determines the abundance of the transient receptor potential channel TRPP2. *J Biol Chem* **289**, 14854-14867, doi:10.1074/jbc.M114.562264 (2014).

224 Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**, 234-246, doi:10.1038/nrg3663 (2014).

225 Hanssen, L. L. P. *et al.* Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol* **19**, 952-961, doi:10.1038/ncb3573 (2017).

226 Bailey, S. D. *et al.* ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat Commun* **2**, 6186, doi:10.1038/ncomms7186 (2015).

227 Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121, doi:10.1016/j.cell.2016.02.007 (2016).

228 Ye, B. Y. *et al.* [ZNF143 is involved in CTCF-mediated chromatin interactions by cooperation with cohesin and other partners]. *Mol Biol (Mosk)* **50**, 496-503, doi:10.7868/S0026898416030034 (2016).

229 Osei, Y. D. & Churchich, J. E. Screening and sequence determination of a cDNA encoding the human brain 4-aminobutyrate aminotransferase. *Gene* **155**, 185-187 (1995).

230 Medina-Kauwe, L. K. *et al.* 4-Aminobutyrate aminotransferase (GABA-transaminase) deficiency. *J Inherit Metab Dis* **22**, 414-427 (1999).

231 Pfeiffer, J. R., McAvoy, B. L., Fecteau, R. E., Deleault, K. M. & Brooks, S. A. CARHSP1 is required for effective tumor necrosis factor alpha mRNA stabilization and localizes to processing bodies and exosomes. *Mol Cell Biol* **31**, 277-286, doi:10.1128/MCB.00775-10 (2011).

232 Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res* **6**, 986-994 (1996).

233 Freeman, W. M., Walker, S. J. & Vrana, K. E. Quantitative RT-PCR: pitfalls and potential. *BioTechniques* **26**, 112-125 (1999).

234 Mikkilineni, V. *et al.* Digital quantitative measurements of gene expression. *Biotechnology and bioengineering* **86**, 117-124 (2004).

235 Baker, M. Digital PCR hits its stride. *Nat Meth* **9**, 541-544 (2012).

236 Vogelstein, B. & Kinzler, K. W. Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-9241 (1999).

237 Pohl, G. & Shih, I.-M. Principle and applications of digital PCR. *Expert review of molecular diagnostics* **4**, 41-47 (2004).

238 Yan, H. *et al.* Small changes in expression affect predisposition to tumorigenesis. *Nature genetics* **30**, 25-26 (2002).

239 Mansier, O. *et al.* Quantification of the Mutant CALR Allelic Burden by Digital PCR: Application to Minimal Residual Disease Evaluation after Bone Marrow Transplantation. *The Journal of molecular diagnostics : JMD*, doi:10.1016/j.jmoldx.2015.07.007 (2015).

240 Bjerregaard, H., Pedersen, S., Kristensen, S. R. & Marcussen, N. Reference genes for gene expression analysis by real-time reverse transcription polymerase chain reaction of renal cell carcinoma. *Diagnostic Molecular Pathology* **20**, 212-217 (2011).

241 Beavon, I. The E-cadherin–catenin complex in tumour metastasis: structure, function and regulation. *European journal of cancer* **36**, 1607-1620 (2000).

242 Roitbak, T. *et al.* A polycystin-1 multiprotein complex is disrupted in polycystic kidney disease cells. *Mol Biol Cell* **15**, 1334-1346, doi:10.1091/mbc.E03-05-0296 (2004).

243 Gottardi, C. J., Wong, E. & Gumbiner, B. M. E-cadherin suppresses cellular transformation by inhibiting beta-catenin signaling in an adhesion-independent manner. *J Cell Biol* **153**, 1049-1060 (2001).

244 Tian, X. *et al.* E-cadherin/beta-catenin complex and the epithelial barrier. *J Biomed Biotechnol* **2011**, 567305, doi:10.1155/2011/567305 (2011).

245 Presnell, S. C. *et al.* Isolation, characterization, and expansion methods for defined primary renal cell populations from rodent, canine, and human normal and diseased kidneys. *Tissue Engineering Part C: Methods* **17**, 261-273 (2010).

246 Ståhlberg, A., Håkansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the reverse transcription reaction in mRNA quantification. *Clinical chemistry* **50**, 509-515 (2004).

247 Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11-23 (2001).

248 Russell, J. E., Morales, J. & Liebhaber, S. A. The role of mRNA stability in the control of globin gene expression. *Progress in nucleic acid research and molecular biology* **57**, 249 (1997).

249 Tarentino, A. L., Trimble, R. B. & Plummer, T. H. Enzymatic approaches for studying the structure, synthesis, and processing of glycoproteins. *Methods in cell biology* **32**, 111-139 (1989).

250 Tarentino, A. L. & Plummer, T. H. [4] Enzymatic deglycosylation of asparagine-linked glycans: Purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. *Methods in enzymology* **230**, 44-57 (1994).

251 Trimble, R. B., Tarentino, A. L., Plummer, T. H., Jr. & Maley, F. Asparaginyl glycopeptides with a low mannose content are hydrolyzed by endo-beta-N-acetylglucosaminidase H. *J Biol Chem* **253**, 4508-4511 (1978).

252 Gromada, J. *et al.* Glucagon-like peptide I increases cytoplasmic calcium in insulin-secreting beta TC3-cells by enhancement of intracellular calcium mobilization. *Diabetes* **44**, 767-774 (1995).

253 Hanson, M. S. *et al.* A simplified approach to human islet quality assessment. *Transplantation* **89**, 1178-1188, doi:10.1097/TP.0b013e3181d54bce (2010).

254 Ghorbani, R., Jalili, C., Salahshoor, M. R. & Shiasi, M. The effect of time and temperature on viability and performance of Langerhans islets separated from Balb/c mouse after death. *Adv Biomed Res* **4**, 93, doi:10.4103/2277-9175.156657 (2015).

255 Papas, K. K., Suszynski, T. M. & Colton, C. K. Islet assessment for transplantation. *Curr Opin Organ Transplant* **14**, 674-682, doi:10.1097/MOT.0b013e328332a489 (2009).

256 Abdelli, S. *et al.* Intracellular stress signaling pathways activated during human islet preparation and following acute cytokine exposure. *Diabetes* **53**, 2815-2823 (2004).

257 Conti, L. R., Radeke, C. M. & Vandenberg, C. A. Membrane targeting of ATP-sensitive potassium channel effects of glycosylation on surface expression. *Journal of Biological Chemistry* **277**, 25416-25422 (2002).

258 Flanagan, S. E. *et al.* Update of mutations in the genes encoding the pancreatic beta-cell K(ATP) channel subunits Kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Hum Mutat* **30**, 170-180, doi:10.1002/humu.20838 (2009).

259 Jenkins, H. Le Chatelier's Principle. *Chemical Thermodynamics at a Glance*, 160-163 (2008).

260 Meda, P. *et al.* Rapid and reversible secretion changes during uncoupling of rat insulin-producing cells. *Journal of Clinical Investigation* **86**, 759 (1990).

261 Meda, P. Protein-mediated interactions of pancreatic islet cells. *Scientifica (Cairo)* **2013**, 621249, doi:10.1155/2013/621249 (2013).

262 Green, A. D., Vasu, S., McClenaghan, N. H. & Flatt, P. R. Pseudoislet formation enhances gene expression, insulin secretion and cytoprotective mechanisms of clonal human insulin-secreting 1.1B4 cells. *Pflugers Arch* **467**, 2219-2228, doi:10.1007/s00424-014-1681-1 (2015).

263 Qureshi, F. M., Dejene, E. A., Corbin, K. L. & Nunemaker, C. S. Stress-induced dissociations between intracellular calcium signaling and insulin secretion in pancreatic islets. *Cell Calcium* **57**, 366-375, doi:10.1016/j.ceca.2015.03.002 (2015).

264 Bottino, R. *et al.* Response of human islets to isolation stress and the effect of antioxidant treatment. *Diabetes* **53**, 2559-2568 (2004).

265 Knowles, B. B., Howe, C. C. & Aden, D. P. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science* **209**, 497-499 (1980).

266 Costantini, S., Di Bernardo, G., Cammarota, M., Castello, G. & Colonna, G. Gene expression signature of human HepG2 cell line. *Gene* **518**, 335-345, doi:10.1016/j.gene.2012.12.106 (2013).

267 Kouranova, E. *et al.* CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. *Hum Gene Ther* **27**, 464-475, doi:10.1089/hum.2016.009 (2016).

268 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

269 Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods* **12**, 982-988, doi:10.1038/nmeth.3543 (2015).

270 Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* **34**, 933-941, doi:10.1038/nbt.3659 (2016).

271 Wood, A. J. *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307, doi:10.1126/science.1207773 (2011).

272 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

273 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

274 Wang, H. X. *et al.* CRISPR/Cas9-Based Genome Editing for Disease Modeling and Therapy: Challenges and Opportunities for Nonviral Delivery. *Chem Rev* **117**, 9874-9906, doi:10.1021/acs.chemrev.6b00799 (2017).

275 Ma, H. *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature* **548**, 413-419, doi:10.1038/nature23305 (2017).

276 Liang, P. *et al.* Correction of beta-thalassemia mutant by base editor in human embryos. *Protein Cell* **8**, 811-822, doi:10.1007/s13238-017-0475-6 (2017).

277 Zheng, T. *et al.* Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. *Sci Rep* **7**, 40638, doi:10.1038/srep40638 (2017).

278 Inui, M. *et al.* Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci Rep* **4**, 5396, doi:10.1038/srep05396 (2014).

279 Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res* **42**, 7473-7485, doi:10.1093/nar/gku402 (2014).

280 Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407, doi:10.1093/nar/gku410 (2014).

281 Xu, X., Duan, D. & Chen, S. J. CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Sci Rep* **7**, 143, doi:10.1038/s41598-017-00180-1 (2017).

282 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

283 Jones, C. H., Chen, C. K., Ravikrishnan, A., Rane, S. & Pfeifer, B. A. Overcoming

nonviral gene delivery barriers: perspective and future. *Mol Pharm* **10**, 4082-4098, doi:10.1021/mp400467x (2013).

284 Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J Biotechnol* **241**, 136-146, doi:10.1016/j.jbiotec.2016.11.011 (2017).

285 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

286 Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**, 113-139, doi:10.1146/annurev-genet-051710-150955 (2010).

287 Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res* **24**, 1020-1027, doi:10.1101/gr.171264.113 (2014).

288 Lukacs, G. L. *et al.* Size-dependent DNA mobility in cytoplasm and nucleus. *J Biol Chem* **275**, 1625-1629 (2000).

289 Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44-53, doi:10.1016/j.jbiotec.2015.04.024 (2015).

290 Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012-1019, doi:10.1101/gr.171322.113 (2014).

291 Casado, M. *et al.* Mild clinical and biochemical phenotype in two patients with PMM2-CDG (congenital disorder of glycosylation Ia). *Cerebellum* **11**, 557-563, doi:10.1007/s12311-011-0313-y (2012).

292 Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* **12**, 204-213, doi:10.1038/nrg2949 (2011).

293 Kang, S. S., Shin, S. H., Auh, C. K. & Chun, J. Human skeletal dysplasia caused by a constitutive activated transient receptor potential vanilloid 4 (TRPV4) cation channel mutation. *Exp Mol Med* **44**, 707-722, doi:10.3858/emm.2012.44.12.080 (2012).

294 Trifunovic, J., Miller, L., Debeljak, Z. & Horvat, V. Pathologic patterns of interleukin 10 expression--a review. *Biochem Med (Zagreb)* **25**, 36-48, doi:10.11613/BM.2015.004 (2015).

295 Lloyd, S. E. *et al.* A common molecular basis for three inherited kidney stone diseases. *Nature* **379**, 445-449, doi:10.1038/379445a0 (1996).

296 Janne, P. A. *et al.* Functional overlap between murine Inpp5b and Ocrl1 may explain why deficiency of the murine ortholog for OCRL1 does not cause Lowe syndrome in mice. *J Clin Invest* **101**, 2042-2053, doi:10.1172/JCI2414 (1998).

297 Ittisoponpisan, S., Alhuzimi, E., Sternberg, M. J. & David, A. Landscape of Pleiotropic Proteins Causing Human Disease: Structural and System Biology Insights. *Hum Mutat* **38**, 289-296, doi:10.1002/humu.23155 (2017).

298 Lonfat, N., Montavon, T., Darbellay, F., Gitto, S. & Duboule, D. Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**, 1004-1006, doi:10.1126/science.1257493 (2014).

299 Gabbay, M., Ellard, S., De Franco, E. & Moises, R. S. Pancreatic Agenesis due to Compound Heterozygosity for a Novel Enhancer and Truncating Mutation in the PTF1A Gene. *J Clin Res Pediatr Endocrinol* **9**, 274-277, doi:10.4274/jcrpe.4494 (2017).

300 Patel, N. *et al.* A novel mechanism for variable phenotypic expressivity in Mendelian diseases uncovered by an AU-rich element (ARE)-creating mutation. *Genome Biol* **18**, 144, doi:10.1186/s13059-017-1274-3 (2017).

301 de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499-506, doi:10.1038/nature12753 (2013).

302 Ko, J. Y., Oh, S. & Yoo, K. H. Functional Enhancers As Master Regulators of Tissue-Specific Gene Regulation and Cancer Development. *Mol Cells* **40**, 169-177, doi:10.14348/molcells.2017.0033 (2017).

303 Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb Perspect Biol* **2**, a000729, doi:10.1101/cshperspect.a000729 (2010).

304 Liang, J. *et al.* Epstein-Barr virus super-enhancer eRNAs are essential for MYC oncogene expression and lymphoblast proliferation. *Proc Natl Acad Sci U S A* **113**, 14121-14126, doi:10.1073/pnas.1616697113 (2016).

305 Rockman, M. V. & Wray, G. A. Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol* **19**, 1991-2004, doi:10.1093/oxfordjournals.molbev.a004023 (2002).

306 Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195, doi:10.1126/science.1222794 (2012).

307 Dunham, I. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

308 Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113, doi:10.1038/nature11279 (2012).

309 Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308 (1981).

310 Stamatoyannopoulos, J. A. What does our genome encode? *Genome Res* **22**, 1602-1611, doi:10.1101/gr.146506.112 (2012).

311 Knight, J. C., Keating, B. J., Rockett, K. A. & Kwiatkowski, D. P. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet* **33**, 469-475, doi:10.1038/ng1124 (2003).

312 Elgar, G. & Vavouri, T. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet* **24**, 344-352, doi:10.1016/j.tig.2008.04.005 (2008).

313 Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286-290, doi:10.1126/science.1084564 (2003).

314 Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311 (2001).

315 Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57**, 159-197, doi:10.1146/annurev.bi.57.070188.001111 (1988).

316 Sonawane, A. R. *et al.* Understanding Tissue-Specific Gene Regulation. *Cell Rep* **21**, 1077-1088, doi:10.1016/j.celrep.2017.10.001 (2017).

317 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

318 Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237, doi:10.1016/j.tig.2016.01.003 (2016).

319 Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* **111**, 996-1001, doi:10.1073/pnas.1317788111 (2014).

320 Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458, doi:10.1126/science.aad9024 (2016).

321 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).

322 Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 772, doi:10.1038/nrg.2016.147 (2016).

323 Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281-1295, doi:10.1016/j.cell.2013.04.053 (2013).

324 Myslinski, E., Krol, A. & Carbon, P. ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf. *J Biol Chem* **273**, 21998-22006 (1998).

325 Drewes, T., Senkel, S., Holewa, B. & Ryffel, G. U. Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Mol Cell Biol* **16**, 925-931 (1996).

326 Dillon, N. & Sabbattini, P. Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. *Bioessays* **22**, 657-665, doi:10.1002/1521-1878(200007)22:7<657::AID-BIES8>3.0.CO;2-2 (2000).

327 Kleinjan, D. A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* **76**, 8-32, doi:10.1086/426833 (2005).

328 Dong, H. *et al.* Gamma-aminobutyric acid up- and downregulates insulin secretion from beta cells in concert with changes in glucose concentration. *Diabetologia* **49**, 697-705, doi:10.1007/s00125-005-0123-1 (2006).

329 Koenig, M. K. *et al.* Phenotype of GABA-transaminase deficiency. *Neurology* **88**,

1919-1924, doi:10.1212/WNL.0000000000003936 (2017).

330 Lee, S., Wishart, M. J. & Williams, J. A. Identification of calcineurin regulated phosphorylation sites on CRHSP-24. *Biochem Biophys Res Commun* **385**, 413-417, doi:10.1016/j.bbrc.2009.05.096 (2009).

331 Cortini, R. & Filion, G. J. Theoretical principles of transcription factor traffic on folded chromatin. *Nat Commun* **9**, 1740, doi:10.1038/s41467-018-04130-x (2018).

332 Strom, E. H., Stromme, P., Westvik, J. & Pedersen, S. J. Renal cysts in the carbohydrate-deficient glycoprotein syndrome. *Pediatr Nephrol* **7**, 253-255 (1993).

333 Kapoor, R. R., James, C. & Hussain, K. Hyperinsulinism in developmental syndromes. *Endocr Dev* **14**, 95-113, doi:10.1159/000207480 (2009).

334 Shanti, B. *et al.* Congenital disorder of glycosylation type Ia: heterogeneity in the clinical presentation from multivisceral failure to hyperinsulinaemic hypoglycaemia as leading symptoms in three infants with phosphomannomutase deficiency. *J Inherit Metab Dis* **32 Suppl 1**, S241-251, doi:10.1007/s10545-009-1180-2 (2009).

335 Babovic-Vuksanovic, D. *et al.* Severe hypoglycemia as a presenting symptom of carbohydrate-deficient glycoprotein syndrome. *J Pediatr* **135**, 775-781 (1999).

336 Sun, L. *et al.* Congenital disorder of glycosylation id presenting with hyperinsulinemic hypoglycemia and islet cell hyperplasia. *J Clin Endocrinol Metab* **90**, 4371-4375, doi:10.1210/jc.2005-0250 (2005).

337 Bohles, H. *et al.* Hyperinsulinaemic hypoglycaemia--leading symptom in a patient with congenital disorder of glycosylation Ia (phosphomannomutase deficiency). *J Inherit Metab Dis* **24**, 858-862 (2001).

338 Vajravelu, M. E. & De Leon, D. D. Genetic characteristics of patients with congenital hyperinsulinism. *Curr Opin Pediatr*, doi:10.1097/MOP.0000000000000645 (2018).

339 Conti, L. R., Radeke, C. M. & Vandenberg, C. A. Membrane targeting of ATP-sensitive potassium channel. Effects of glycosylation on surface expression. *J Biol Chem* **277**, 25416-25422, doi:10.1074/jbc.M203109200 (2002).

340 Asano, T. *et al.* The role of N-glycosylation of GLUT1 for glucose transport activity. *J Biol Chem* **266**, 24632-24636 (1991).

341 Porath, B. *et al.* Mutations in GANAB, Encoding the Glucosidase IIalpha Subunit, Cause Autosomal-Dominant Polycystic Kidney and Liver Disease. *Am J Hum Genet* **98**, 1193-1207, doi:10.1016/j.ajhg.2016.05.004 (2016).

342 Tham, E. *et al.* A novel phenotype in N-glycosylation disorders: Gillessen-Kaesbach-Nishimura skeletal dysplasia due to pathogenic variants in ALG9. *Eur J Hum Genet* **24**, 198-207, doi:10.1038/ejhg.2015.91 (2016).

343 Janssen, M. J., Waanders, E., Woudenberg, J., Lefeber, D. J. & Drenth, J. P. Congenital disorders of glycosylation in hepatology: the example of polycystic liver disease. *J Hepatol* **52**, 432-440, doi:10.1016/j.jhep.2009.12.011 (2010).

344 Hep G2 is a hepatoblastoma-derived cell line.

345 Sacco, A. e. a. Cycle Reactivation in Skeletal Muscle and Other Terminally Differentiated Cells. *Madame Curie Bioscience Database. Austin (TX): Landes Bioscience. Available from:* https://www.ncbi.nlm.nih.gov/books/NBK6180/ (2000-2013).

346 Kaur, G. & Dufour, J. M. Cell lines: Valuable tools or useless artifacts. *Spermatogenesis* **2**, 1-5, doi:10.4161/spmg.19885 (2012).

347 Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-212, doi:10.1038/ng.2871 (2014).

348 Oudelaar, A. M., Downes, D. J., Davies, J. O. J. & Hughes, J. R. Low-input Capture-C: A Chromosome Conformation Capture Assay to Analyze Chromatin Architecture in Small Numbers of Cells. *Bio Protoc* **7**, doi:10.21769/BioProtoc.2645 (2017).

349 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

350 Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol*, doi:10.1038/nbt.4172 (2018).

351 Gehrke, J. M. *et al.* An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol*, doi:10.1038/nbt.4199 (2018).

352 Giacalone, J. C. *et al.* CRISPR-Cas9-Based Genome Editing of Human Induced Pluripotent Stem Cells. *Curr Protoc Stem Cell Biol* **44**, 5B 7 1-5B 7 22, doi:10.1002/cpsc.46 (2018).

353 Beer, N. L. & Gloyn, A. L. Genome-edited human stem cell-derived beta cells: a powerful tool for drilling down on type 2 diabetes GWAS biology. *F1000Res* **5**, doi:10.12688/f1000research.8682.1 (2016).

354 Lecomte, M. J. *et al.* Aggregation of Engineered Human beta-Cells Into Pseudoislets: Insulin Secretion and Gene Expression Profile in Normoxic and Hypoxic Milieu. *Cell Med* **8**, 99-112, doi:10.3727/215517916X692843 (2016).

355 Green, A. D., Vasu, S. & Flatt, P. R. Cellular models for beta-cell function and diabetes gene therapy. *Acta Physiol (Oxf)* **222**, doi:10.1111/apha.13012 (2018).

356 Mukhopadhyay, A., Deplancke, B., Walhout, A. J. & Tissenbaum, H. A. Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in Caenorhabditis elegans. *Nat Protoc* **3**, 698-709, doi:10.1038/nprot.2008.38 (2008).

Appendix

1. Insulin sceretion assay (Assay 1) human pancreatic islets Batch 1 (absolute values

Table 16 APPENDIX: Single values, means and SEM (abs. values; µU/mL) measured and calculated for experiment 1, Assay 1 (islets Batch 1).
Secreted insulin and intracellular insulin (cell lysates).

Experiment 1 (µU/mL)						
Condition	**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4/Control**	**Mean**	**SEM**
2.8 mM 0 min	173.80	343.56	330.44	378.45	350.82	45.40
2.8 mM 60 min	366.07	585.82	424.86	351.26	432.0	53.68
28 mM 0 min	181.50	32.62	98.16	14.93	81.80	37.75
28 mM 60 min	444.93	287.17	209.95	272.32	303.59	50.00
2.8 mM 0 min	385.95	421.3	472.09	280.06	389.85	40.64
2.8 mM 60 min	366.0	637.46	356.48	382.15	435.52	67.52
0 mM + PNGase F 0 min	199.69	173.51	7.84	/	127.01	60.06
0 mM + PNGase F 60 min	587.03	463.87	331.42	/	460.77	73.80
0 mM 0 min (Control)	/	/	/	318.08	/	/
0 mM 60 min (Control)	/	/	/	532.57	/	/
Cell lysates	421.15	2640.97	455.71	480.37 (Control)	1172.61*	734.25*

*excluding sample 4/control

2.8 mM = 2.8 mM glucose

28 mM = 28 mM glucose

0 mM = 0 mM glucose

PNGase F = 500 nM PNGase

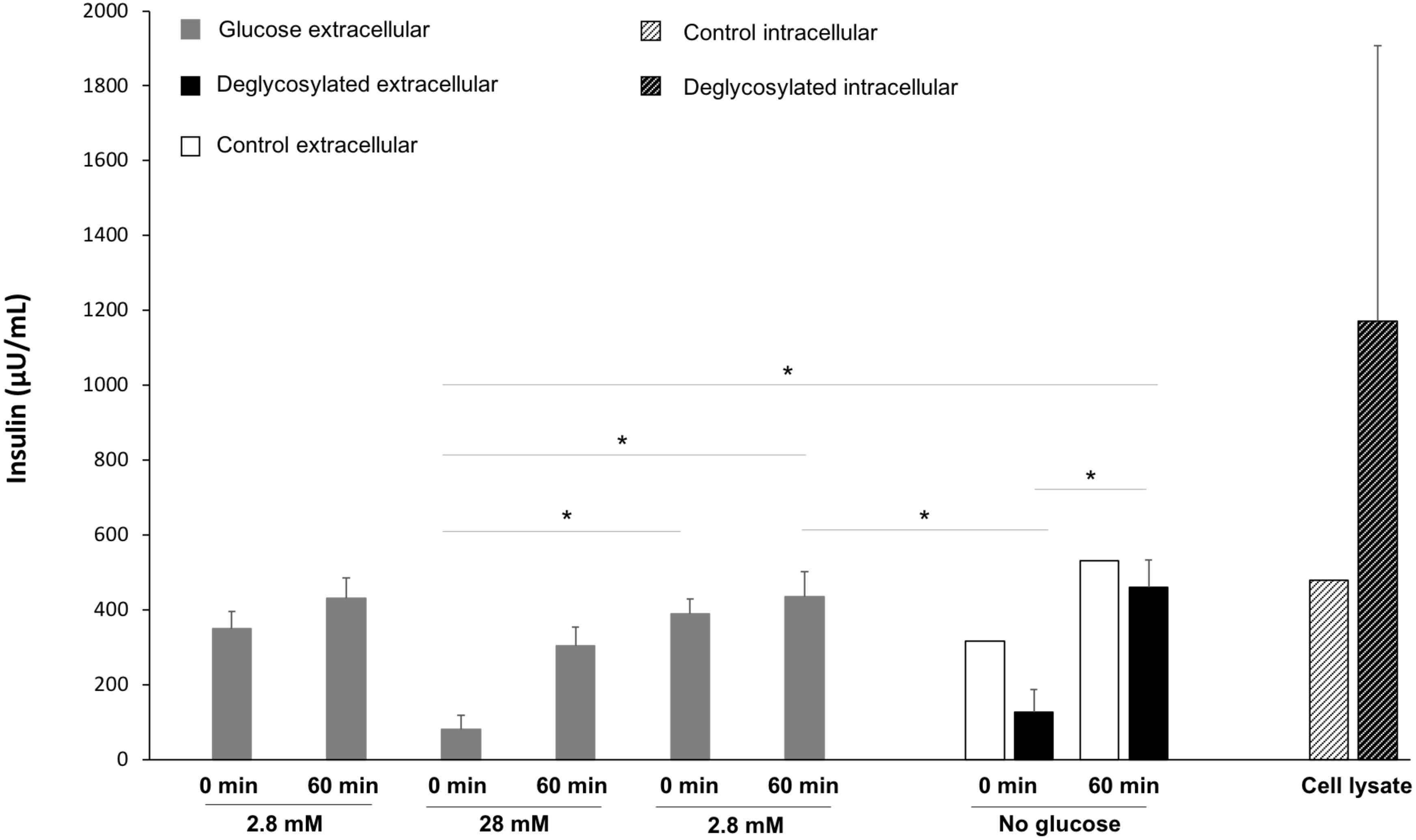


Figure 72: APPENDIX. Assay 1, experiment 1. Extra- and intracellular insulin measurement.
Absolute insulin concentrations (µU/mL) for time points T_0 and T_{60} are shown. All four islets samples were included in the GSIS series (n=4) and stimulated with 2.8 mM, 28 mM and 2.8 mM glucose (Grey bars). After GSIS three of the samples were incubated with 2,000 U PNGase F pre-prepared in its enzyme buffer (Na_3PO_4, 50 mM final concentration) in 0 mM glucose (n=3, black bar). One sample was used as a control only incubated with enzyme buffer also in 0 mM glucose (n=1, white bar). After passing all 4 conditions islets were lysed. Values are means±SEM when applicable. Statistical significance (P<0.05) was tested using a one-way ANOVA followed by a Bonferroni multiple comparisons test to compare between the different conditions.

Table 17: APPENDIX. Single values, means and SEM (abs. values; µU/mL) measured and calculated for experiment 2, Assay 1 (islets Batch 1). Secreted insulin and intracellular insulin (cell lysates).

Experiment 2 (µU/mL)						
Condition	**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4/Control**	**Mean**	**SEM**
2.8 mM 0 min	25.14	37.35	35.24	80.52	44.56	12.28
2.8 mM 60 min	74.16	100.98	97.74	209.99	120.72	30.35
28 mM 0 min	37.74	15.26	113.66	93.64	65.07	23.10
28 mM 60 min	40.98	46.92	68.18	285.91	110.5	58.76
2.8 mM 0 min	96.08	105.58	216.20	102.55	130.10	28.77
2.8 mM 60 min	188.16	152.38	169.19	237.17	186.73	18.33
0 mM + PNGase F 0 min	4.99	4.43	10.54	/	6.65	1.95
0 mM + PNGase F 60 min	157.97	168.61	164.61	/	163.54	2.95
0 mM 0 min (Control)	/	/	/	5.55	/	/
0 mM 60 min (Control)	/	/	/	178.12	/	/
Cell lysates	283.72	257.3	299.72	290.61*	280.25	12.37

*excluding sample 4/control

2.8 mM = 2.8 mM glucose

28 mM = 28 mM glucose

0 mM = 0 mM glucose

PNGase F = 500 nM PNGase F

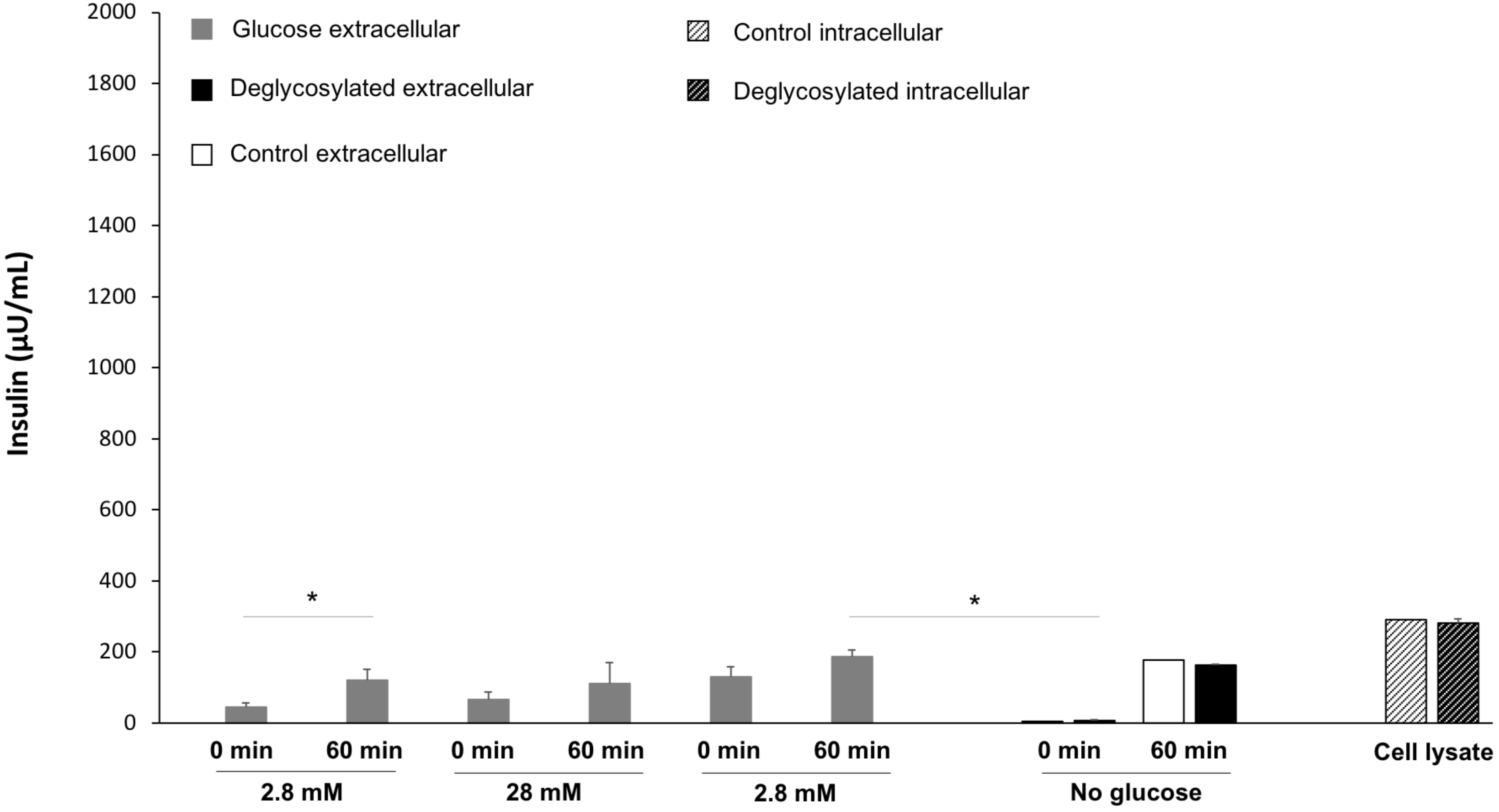


Figure 73: APPENDIX. Assay 1, experiment 2. Extra- and intracellular insulin measurement.
Absolute insulin concentrations (µU/mL) for time points T_0 and T_{60} are shown. All four islets samples were included in the GSIS series (n=4) and stimulated with 2.8 mM, 28 mM and 2.8 mM glucose (Grey bars). After GSIS three of the samples were incubated with 2,000 U PNGase F pre-prepared in its enzyme buffer (Na_3PO_4, 50 mM final concentration) in 0 mM glucose (n=3, black bar). One sample was used as a control only incubated with enzyme buffer also in 0 mM glucose (n=1, white bar). After passing all 4 conditions islets were lysed. Values are means±SEM when applicable. Statistical significance (P<0.05) was tested using a one-way ANOVA followed by a Bonferroni multiple comparisons test to compare between the different conditions.

2. Insulin sceretion assay (Assay 2) human pancreatic islets Batch 2 (absolute values)

Table 18: APPENDIX. Insulin secretion from Assay 2 with islets Batch 2. Single values, means and SEM (abs. values; µU/mL) measured and calculated.

Secreted Insulin (µU/mL)					
Group	**Condition**	**Experiment 1**	**Experiment 2**	**Means**	**SEM**
Without PNGase F and Endo H	2 mM glucose	914.55 977.08	1003.6 1025.23	980.12	23.97
	20 mM glucose	1273.01 1081.01	1077.76 1202.16	1158.48	47.91
	20 mM glucose + 500 nM PMA	1489.49 1278.67	1377.89 1420.35	1391.6	44.11
With PNGase F and Endo H	2 mM glucose	11697.72* 988.48	1112.72 1041.06	1047.42	36.01
	20 mM glucose	1150.47 1106.69	1400.16 1357.33	1253.66	73.29
	20 mM glucose + 500 nM PMA	1375.72 1369.24	1943.95 3899.25*	1562.97	190.50

*omitted

Table 19: APPENDIX. Intracellular insulin from Assay 2 with islets Batch 2. Single values, means and SEM (abs. values; µU/mL) measured and calculated.

Intracellular Insulin (Lysates) (µU/mL)					
Group	**Condition**	**Experiment 1**	**Experiment 2**	**Means**	**SEM**
Without PNGase F and Endo H	2 mM glucose	26729.45 50507.07	16674.79 25997.29	22477.15	2380.86
	20 mM glucose	40018.63 37073.25	13760.08 17090.49	26985.61	6735.78
	20 mM glucose + 500 nM PMA	30457.48 27376.21	15443.09 17529.09	22701.47	3667.95
With PNGase F and Endo H	2 mM glucose	43204.21 46066.01	1500.19 14473.35	34581.19	10087.81
	20 mM glucose	38633.77 40741.23	15688.22 17322.04	28096.32	6714.28
	20 mM glucose + 500 nM PMA	21523.8 35140.43	26863.72 16178.61	24926.64	4043.32

*omitted

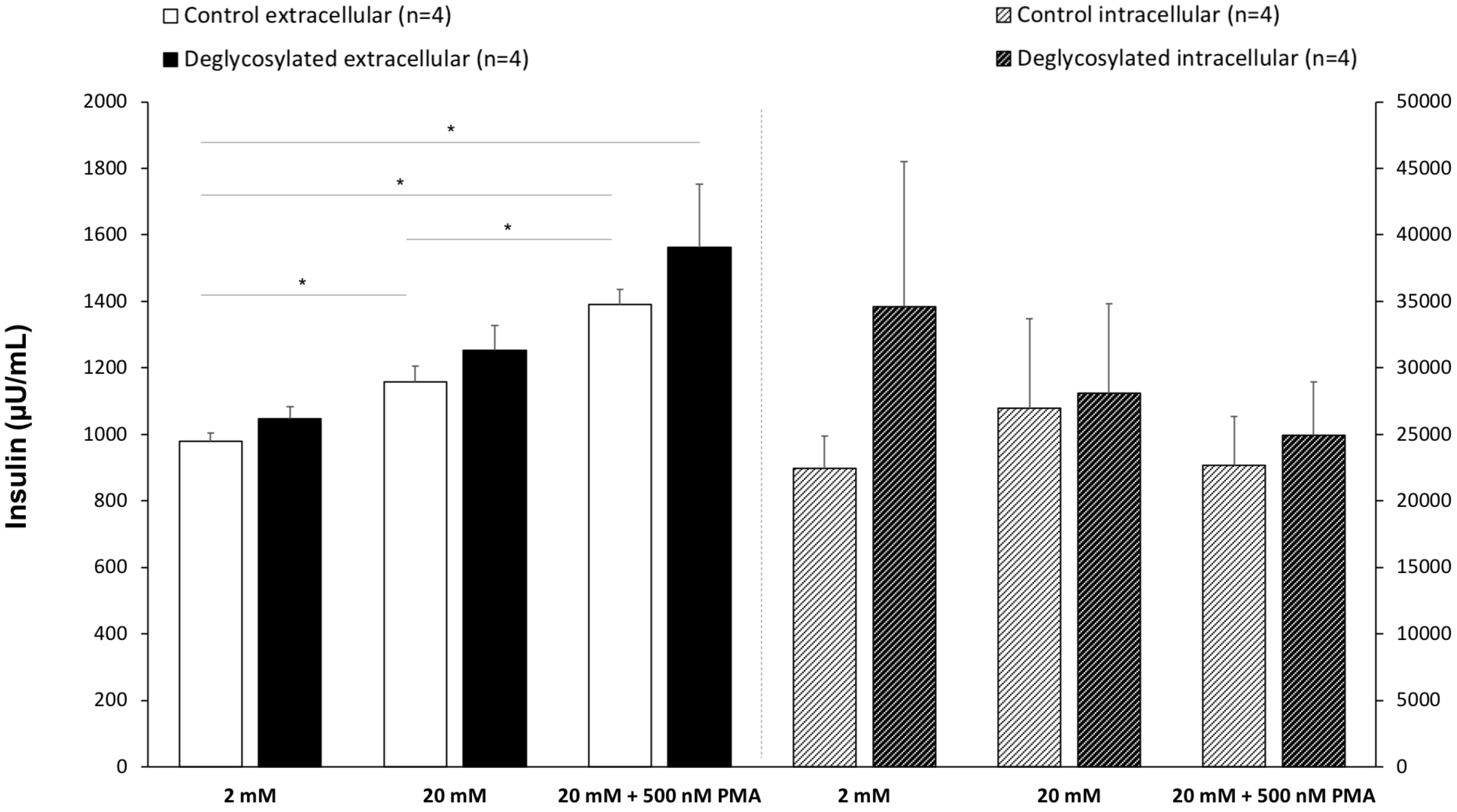


Figure 74: APPENDIX. Assay 2. Extra- and intracellular insulin measurement.

Absolute insulin concentrations (µU/mL) for each condition for non-glycosylated control (white bars) and deglycosylated samples (black bars) are shown. Six samples were treated with PNGase F and Endo H for deglycosylation and six were kept glycosylated in a first incubation step at 2 mM glucose (120 minutes). In total, three different conditions, 2 mM glucose, 20 mM glucose and 20 mM glucose plus 500 nM PMA were tested and completed in BSS (Balanced salt solution) in a second step (60 minutes incubation). Two islets samples were included per condition per experiment. After incubation islets were lysed (hatched bars). Values are means±SEM when applicable. Statistical significance (P<0.05) was tested using a one-way ANOVA followed by a Bonferroni multiple comparisons test to compare between the different conditions.

3. 100 bp and 1 kb DNA standards (NEB)

100 bp standard:

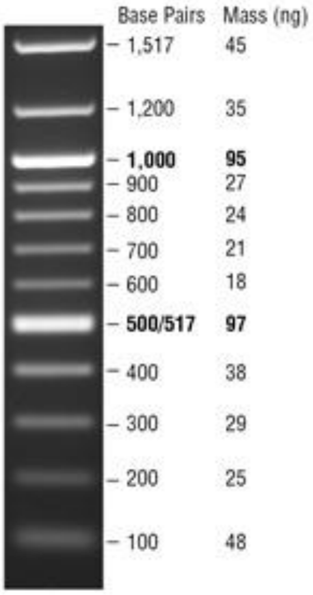


Figure 75: APPENDIX. 100 bp DNA standard.
A number of proprietary plasmids are digested to completion with appropriate restriction enzymes to yield 12 bands suitable for use as molecular weight standards for agarose gel electrophoresis. The digested DNA includes fragments ranging from 100-1,517 base pairs. The 500 and 1,000 base pair bands have increased intensity to serve as reference points. The approximate mass of DNA in each of the bands is provided (assuming a 0.5 µg load) for approximating the mass of DNA in comparably intense samples of similar size. 100 bp DNA Ladder visualised by ethidium bromide staining on a 1.3% TAE agarose gel. Mass values are for 0.5 µg/lane. Information taken from NEB.

1 kb standard:

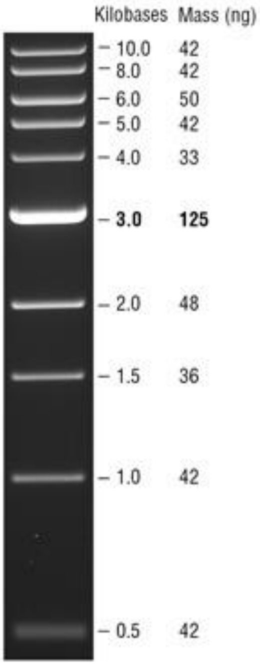


Figure 76: APPENDIX. 1 kb DNA standard.
A number of proprietary plasmids are digested to completion with appropriate restriction enzymes to yield 10 bands suitable for use as molecular weight standards for agarose gel electrophoresis. The digested DNA includes fragments ranging from 0.5-10.0 kilobases (kb). The 3.0 kb fragment has increased intensity to serve as a reference band. The approximate mass of DNA in each of the bands is provided (assuming a 0.5 µg load) for approximating the mass of DNA in comparably intense samples of similar size. 1 kb DNA Ladder visualized by ethidium bromide staining on a 0.8% TAE agarose gel. Mass values are for 0.5 µg/gel lane. Information taken from NEB.

4. Ethical considerations

DNA and tissue samples (if applicable) were obtained after written consent from the patient and/or parents according to protocols approved by the Institute of Child Health/Great Ormond Street Hospital Research Ethics Committee (REC reference number: 05/Q0508/6).

List of publications

Cabezas OR, Flanagan SE, Stanescu H, García-Martínez E, Caswell R, Lango-Allen H, Antón-Gamero M, Argente J, Bussell AM, Brandli A, Cheshire C, Crowne E, Dumitriu S, Drynda R, Hamilton-Shield JP, Hayes W, Hofherr A, Iancu D, Issler N, Jefferies C, Jones P, Johnson M, Kesselheim A, Klootwijk E, Koettgen M, Lewis W, Martos JM, Mozere M, Norman J, Patel V, Parrish A, Pérez-Cerdá C, Pozo J, Rahman SA, Sebire N, Tekman M, Turnpenny PD, Hoff WV, Viering DHHM, Weedon MN, Wilson P, Guay-Woodford L, Kleta R, Hussain K, Ellard S, Bockenhauer D. Polycystic Kidney Disease with Hyperinsulinemic Hypoglycemia Caused by a Promoter Mutation in Phosphomannomutase 2. J Am Soc Nephrol 28, 2529-2539, doi:10.1681/ASN.2016121312 (**2017**).

Kesselheim A, Ashton E & Bockenhauer D. Potential and pitfalls in the genetic diagnosis of kidney diseases. Clin Kidney J 10, 581-585, doi:10.1093/ckj/sfx075 (**2017**).

Reichold M, Klootwijk ED, Reinders J, Otto EA, Milani M, Broeker C, Laing C, Wiesner J, Devi S, Zhou W, Schmitt R, Tegtmeier I, Sterner C, Doellerer H, Renner K, Oefner PJ, Dettmer K, Simbuerger JM, Witzgall R, Stanescu HC, Dumitriu S, Iancu D, Patel V, Mozere M, Tekman M, Jaureguiberry G, Issler N, Kesselheim A, Walsh SB, Gale DP, Howie AJ, Martins JR, Hall AM, Kasgharian M, O'Brien K, Ferreira CR, Atwal PS, Jain M, Hammers A, Charles-Edwards G, Choe CU, Isbrandt D, Cebrian-Serrano A, Davies B, Sandford RN, Pugh C, Konecki DS, Povey S, Bockenhauer D, Lichter-Konecki U, Gahl WA, Unwin R, Warth R, Kleta R. Glycine Amidinotransferase (GATM), Renal Fanconi Syndrome, and Kidney Failure. J Am Soc Nephrol 29, 1849-1858, doi:10.1681/ASN.2017111179 (**2018**).

Attended meetings

A. Kesselheim, E. Klootwijk, T. Dennison, V. Patel, Ben Walsh, R. Kleta, D. Bockenhauer. Understanding an unusual form of Renal Fanconi Syndrome. Annual Division of Medicine Research Retreat. University College London, London, **2016**. **Poster**.

A. Kesselheim. The R85W mutation in HNF4A and its role in Renal Fanconi Syndrome. Renal Academic Seminar Division of Medicine. Royal Free Campus University College London, London, **2016**. **Oral communication**.

A. Kesselheim, C. Cheshire, H. Stanescu, V. Patel, Ben Davies, R. Kleta, D. Bockenhauer. Generation of a cell model . Annual Division of Medicine Research Retreat. University College London, London, **2018**. **Poster**.

A. Kesselheim. Generating a model system for HIPKD. Renal Academic Seminar Division of Medicine. Royal Free Campus University College London, London, **2017**. **Oral communication**.