# MODELLING LIVER DISEASE IN VITRO AND IN SILICO

Li, Chaozheng

*Awarding institution:*
King's College London

# MODELLING LIVER DISEASE *IN VITRO* AND *IN SILICO*

**Chaozheng (Rex) Li**

This dissertation is submitted to King's College London for the Degree of Doctor of Philosophy in Stem Cells & Regenerative Medicine

Faculty of Life Sciences and Medicine

King's College London

DEC 2021

Supervisors:

1. Dr Tamir Rashid

   Clinical Reader in Liver Regeneration

2. Prof Richard Thompson

   Professor of Molecular Hepatology

3. Dr Inês Cebola

   Advanced Research Fellow

This thesis, submitted for the degree of Doctor of Philosophy, was completed under the supervision of Dr Tamir Rashid, Prof Richard Thompson and Dr Inês Cebola. The work described within was carried out as part of my doctoral training across two laboratories:

Centre for Stem Cells & Regenerative Medicine, King's College London,
Guy's Hospital, Tower Wing 28th Floor, London, SE1 9RT

Department of Metabolism, Digestion and Reproduction, Imperial College London,
ICTEM, Hammersmith Campus, London, W12 0NN

I, Chaozheng (Rex) Li, 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 text.

In memory of my grandparents
De Li and Shunling Chen

# Abstract

Although the first progressive familial intrahepatic cholestasis (PFIC) secondary to the *TJP2* mutation patient was reported in early 2014, an effective *in vitro* model for this disease is still lacking. During my PhD, I generated a human-induced Pluripotent Stem Cell (iPSC)-derived hepatocyte (iHEPs)-based high-throughput functional assay, which mimicked the disease phenotype in patients, and hence, it could be used for drug screening and mechanistic studies. Both PFIC patients with tight junction protein 2 (*TJP2*) mutation fibroblast-reprogrammed iPSCs and clustered regularly interspaced short palindromic repeats-CRISPR-associated protein (CRISPR-Cas9)-engineered iPSC isogenic control clones with *TJP2* knock out (KO) were successfully generated. Additionally, an ATP Binding Cassette Subfamily B Member 4 (*ABCB4*) homozygous knock-out iPSC clone was generated by using CRISPR-Cas9, but it needs to be characterised further. Both disease iPSC lines were successfully differentiated to hepatocyte-like cells. In addition, functional bile salt export pump (BSEP)/Multidrug-Resistant protein 2 (MRP2) substrates 5-(and-6)-carboxy-2',7' – dichlorofluorescein diacetate (CDFDA) – transportation and chicken-wire-like canaliculi structures were observed in healthy iHEP cultured in a Matrigel sandwich. Interestingly, both disease iHEPs demonstrated abnormal CDFDA transportation and an irregular canaliculi shape, which partially resembled the patients' liver phenotypes. Furthermore, as genome-wide association study (GWAS) data of drug-induced liver injury (DILI) patients and Encyclopedia of DNA Elements (ENCODE) project phase 3 information became available in 2020, *in silico* analysis was used to curate a list of DILI-related single nucleotide polymorphisms (SNPs) (excluding human leukocyte antigen (HLA)-related SNPs) located within cis-regulated elements (CREs), from which a CRISPR activator/inhibitor (CRISPRa/i) screening library was designed to investigate how those SNPs in CREs caused DILI. There were 26 novel SNPs identified in CRE which could contribute to DILI, but they were unable to be used in the CRISPR screen due to their relatively low quantity. Overall, this study suggested iHEPs can be used to model PFIC with *TJP2* mutation, as it demonstrated the relevant patient liver disease phenotype. Moreover, iHEPs combined with *in silico* analysis and CRISPRa/i screening can be used to model complex disorders, such as DILI.

# Acknowledgments

I first wish to thank my supervisors, Dr Tamir Rashid, Prof Richard Thompson and Dr Inês Cebola, for the opportunity to join their laboratories, their invaluable contributions to this project, as well as their continuous support to my personal development as a scientist and an entrepreneur. This body of work would also not have been possible without the funding provided by the King's Medical Research Trust/Joint Research Committee (KMRT/JRC).

I would also like to recognize the contributions, help, and mentorship given to me by Dr. Toru Hiratsuka, Dr. Aamir Ahmed, and Dr. Soon Seng Ng. Moreover, their guidance and support helped me to get over disappointments. and more importantly boosted my confidence to face all the challenges in front of me. Additionally, I must thank the fellow Ph.D. students and research staff whom worked alongside me during these past years. Without your help, this thesis could hardly existed. Finally, I must thank Prof. Fiona Watt and Dr. Kif Liakath-Ali for the foundational scientific training and more importantly to allow me to work in the world-leading stem cell research center.

Finally, and most importantly, I would like to acknowledge my whole family as their support is the foundation of all I achieve. Thank you for all the financial and emotional support over the past four years. The unconditional love from both of my parents has been the safe harbor for me whenever I needed help. In addition, I would like to thank my girlfriend Jessie Xu for all the great supports at the end of my PhD study. This Ph.D. journey truly justified the family philosophy you had taught me throughout my life, which ultimately leads to a fruitful and rewarding 4 years. I am eternally grateful to both of you.

# Publications (both under review)

1) **Li, C.Z**., Ogawa, H., Ng, S.S., Baker, O., Kishimoto, E., Sakabe, K., Hellmann, J., Soares, F., Tasnova, N.L., Chen, X., Ma, L., Xiao, F., Miethke, A., Tang, Z., Danovi, D., Blackford, S.J.I., Sambrotta, M., Tavabie, O., Thompson, R.J., Rashid, S.T., Asai, A. (2020). Human iPSC models cholestasis with Tight Junction Protein 2 (TJP2) deficiency. *Journal of Hepatology Reports*.

2) Kent, D., Ng, S.S., Khoshkenar, P., Zieger, M., Greer, C., Hatch, S., Segal, J.M., Blackford, S.J.I., **Li, C.Z.**, Chowdary, V., Ismali, T., Sahdeo, S., Danovi, D., Ebner, D., Mueller, C., Rashid, S,T. (2020). CZC-25146, a small molecule inhibitor of LRRK2, reduces polymers and inflammation in patient-derived iPS-hepatocytes and mouse models of Alpha-1 Antitrypsin Deficiency. *Journal of Clinical Investigation.*

# Table of Contents

# List of figures

# List of tables



# Appendix 1 contents

# List of abbreviations

| | |
|---|---|
| 4-PBA: | 4-phenylbutyric acid |
| A1ATD: | Alpha1 Antitrypsin Deficiency |
| ABC: | ATP Binding Cassette Subfamily |
| ABCB11: | ATP Binding Cassette Subfamily B Member 11 |
| ABCB11: | ATP Binding Cassette Subfamily B Member 11 |
| ABCB4: | ATP Binding Cassette Subfamily B Member 4 |
| ABCC2: | ATP Binding Cassette Subfamily C Member 2 |
| ALB: | Albumin |
| AMPK: | 5’ AMP-activated protein kinase |
| ASGR2: | Asialoglycoprotein receptor 2 |
| ATAC seq: | Assay for Transposase-Accessible Chromatin using sequencing |
| ATP7b: | Wilson disease protein |
| ATP8B1: | Probable phospholipid-transporting ATPase IC |
| BA: | Biliary Atresia |
| BCL2L11: | BCL2 like protein 11 |
| BEI: | Biliary Excretory Index |
| BMP4: | Bone Morphogenetic Protein4 |
| BSA: | Bovine Serum Albumin |
| BSEP: | Bile Salt Export Pump |
| c-MET: | Tyrosine-protein kinase Met |
| C1 to C8: | Colony1 to Colony8 |
| cAMP: | cyclic Adenosine Monophosphate |
| CARM1: | Coactivator-Associated Arginine Methyltransferase 1 |
| CBZ: | Carbamazepine |
| CDC123: | Cell Division Cycle 123 |
| CDFDA: | 5-(and-6)-carboxy-2′,7′-dichlorofluorescein |
| CDH9: | Cadherin 9 |
| CDM: | Chemically Defined Medium |
| CGAS: | Candidate Gene Association Study |
| ChIP-seq: | Chromatin Immunoprecipitation-sequencing |

| | |
|---|---|
| CHOP: | C/EBP Homologous Protein |
| CIX: | Cholestatic Index |
| CK18: | Anti-Cytokeratin18 |
| CK7: | Cytokeratin-7 |
| CLDN1: | Claudin1 |
| CLF: | Cholyl-Lysyl-Fluorescein |
| CPZ: | Chlorpromazine |
| CRE: | Cis-Regulatory Element |
| CRISPRa/i: | CRISPR activator/inhibitor |
| CRISPR-Cas9: | Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated protein |
| crRNA: | CRISPR RNA |
| CU: | Copper |
| CYP2E1: | Cytochrome 2E1 |
| *CYP3A4*: | *Cytochrome P450 3A4* |
| dCAS9–SAM: | dCAS9 Synergistic Activation Mediator |
| DDW: | Deuterium Depleted Water |
| DFNA51: | Deafness, Autosomal Dominant 51 |
| DHS: | DNA Hypersensitive Regions |
| DILI: | Drug-Induced Liver Injury |
| DNA: | Deoxyribonucleic acid |
| DS: | Donkey Serum |
| E-BOX: | Enhancer Box |
| E6: | Essential 6 media |
| E8: | Essential 8 media |
| EBP: | Emopamil Binding Protein |
| eCLIP: | Enhanced version of the crosslinking and immunoprecipitation assay |
| ECM: | Extracellular Matrix |
| EGFR: | Epidermal Growth Factor Receptor |
| ELK-1: | ETS Like-1 protein |
| EMX1: | Empty Spiracles Homeobox 1 |

| | |
|---|---|
| ENCODE: | Encyclopedia of DNA Elements |
| EP: | Endocochlear Potential |
| Epac: | Exchange factor directly activated by cAMP1 |
| ESC: | Embryonic Stem cells |
| FACS: | Fluorescence Activated Cell Sorting |
| FBXO15: | F-box protein 15 |
| FDA: | Food & Drug Administration |
| FGF2: | Fibroblasts Growth Factor 2 |
| Foxa2: | Fork – head box protein A2 |
| FRAP: | Fluorescence Recovery After Photobleaching |
| FSC: | Forward Scatter |
| FWD: | Forward |
| FXR: | Farnesoid X Receptor |
| GABRB2: | Gamma-Aminobutyric Acid Type A Receptor Subunit Beta2 |
| Gamma-GGT: | Gamma-Glutamyl Transferase |
| GAPDH: | Glyceraldehyde 3-phosphate dehydrogenase |
| gRNA: | guide RNA |
| GSK-3β: | Glycogen Synthase Kinase 3 beta |
| GWAS: | Genome-Wide Association Study |
| H3K27ac: | Histone H3 lysine 27 acetylation |
| H3K27me3: | Histone 3 lysine27 methyl group 3 |
| H3K4 me3: | Histone3 lysine4 methyl group 3 |
| HBL-MPS-DB: | Human Biomimetic Liver-Microphysiology Database |
| HBSS: | HANKs Balanced Salt Solution |
| HDAC: | Histone Deacetylases |
| HDR: | Homologous Directed Repair |
| HIV: | Human Immunodeficiency Virus |
| HLA-DRB1: | Major histocompatibility complex, class2, DR Beta1 |
| HLA: | Human Leukocyte Antigen |
| HLO: | Human iPSC-Derived Liver Organoid |
| HNF4: | Hepatocyte Nuclear Factor 4 Alpha |

| | |
|---|---|
| HuH-7: | Human Hepatocarcinoma Cell 7 |
| ICC: | Inverted Colloidal Crystals |
| iHEP: | iPSC-derived Hepatocyte-like cells |
| INK4a: | P16 |
| iPSC: | induced Pluripotent Stem Cell |
| iRSC: | intermediately Reprogrammed Stem Cell |
| JMJD3: | Lysine Demethylase 6B |
| *Klf4*: | Kruppel-like factor 4 |
| KO: | Knock Out |
| KRAB: | Kruppel-Associated Box |
| LATS: | Large tumour suppressor kinase |
| LC/MS: | Liquid Chromatography/Mass Spectrometry |
| LIF: | Leukaemia Inhibitor Factor |
| LKB1: | Liver Kinase B1 |
| LSEC: | Liver Sinusoidal Endothelial Cells |
| MDCK cell: | Madin-Darby Canine Kidney |
| *MDR2*: | Multidrug-resistant protein 2 |
| *MDR3*: | Multidrug-resistant protein 3 |
| MEFs: | Mouse Embryonic Fibroblasts |
| MEK: | Mitogen-activated protein kinase |
| MET: | Mesenchymal-to-Epithelial Transition |
| MHC-1: | Major Histocompatibility Complex-1 |
| Mins: | Minutes |
| MPCCs: | Micropatterned Co-cultures |
| mRNA: | messenger RNA |
| MRP2: | Multidrug-Resistant Protein 2 |
| MRP4: | Multidrug-Resistant Protein 4 |
| MST/STK4: | Serine/Theronine Kinase 4 |
| NAT2: | N-Acetyltransferase 2 |
| NET-seq: | Native Elongating Transcript sequencing |
| NFKB: | Nuclear factor kappa-light-chain-enhancer of activated B cells |

| | |
|---|---|
| NGS: | Next-Generation Sequencing |
| NHEJ: | Nonhomologous End Joining |
| NLRV: | Nonpotential Liver Regulatory Variants |
| NLS: | Nuclear Localisation Signal |
| NMD: | Non-sense Mediated mRNA Degradation |
| NPC: | Neural Progenitor Cells |
| NSC: | Neural stem Cells |
| NTCP: | Sodium/taurocholate transporting polypeptide |
| OATP: | Organic-Anion-Transporting Polypeptide |
| Oatp2: | Organic-Anion-Transporting Polypeptides 2 |
| OCT4: | Octamer-binding transcription factor 4 |
| OHC: | Outer Hair Cells |
| oriP/EBNA1: | the origin of viral replication/Epstein–Barr virus nuclear antigen 1 |
| PAM: | Protospacer Adjacent Motif |
| PBDE-99: | Polybrominated Diphenyl Ethers-99 |
| PBMC: | Peripheral Blood Mononuclear Cell |
| PBS: | Phosphate-Buffered Saline |
| PC: | Phosphatidyl Choline |
| PCL: | Precision-Cut Liver |
| PCR: | Polymerase Chain Reaction |
| PDD: | Phenotypic Drug Discovery |
| PDMS: | Polydimethylsiloxane |
| PDX1: | Pancreatic and Duodenal Homeobox 1 |
| PDZ: | Post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (ZO-1) |
| PEG: | Polyethylene Glycol |
| pegRNA: | prime editing guide RNA |
| PFA: | Paraformaldehyde |
| PFIC: | Progressive Familial Intrahepatic Cholestasis |
| PI: | Propidium Iodide |
| PKA: | Protein Kinase A |

| | |
|---|---|
| PPARG: | Peroxisome Proliferator-Activated Receptor gamma |
| PRMT1: | Protein arginine methyltransferase |
| PSC: | Primary Sclerosing Cholangitis |
| qPCR: | Quantitative polymerase chain reaction |
| Rap1: | Ras related protein-1 |
| Ras: | Small GTPase |
| REV: | Reverse |
| RNA: | Ribonucleic acid |
| RNP: | Ribonucleoprotein complex |
| ROCKi: | Rho-associated protein kinase inhibitor |
| ROI: | Region of Interest |
| RPS3A: | 40S Ribosomal Protein S3a |
| RT-PCR: | Reverse Transcription-Polymerase Chain Reaction |
| RXRa: | Retinoid X Receptor alpha |
| SAF-B: | Scaffold Attachment Factor-B |
| SCH: | Sandwich Culture Hepatocytes |
| SCNT: | Somatic Cell Nuclear Transfer |
| SERPINA1: | Serpin family A member 1 |
| SH3D19: | SH3 Domain Containing 19 |
| shRNA: | short hairpin RNA |
| SNP: | Single Nucleotide Polymorphism |
| Sox2: | Sex determining region Y 2 |
| SQL-SAL: | Sequentially Self-Assembly Liver Model |
| SSC: | Side Scatter |
| ssODN: | single strand Oligo Nucleotide |
| SYT14: | Synaptotagmin 14 |
| T2D: | Type 2 Diabetes |
| TAD: | Topological/Topologically Associated Domain |
| TALEN: | Transcription Activator-Like Effector Nucleases |
| TAOK: | TAO Kinase |
| TAZ: | Transcriptional activator with PDZ-binding domain |

| | |
|---|---|
| TDD: | Target-Based Drug Discovery |
| TEAD: | Transcriptional enhancer fact1or TEF-1 |
| TEM: | Transmission Electron Microscopy |
| TGN: | Transgolgi Network |
| TJP2: | Tight Junction Protein 2 |
| TNT4: | Cellosaurus cell line MSU-TnT4 |
| TPD52: | Tumour Protein D52 |
| tracrRNA: | Trans-activating crRNA |
| trxG: | Trithorax complex |
| UCCA: | Microscale cell culture analogue |
| UDCA: | Ursodeoxycholic acid |
| USP53: | Ubiquitin Specific Peptidase 53 |
| UV: | Ultraviolet |
| VDR: | Vitamin D Receptor |
| VGLL4: | Vestigial-like family member 4 |
| Wdr5: | WD-repeat protein-5 |
| WES: | Whole Exon Sequencing |
| WIF-B9: | Rat hepatoma/human fibroblast hybrid cell line |
| WT: | Wild Type |
| YAP: | Yes-Associated Protein |
| ZASP: | ZO-2 Associated Speckle Protein |
| ZNF: | Zinc Finger |
| ZO2: | Zonula Occludin-2 |
| β-geo: | β-galactosidase |

# Chapter 1: Making PFIC-relevant iPSC models by CRISPR-Cas9 engineering or by reprogramming patient fibroblasts

# 1.1 Introduction

## 1.1.1 Cholestasis secondary to *TJP2* and *ABCB4* mutation backgrounds

Human Zonula Occludins-2 (ZO2) is a 133958 Da protein encoded by the gene *TJP2* located on chromosome 9q (9q21.11) (Duclos et al., 1994). The human *TJP2* gene spans more than 80 kb in length and translates into a protein that is 1,190 amino acids long (Fig1.1Ai). There are two isoforms, ZO2A and ZO2C, which were transcribed from separate promoters. ZO2C isoforms do not have the N terminal-23 amino acid peptides present in ZO2A. Northern blot results reveal ZO2A has higher expression in heart and brain tissues; ZO2C, however, has higher expression in the kidneys, pancreas, heart and placenta (Chlenski et al., 2000). The ZO2 protein encoded by the *TJP2* gene has been reported to have multiple functions in various cell types: first, immunoprecipitation results from ZO2-deficient mice epithelial cells, suggesting that ZO2 can be recruited to the tight junction and interact with Claudin1 (CLDN1) through Claudin8 (CLDN-8) through its PSD95, Dlg1, ZO-1(PDZ1) domain (Fig1.1Aii) (Itoh et al., 1999). Second, mammalian epithelial cell evidence suggests that ZO2 interacts with DNA-binding protein Scaffold Attachment Factor-B (SAF-B) in the nucleus (Traweger et al., 2003). Furthermore, over-expression of *TJP2* in human hair cells leads to decreased phosphorylation of Glycogen Synthase Kinase 3 beta (GSK-3β) and altered expression of genes that regulate apoptosis. This helps to explain the mechanism of progressive nonsyndromic hearing loss *Deafness, Autosomal Dominant 51* (*DFNA51*) with *TJP2* gene duplication (Walsh et al., 2010). More recently, ZO-2 Associated Speckle Protein (ZASP) has been id entified to interact with ZO2 in Madin-Darby Canine Kidney (MDCK) cells by immunoprecipitation assay. ZASP blocks the inhibitory activity of ZO2 on cyclin D1 gene transcription (Lechuga et al., 2010). Overall, ZO2 protein is located within the cells' tight junction and links cell membrane protein claudin with intracellular actin filaments (Fig1.1Aiii). ZO2's presence is essential for maintaining normal hepatocyte paracellular space physiology (Sambrotta et al., 2014a).

PFIC is an umbrella term that describes a group of several genetically discrete cholestasis diseases. The molecular mutations in PFIC patients lead to defective canalicular transport, tight junction formation, vesicular trafficking, and membrane maintenance in the liver. This consequently leads to a failure of bile acid and lipid homeostasis, which are the key causes of intrahepatic cholestasis

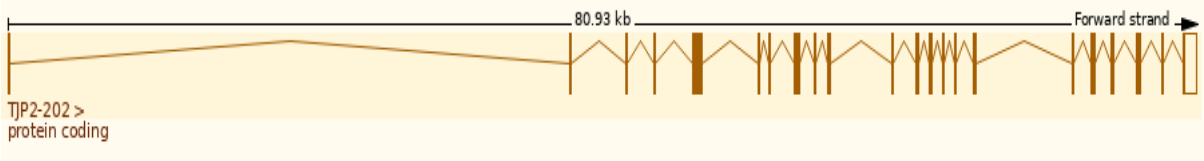
(Bull and Thompson, 2018). PFIC secondary to *TJP2* deficiency was discovered using high-throughput Next-Generation Sequencing (NGS) and Whole Exon Sequencing (WES) approaches (Sambrotta et al., 2014a). Bi-allelic *TJP2* mutations reported in PFIC patients lead to complete loss of ZO2 protein expression and function. All patients with complete loss of ZO2 function have early-onset progressive liver disease, with liver transplantation being the only effective treatment option. If left untreated, PFIC could lead to death due to liver failure. However, due to a shortage of organ donors and the inconvenience of lifelong immunosuppressants post-transplantation, there is an urgent need to understand the mechanism of the disease and develop novel therapies. Attempts to KO ZO2 in mouse embryos have been lethal. This contrast with patients with *TJP2* mutation further suggests the interspecies difference between humans and mice. Interestingly, the *TJP2* mutation is only associated with cholestasis, not cholangiopathy, which suggests the tight junction barrier function is not badly disrupted. Transmission electron microscopy (TEM) evidence suggests there is a disruption of tight junction structure in patient liver samples; missing ZO2 also leads to a dislocation of the other tight junction protein, CLDN1, as confirmed by immunohistochemistry (Bull and Thompson, 2018; Sambrotta et al., 2014a).

Multidrug-resistant protein 3 (MDR3) is a 141 kDa protein encoded by the *ABCB4* gene, located on chromosome 7 (7q21.12) (Van der Bliek et al., 1987). The *ABCB4* gene spans 73.69 kb in length and translates into a protein that is 1,286 amino acids long (Fig1.1Bi) (Van der Bliek et al., 1987; Lincke et al., 1991). There are three main isoforms of human MDR3 with a slight difference in amino acid sequences. Isoform A localises to the cell membrane of human hepatic cells, such as Human Hepatocarcinoma Cell 7 (HuH-7), whereas Isoforms B and C are expressed only in the cytoplasm (Weber et al., 2019). The first functional study in mice confirmed Multidrug-resistant protein 2 (mdr2) (the homologue of human MDR3) acts as a lipid translocase that is responsible for Phosphatidyl Choline (PC) transportation from the inner to the outer leaflet of the membrane (Fig1.1Bii). This process is strictly adenosine triphosphate (ATP) and magnesium ion-dependent (Ruetz and Gros, 1994). In mice with bi-allelic *ABCB4* gene disruption, phospholipid secretion is negligible at every bile salt output rate tested. This suggests that mdr2 transports PC from the inner to the outer leaflet of the canalicular membrane, where it forms a protective layer for the bile micelles to prevent bile acid damage to the canaliculi wall (Fig1.1Bii). Overall, MDR3 plays a crucial role in maintaining phospholipid homeostasis on hepatocytes' apical/canalicular

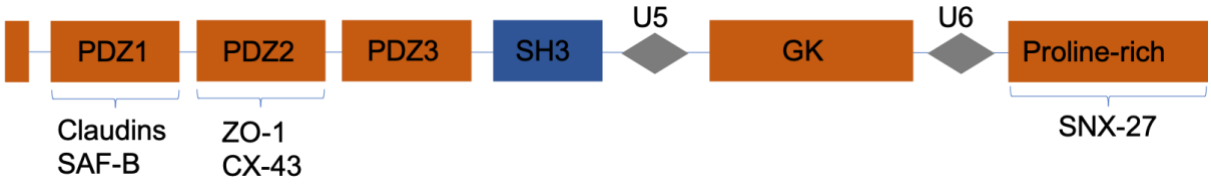
membranes by translocating phospholipids such as PC (Fig1.1Biii). More importantly, MDR3 transporting PCs in canaliculi lumen could create harmless mixed micelles with bile acid and cholesterol; therefore, lumen surfaces are protected from pure bile acids with detergent natures.

Patients with MDR3 deficiency often have no *ABCB4* messenger RNA (mRNA) and are associated with increased serum levels of Gamma-Glutamyl Transferase (GGT). Most patients develop cholangiopathy, and even 50% MDR3 loss of function is enough for damage to occur. Currently, both transplantation and small molecule treatments are options to treat PFIC with MDR3-deficient patients. For example, Ursodeoxycholic acid (UDCA) supplementation could help patients with milder symptoms to reduce the detergent nature of bile (Bull and Thompson, 2018). However, patients with severe symptoms still require liver transplantation to survive.

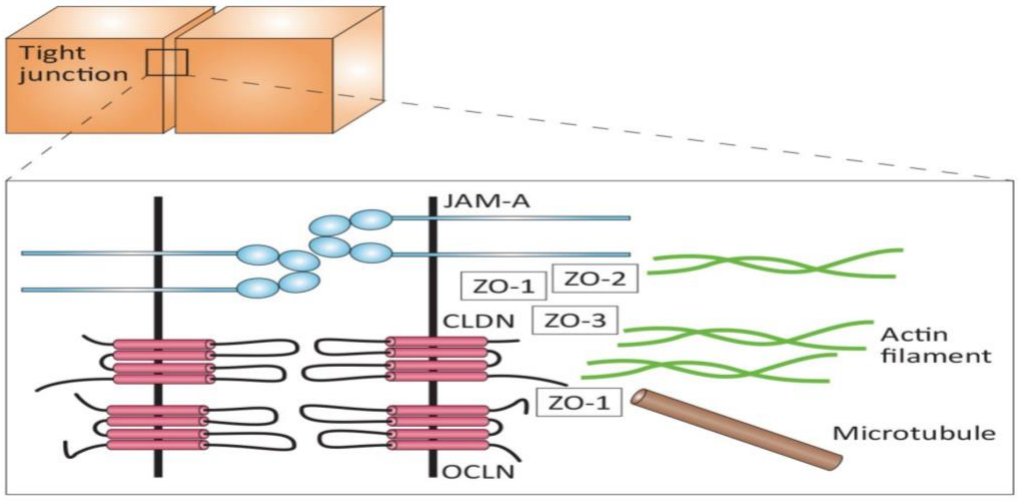
(Ai)



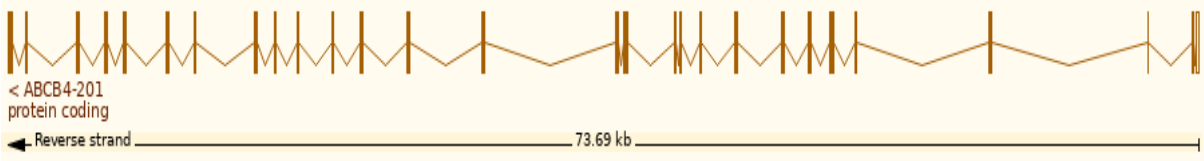

(Aii)



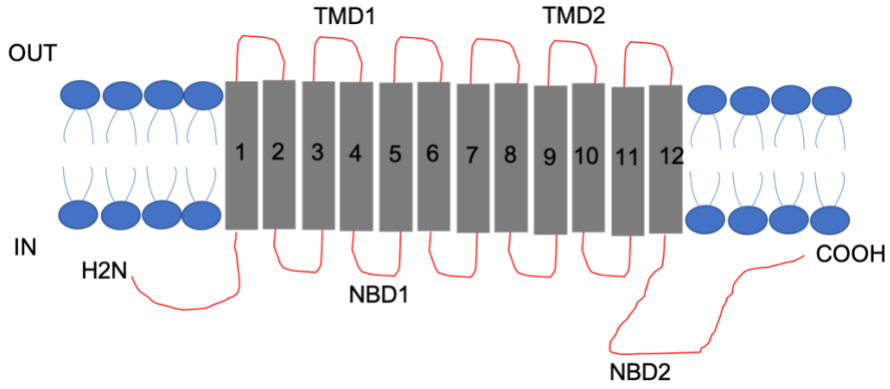

(Aiii)



(Bi)


< ABCB4-201
protein coding
Reverse strand
73.69 kb

(Bii)


TMD1
TMD2
OUT
IN
1
2
3
4
5
6
7
8
9
10
11
12
H2N
COOH
NBD1
NBD2

(Biii)

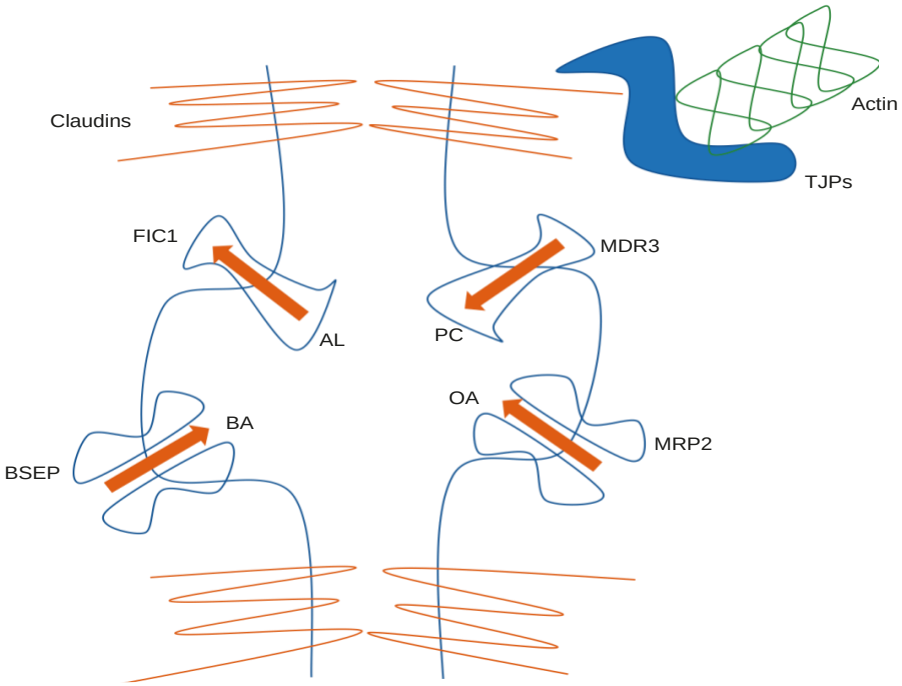

Claudins
Actin
TJPs
FIC1
AL
MDR3
PC
BA
BSEP
OA
MRP2

**Figure 1.1 TJP2 and ABCB4 summary** (Ai) Schematic illustration of *TJP2*-202 transcript, which contains 23 exons (yellow rectangular box). The *TJP2* gene has a total length of 80.93 kb. Figure adapted from Ensembl.org (Aii) The structure of ZO2 includes three PDZ domains, a SH3 domain, a Guanylate Kinase (GK)-like domain and a proline-rich region in the C-terminus. There are two unique (U5-U6) regions identified. ZO2's PDZ1 region is shown to interact with claudins and SAF-B proteins. ZO2's PDZ2 region is shown to interact with ZO1 and CX43 (connexin 43). ZO2's proline-rich region is shown to interact with Sortin Nexin 27 (SNX27). Figure was adapted from (Bauer et al., 2010) (Aiii) Simplified representation of tight junction structure in epithelial cells. Tight junctions are composed of integral proteins, such as claudins, which link to the actin cytoskeleton through a complex cytoplasmic structure. Tight junction proteins ZO1 and ZO2 are the elements represented. Figure was adapted from (Mirjam et al., 2019) (Bi) Schematic illustration of *ABCB4*-201 transcript, which contains 28 exons (yellow rectangular box). The *ABCB4* gene has a total length of 73.69 kb. Figure adapted from Ensembl.org (Bii) The structure of MDR3 includes one N terminal and one C terminal. There are two trans membrane domains (TMDs), TMD1 and TMD2, in MDR3. In addition, two nucleotide-binding domains (NBDs) are presented in MDR3. Figure was adapted from (Rosmorduc and Poupon, 2007) (Biii) Schematic overview of major apical membrane proteins of hepatocytes involved in genetic cholestasis. BSEP is the major bile acid transporter. MRP2 transports organic anion (OA) conjugates. MDR3 flops PC from the inner canaliculi membrane to the outer canaliculi membrane. Familial intrahepatic cholestasis 1 (FIC1) flips aminophospholipids (ALs). Two adjacent hepatocyte apical membranes create canalicular space sealed by tight junction complexes. Arrows indicate the direction of the apical membrane protein's transportation. Figure was adapted from (Sambrotta and Thompson, 2015a).

### 1.1.2 iPSC history and its application in modelling liver diseases

Waddington's landscape, which has long been accepted as the basic rule of developmental biology, began in the twentieth century. As Waddington said, 'You were looking at an incline called the Hump. The wagons were pushed over the Hump and went running downhill and were sorted out by the systems of points into the various sidings. Now an embryo was in some ways analogous to a set of trucks sliding down the Hump' (Allen, 2015). A big assumption was derived from this theory: cellular differentiation is a one-direction process – from fertilised egg to mature somatic cells. The evidence of reprogramming a somatic cell to its embryonic stage was first shown by Sir John Gurdon, who used nucleus transfer technology to remove the nucleus from a presumptive donor frog somatic cell and put it into a recipient enucleated frog egg. The resulting egg was able to bear a tadpole which was genetically identical to the donor (Fig1.2) (GURDON, 1962). This demonstrated that the cytoplasm of the enucleated egg had changed the presumptive somatic cell's nucleus into a stem cell-like cell as it could give rise to all tadpole cell types.

The ‘Somatic Cell Nuclear Transfer’ (SCNT) technology was applied to more complex mammals; the most famous example was ‘Dolly the sheep’. At the Roslin Institute, scientists took the nucleus from an established sheep somatic cell line, Cellosaurus cell line MSU-TnT4 (TNT4), then transferred it into an enucleated sheep oocyte. The new nucleus-transferred egg cells were able to give rise to a new lamb, and microsatellite data suggested all cells from the new lamb were coming from a single cell population (Fig1.2) (Campbell et al., 1996). Due to ethical restraint, there is no human ‘cloning’ reported, but researchers are trying to use the cytoplasm of human embryonic stem cells (hESCs) to reprogram human somatic cells.

Prof Chad Cowen first showed this in 2005 by fusing human fibroblasts with hESCs. The resulting hybrid cells resembled embryonic stem cell (ESC) characteristics. More complex analysis revealed that the new hybrid cells and the hESC shared common genome-wide transcriptional activity and DNA Methylation patterns. Together, all the data from this study suggests that hESC’ cytoplasm could reprogram the transcriptional state of somatic nuclei. This established the foundation for Yamanaka’s discovery of iPSC (Cowan et al., 2005).

Prof Shinya Yamanaka proposed the hypothesis that there are certain factors in the cytoplasm of human ESCs which are responsible for reprogramming. To test this hypothesis, his colleagues first established Mouse Embryonic Fibroblast (MEF) cell lines, where the β-geo cassette (a fusion of the β-galactosidase and neomycin resistance genes) was inserted into the F-box protein 15 (*FBXO15*) gene by homologous recombination, and ESCs homozygous for βgeo knock in would show resistance to aminoglycoside antibiotic (G418). Then, they transduced a mini library of candidate genes into the MEF cells. Surprisingly, all 24 factors transduced together resulted in a few clones with resembling characteristics of the ESCs, and after a further selection process, they confirmed octamer-binding transcription factor 4 (*OCT4*), Kruppel-like factor 4 (*KLF4*), Sex determining region Y 2 (*SOX2*) and cellular MYC (*C-MYC)*, which demonstrated resistance to G418, were the four factors required to generate new cells; they were named iPSCs (Fig1.2) (Takahashi and Yamanaka, 2016).

During the next 12 months, Okita et al also improved the reprogramming techniques and achieved germline transmission (the ability of stem cells to give rise to progeny when transplanted into a

recipient blastocyst). This suggests that the mouse iPSC derived from mouse somatic cells were able to give rise to an entirely new mouse (Okita et al., 2007). By using the same approach, human iPSCs were soon established from various human somatic cell lines (Takahashi et al., 2007; Yu et al., 2007). From 2007 onwards, iPSC reprogramming research has focused on improving reprogramming efficiency, including adding a tumour protein p53 inhibitor and improving safety to make sure there is no disruption of the genetic content of the iPSC (Hong et al., 2009).

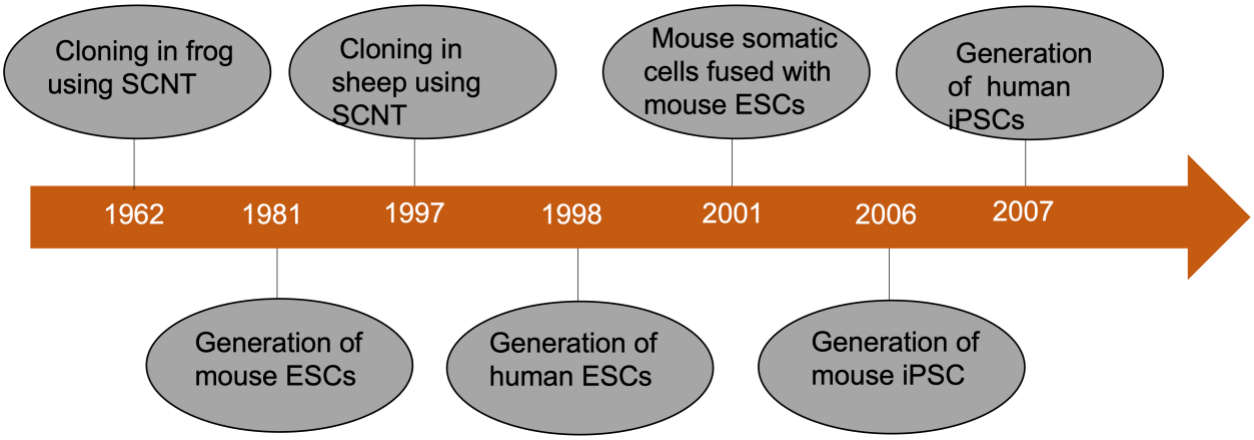



**Figure 1.2 Summary of iPSCs' 10-year developmental history** Schematic illustration of key events during iPSC development. The idea of reprogramming a mature somatic cell can be traced back to 1962 when Sir John Gurdon performed the SCNT in frog egg cells. In 1981 and 1998, the two landmark experiments performed by Sir Martin Evans's team and Prof James Thompson's team led to the discovery of mouse and human ESCs independently. The birth of 'Dolly the sheep' in 1997 marked the success of the first mammal generated by somatic cloning. Later, in early 2000, Yue et al fused mouse somatic cells with mouse ESCs to reactivate pluripotency genes(Yue et al., 2010). Around 2006, Prof Shinya Yamanaka and Prof James Thompson's team discovered different reprogramming factors which successfully generated mouse and human iPSCs from somatic cells independently. Figure was adapted from (Omole and Fakoya, 2018)

### 1.1.3 iPSC and ESC comparison

At the macrolevel, iPSC and ESC both have self-renewal functions and the potential to differentiate to all cell lineages. The two cell types are derived from two different sources, which further shape their downstream applications. In 1998, Prof James Thompson established the first human ESC lines from the inner cell mass of an embryo at the blastocyst stage of development. Since then, stem cell research has been developing at a breathtaking rate (Fig1.2) (Thomson et al , 1998). However, this process requires the destruction of an embryo to isolate the ESCs, and therefore, ethical concerns and a shortage of donors have become the main obstacles for ESC research. However, Prof Yamanaka has derived iPSCs from human somatic cells, which bypasses the ethical problems and increases potential donors compared to human ESCs. Due to the huge potential of ESCs for cell therapy and disease modelling, many people have started to wonder if iPSCs could replace ESCs.

Looking at their global gene expression profiles, Marchetto et al transfected episomal vectors containing Yamanaka factors to foetal neural progenitor cells (NSCs), which successfully created iPSCs, which shared transcriptional signatures with their ESCs counterparts (Marchetto et al., 2009a). However, the study also identified three groups of unique cell type expression. First, early embryonic development genes are specifically expressed in ESCs but not in iPSCs. Second, genes associated with neuronal lineage are not sufficiently suppressed in iPSCs. And, third, reprogramming induced gene expression specifically in iPSCs but not in ESCs or NSCs. Overall, from this transcriptome analysis, iPSCs were found to be globally like ESCs, but they were not indistinguishable (Marchetto et al., 2009b). There was also a study carried out on the global epigenetic profile comparison between iPSCs and ESCs. Chin et al performed genome-wide Chromatin Immunoprecipitation-sequencing (ChIP-seq) on Histone 3 Lysine27 methyl group 3 (H3K27me3) and found promoters' region methylation patterns were almost identical between ESCs and iPSCs. This suggests that during reprogramming of human fibroblast to iPSC, all the H3K27me3 patterns reset back to an ESC state almost completely (Chin et al., 2009).

The next interesting question is to find out if human ESCs and iPSCs use the same molecular signalling pathway to maintain their pluripotency. Prof Ludovic Vallier generated iPSCs by transducing Yamanaka factors in viral vectors and then culturing them in Chemically Defined

Medium (CDM) supplemented with activin and Fibroblast Growth Factor 2 (FGF2). This medium allowed both iPSCs and ESCs to maintain their pluripotency. By using activin receptor-specific inhibitors or by removal of activin from the medium, iPSCs and ESCs started to generate extraembryonic and neuroectoderm differentiation, and thus, their pluripotency was no longer maintained. Later, ChIP and luciferase assay results revealed activin/nodal signalling maintained pluripotency of human induced Pluripotent Stem Cells (hiPSCs) by directly controlling the expression of *NANOG*, as it did in hESCs (Vallier et al., 2009). Most importantly, iPSCs in a hepatocyte-like cell-differentiation protocol could be widely applied to a variety of current good manufacturing practice (cGMP) grade iPSCs and ESCs, suggesting that the protocol could be shared between the two cell types (Blackford et al., 2019a).

Finally, one study reported that hiPSCs preserved epigenetic memory from their donor cell, which, thus, affected its differentiation propensity. Therefore, for certain cell lineage differentiation, hiPSCs could be a better source than hESCs. Bar-Nur et al reprogrammed iPSCs from human pancreatic islet beta cells, and they found such beta cell derived iPSCs (BiPSC) maintained beta cell-like open chromatin structures (marked by H3 acetylation) at key beta-cell genes, such as *INS* and *PDX1*, together with a unique DNA Methylation signature that distinguished them from other PSCs. Scientists then followed the differentiation protocol to turn hESCs and BiPSCs into pancreatic endocrine progenitors and transplanted them into SCID mice by kidney capsule. BiPSC-derived pancreatic progenitors demonstrated detectable serum levels of human C-peptide much faster and in a higher quantity compared to their ESC counterparts (Bar-Nur et al., 2011). This suggests that it might be beneficial to use ‘tissue relevant iPSC’ for a large amount of ‘tissue-relevant somatic cell’ generation. As iPSC reprogramming and ESC generation methods improve, an increasing number of pluripotent cell lines will need to be evaluated for their quality and utility. Therefore, a high-throughput system to assess the new cell lines would be highly desirable.

Scientists developed a scorecard based on new pluripotent stem cells’ global DNA Methylation or gene expression pattern to bioinformatically score the cells based on their performance. First, scientists compare new iPSCs’ DNA Methylation and gene expression profiles with an ESC reference to identify iPS cell lines with epigenetic or transcriptional problems that might interfere with potential desired somatic cell differentiation. Second, each gene’s DNA Methylation and

expression levels are observed for whether they fall within the range observed among ESC lines. Third, genes outside the ESC range are flagged and outlier detection is used to make a 'deviation scorecard'. Last, scientists look for known genes that might specifically interfere with desired somatic cell differentiation from the deviation scorecard. After bioinformatic inspection, the new pluripotent stem cell is assessed using an Embryoid bodies (EB) differentiation assay to quantify the cell's differentiation propensity. After 16 days of nondirected EB differentiation, mature EB are subjected to expression profiling for 500 lineage marker genes and quantified for the expression differences versus reference. Finally, scientists carry out gene set enrichment analysis for lineage marker genes and generate a differentiation propensity scorecard for the pluripotent stem cell (Bock et al., 2011).

To conclude, future iPSC and ESC comparison will have to be more standardised in methods of reprogramming, isolation, or differentiation and, more importantly, a reproducible high-throughput score system to assess their quality and differentiation propensities. The two cell types from the current analysis suggest they are like identical twins – very hard to tell apart from afar, but looking closely, it is possible to tell the difference. I believe such differences are fundamentally important to all pluripotent stem cell downstream applications, such as cell therapy and disease modelling. Only deep studies like these can give us sufficient insight into the potential efficacy and safety of stem cell-related products and which cell type will be more favourable to future clinical applications (Puri and Nagy, 2012).

### 1.1.4 iPSC reprogramming mechanism

Human iPSC reprogramming is an extremely inefficient and stochastic process. In order to use iPSC to make personalised disease models or cell therapy products, an improved understanding of human iPSC reprogramming is required to improve the efficiency and safety of iPSC generation. However, current knowledge of the human iPSC reprogramming mechanism is far from adequate. This is mainly due to the low success rate of reprogramming (< 5%) and the lack of specific markers to select cells undergoing full reprogramming.

The entire human iPSC reprogramming process could be roughly divided into three stages: loss of somatic memory, Mesenchymal-to-Epithelial Transition (MET) and gain of pluripotency. Teshigawara et al generated intermediately reprogrammed stem cells (iRSCs) by transducing the Yamanaka factor and the DsRed gene to human fibroblasts. Then, they selected mesenchymal-like cells as candidates, and the iRSCs were validated by global gene expression profile. iRSCs are transcriptionally similar to fibroblasts and iPSCs. More importantly, just culturing iRSCs at a confluent density could turn iRSCs to iPSCs. By using CRISPR-Cas9 technology, one green fluorescent protein (GFP) allele is inserted into the Octamer-binding transcription factor 4 (*OCT4)* genes of iRSC. Using this new cell line, scientists have observed endogenous *OCT4* expressed only before MET transition and conditioned to yamanaka factors silence ectopically in the cell (Teshigawara et al., 2016). This process is also described as the 'stochastic phase', in which OKSM leads to multiple gene activation. This is then followed by a 'deterministic phase' that commences with the activation of *SOX2*. The activation of the *SOX2* locus can occur because of either direct activation or sequential gene activation (Fig1.3) (Buganim et al., 2012).

It is also crucial to understand the function of each reprogramming factor of ESCs and iPS cells to elucidate the mechanism of somatic cell direct reprogramming. Boyer et al first established OCT4's, SOX2's and NANOG's core transcriptional regulatory circuitries in human ESCs (Boyer et al., 2005a). ChIP and DNA microarray data first suggested OCT4, SOX2 and NANOG co-occupied many genes, some of which are active genes playing crucial roles in self-renewal and pluripotency maintenance. Others are repressed genes that are key to developmental processes. Furthermore, the team also suggested that OCT4, SOX2 and NANOG form the core transcription regulatory network via autoregulation and forward feed mechanisms.

In practice, transcription factors activate complex chromatin-remodelling genes to fine-tune chromatin structures. This, in turn, expands the transcriptional regulatory network to maintain the ESC's pluripotency and repressed differentiation capacity (Boyer et al., 2005b). One example from mouse ESCs revealed OCT4 could interact with WD-repeat protein-5 (WDR5) with the help of SOX2 and NANOG. The transcription factors' circuitry cooperates with Trithorax complex (trxG) to add the H3K4 me3 marks to open up the chromatin region, hence allowing the transcription factors to initiate transcription of self-renewal genes, such as stage specific embryonic antigen-1

(*SSEA1)*. The study also demonstrates that WDR5 is needed for efficient iPSC reprogramming, as *WDR5* short hairpin RNA (shRNA) knock down in mouse firbroblasts leads to a reduced number of iPSC colony numbers. However OCT4, SOX2 and NANOG's epigenetic regulation mechanisms in mouse iPSC were not available in this study (Ang et al., 2011).

Another barrier to human iPSC reprogramming is that the expression of Yamanaka factors can lead to oxidative stress and DNA damage; this activates cell senescence. Banito et al demonstrate expression of *OCT4*, *SOX2*, *KLF4* and *C-MYC* in human fibroblasts up-regulated senesence effectors p53, $p16^{INK4a}$ and $p21C^{IP1}$(Banito et al., 2009a) ChIP data suggests there is increased recruitment of histone demethylase Lysine Demethylase 6B (JMJD3) to P16 (*INK4a*) promoter during early reprogramming. This removes the histone repressive markers and makes *INK4a* promoter more accessible to transcription factors. Most importantly, shRNA knock down of senescence effector p53, $p16^{INK4a}$ leads to improved iPSC reprogramming efficiency. This observation has been replicated in multiple cell lines and external studies (Banito et al., 2009b; Hong et al., 2009).

A deep understanding of the human iPSC reprogramming mechanism would ultimately benefit clinical translation. There are a few things that need to be taken into consideration for further mechanistic study. First, OKSM's epigenetic modulation during iPSC reprogramming is a complex and continuous process. Therefore, chromatic structure changes need to be studied at multiple time points. Second, permanent suppression of senescence effectors, such as p53 in iPSC, is dangerous to the cells, as they lose the protection mechanism of switching cancer cells to a senescent state. Future studies should focus on the transient expression of p53 inhibitors or silencers.

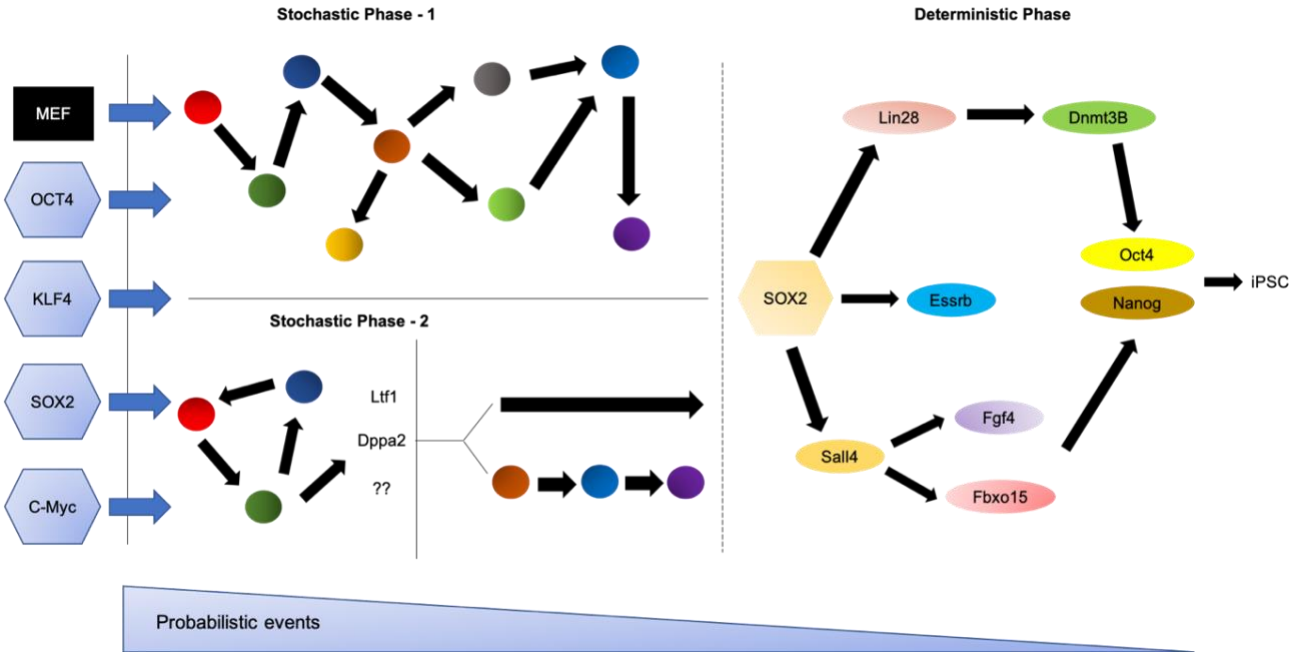



**Figure 1.3 Somatic cell to iPSC reprogramming mechanism** iPSC reprogramming can be divided into two main phases: stochastic and deterministic. In the early stochastic phases, somatic cells undergo MET. In this phase, OSKM initiates a complex gene interaction which could either directly lead to the activation of the *SOX2* locus or to some 'predictive markers', such as undifferentiated embryonic cell transcription factor 1 (*UTF1*), Estrogen Related Receptor Beta (*ESRRB*), Developmental Pluripotency Associated 2 (*DPPA*), Lin-28 Homolog A (*LIN28*), which would be activated first to switch on *SOX2*. Either way, the stochastic process is highly random. During the deterministic phase, SOX2 is responsible for initiating a series of gene activation which eventually switches on pluripotency genes and switches off somatic genes to generate iPSC. Figure adapted form (Buganim et al., 2012).

### 1.1.5 iPSC-based disease modelling

Human iPSC has a key research advantage compared to its ESC counterpart in that it has overcome the ethical concerns of using cells originating from blastocysts. Specifically, in the field of disease modelling, patient-derived iPSCs store all genetic information of an individual patient, bringing hope to personalised medicine. In combination with different differentiation protocols, from stem cells to somatic cells, for the first time, scientists have the potential to generate an unlimited amount of patient-specific disease-relevant cells for mechanistic study, drug screening and in vivo study (Fig 1.4) (Soldner and Jaenisch, 2018). However, until iPSC-derived somatic cells demonstrate key disease phenotypes, the cells are not 'clinically relevant'.

The first exciting breakthrough came as Dr Tamir Rashid whilst in Cambridge successfully used patients' iPSC-derived hepatocyte cells to model Alpha1 Antitrypsin Deficiency (A1ATD). Here, multiple patients' somatic cells were collected and reprogrammed into iPSCs. Then, these cells followed a 21-day differentiation protocol to generate hepatocyte-like cells. Most iHEPs expressed hepatic marker albumin and preserved normal Cytochrome P450 3A4 (CYP3A4) functions. Nevertheless, the author acknowledges that the iHEPs generated were between the end of the first trimester of foetal embryonic development and fully adult cells.

In this study, A1ATD patients had a homozygous Z mutation in the Serpin Family A Member 1 (*SERPINA1*) gene that led to the accumulation of misfolded A1AT in the endoplasmic reticulum (ER) of the human hepatocytes. Immunofluorescence results suggested there were antibodies specific to the A1AT polymer accumulation in the patient's iHEP but not in healthy iHEP (Rashid et al., 2010a). The iHEP-based functional assay has huge potential in large-scale drug screening due to its unlimited cell supply and accurate resemblance to patient disease phenotypes. Indeed, when such a system was integrated with high-throughput imaging assay, a team from Johns Hopkins demonstrated carbamazepine (CBZ) was able to reduce the A1AT polymer load most effectively in multiple patient-derived iHEPs from a large compound library screening (Choi et al., 2013). Although the mechanisms for A1ATD are not fully understood yet, iHEP provides the pharmaceutical industry an opportunity to perform phenotypic drug discovery (PDD) instead of labour-intensive and time-consuming target-based drug discovery (TDD) (Soldner and Jaenisch, 2018).

Moreover, another group from Japan successfully used patient Peripheral Blood Mononuclear Cell (PBMC)-derived iHEP as a functional assay to model PFIC2. Under Matrigel sandwich culturing conditions, the iHEP preserved the mutation in ATP Binding Cassette Subfamily B Member 11 (*ABCB11*) gene which encodes the bile salt export pump (BSEP), which led to aberrant splicing and BSEP protein low expression. Consequently, the BSEP was not expressed on iHEPs' apical membranes. Moreover, the bile acid-excretory rate from the hepatocyte to the canaliculi was much lower in the diseased iHEP compared to the healthy control. On top of that, one of the patient-derived iHEPs was given 4-phenylbutyric acid (4PBA) as a treatment, and the biliary excretion

rate significantly recovered compared to the control. The establishment of the key pathological feature and the initial success of disease phenotype rescued by 4-PBA demonstrate the importance and potential of the iHEP-based cholestatic disease model (Imagawa et al., 2017).

However, Matrigel has been known for its batch-to-batch variability, and this paper did not show polarised canaliculi formed from iHEP. The future iHEP-based cholestatic liver disease model should resemble both functionality and the structural organisation of iHEP to the hepatic lobule as closely as possible (more cholestasis disease models will be discussed in Chapter 2. Prof Takanori Takebe has shown how human iPSC-derived liver organoid (HLO) can be used as a potential assay system in liver toxicity studies. HLO has a clear advantage in differentiation compared to 2D iHEP differentiation. During differentiation, foregut cells form pre-organoids which can be stored in -80 °C freezer. Later, revived organoids can be further matured into HLOs. This protocol is more time-efficient than fresh 2D iPSC differentiation.

Real-time qPCR analysis suggests the gene expression profiles between iPSC-derived HLO and primary hepatocytes are similar. Single-cell RNA sequencing data reveals distinct populations consisting of a parenchymal (74.41%) and a nonparenchymal population (25.59%) in HLOs that express characteristic markers of hepatic stellate cells, portal endothelial cells and cholangiocytes in primary samples. After confirming HLO hepatic functions and cell composition, HLO acquires key canaliculi structures between the iHEPs (hepatocyte polarity will be discussed further in Chapter 2). Another important character of HLO is its active bile transportation. After the addition of BSEP-specific substrate Cholyl-Lysyl-Fluorescein (CLF), Wild Type (WT) HLO shows CLF is transported to the centre ‘canaliculi’, while the BSEP CRISPR-Cas9-KO HLO loses this ability.

The 384 well plate HLO-based high-throughput imaging assay was challenged with several known drugs that cause DILI at their clinical maximum concentration (cMAX) to see how well HLOs could predict toxicity. As a result, HLO achieved 88.7% and 88.9% in sensitivity and specificity, respectively. Conveniently, the DILI score could be calculated based on CLF fluorescence intensity in the ‘canaliculi’ of HLO. Last, patient iPSCs with gene variant cytochrome P450 enzymes 2C9∗2 (CYP2C9∗2)-derived HLO showed specificity to bosentan-induced cholestasis (Shinozawa et al., 2020). However, this study did not show dose-dependent DILI; every drug can

cause toxicity at a high dose. Therefore, it is highly important to consider if the iHEP-based DILI disease model can detect compounds' toxicity within the human physiological context. There are also other iPSC-derived somatic cell-based disease models available. However, most suffer from immature cell differentiation. Therefore, future iPSC-derived disease models should focus more on improving cell hepatic functionality (Hu et al., 2020; Zhang et al., 2011).

Overall, I tried to build an iPSC-derived hepatocyte-like cell disease model for PFIC secondary to *TJP2* mutation in this PhD project. This model satisfies three criteria: first, it is a scalable model, to which compound interests can be applied; second, it is a cheap and rapid assay that detects the desired biological endpoints; third, it has a robust and unambiguous endpoint indicating the state of the biological process of interest (Corbett and Duncan, 2019). The effort will surely be worthwhile because human iPSC-based platforms offer a unique opportunity for patient-specific research that is otherwise unachievable in other disease models.

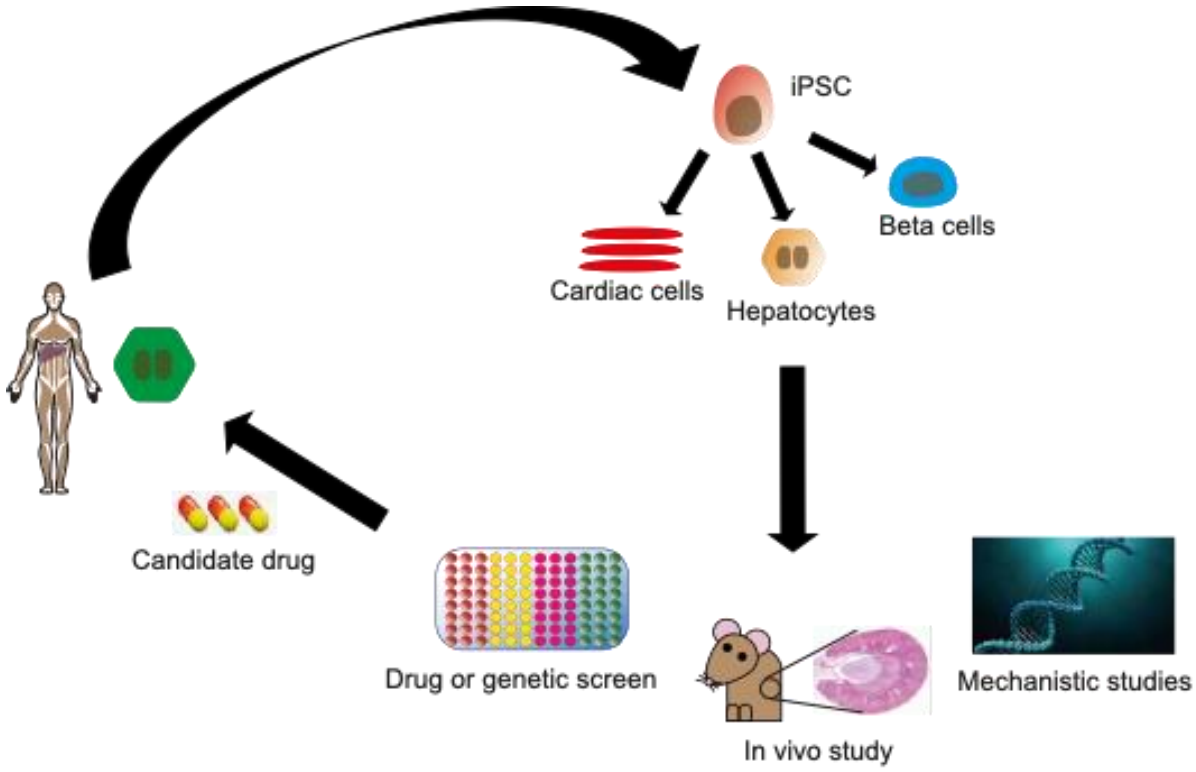



**Figure 1.4 iPSC's potential applications** Schematic overview of using patient iPSCs to model diseases. The patient somatic cells are first collected and reprogrammed into iPSCs. Then, the iPSCs are differentiated into multiple somatic cell lineages. The iPSC-derived somatic cells can be used in the following in vivo disease-mechanistic drug/genetic screen study. Finally, the

positive candidate drugs from the screening system can be used to treat the patient's disease. Figure adapted from mskcc.org.

### 1.1.6 ZINC finger, TALEN and CRISPR-Cas9 genome-editing technology

Ever since the discovery of the DNA double helix structure, scientists and clinicians have been dreaming of rewriting the genome to influence gene expression, as well as the functionality of proteins. The early observations of natural DNA repair mechanisms in bacteria and Homologous Directed Repair (HDR) in mammalian cells revealed that cells used endogenous methods to repair double-strand breaks (DSBs), which were otherwise lethal to the cell if not resolved. Pioneer studies involved packaging oligonucleotides with a DNA cleavage reagent, such as bleomycin, or using self-splicing intron to change DNA sequences. Unsurprisingly, such methods were highly inefficient and stochastic (Doudna and Charpentier, 2014). Later, zinc fingers' (ZNF) protein-mediated DNA binding was shown to be effective when coupled with restriction endonuclease FoKI restriction enzyme to cut Lambda DNA in a sequence-specific manner (Kim et al., 1996). However, ZNF protein generation was time-consuming, and the targeting was often context-dependent (Table1.1).

In early studies, individual ZNF used three 'fingers' to bind nine base-pair targets, and there were two such ZNFs on either side of the cleavage domain. Recent studies have added more fingers (up to six per ZNF) to bind larger cleavage targets (Urnov et al., 2010). Yusa et al. used ZNF and piggyBac technology to engineer a bi-allelic correction of a point mutation (Glu342Lys) in the α1-antitrypsin gene without a genomic scar. This was the proof of concept study for ZNF, used as an engineering tool for gene therapy to treat α1-antitrypsin deficiency (A1ATD) (Yusa et al., 2011).

Transcription activator-like (TAL) effector nuclease (TALEN) technology has improved gene-editing efficiency, as TALEN's DNA binding sites, in theory, have no restrictions, and they can be rapidly designed according to 'protein-DNA code'. TALEN is made up of the N-terminal domain, TAL effector repeat domain, C-terminal domain and FoKI nuclease domain. TALE repeat domain is composed of proteins with highly conserved amino acid repeat sequences; it can also be customised to bind different DNA sequences (Table 1.1) (Joung and Sander, 2013). Prof Chad Cowen's team has shown TALEN can be used to generate disease-specific mutations in human ES and iPSCs, with very few off-target effects, upon differentiation into various somatic cell types.

Key disease phenotypes, such as dyslipidemia, insulin resistance and motor-neuron death have been captured. This study showcases how TALEN could be used to generate disease models (Ding et al., 2013).

The discovery of CRISPR-Cas9 brought the gene-editing field into a 'golden era'. Realising CRISPR-Cas9's full potential will likely have huge impacts in areas such as human gene therapy, agriculture and disease modelling. The idea of CRISPR-Cas9 was first introduced by a group of Japanese researchers, who reported a series of short direct repeats interspaced with short sequences in the genome of Escherichia Coli. Later, such sequences were seen across many bacteria. Around the same time, key evidence suggested this CRISPR sequence had viral origins. Furthermore, this CRISPR-associated sequence (Cas) encoded proteins with putative nuclei and helicase domains. Together with CRISPR sequences in bacteria, scientists proposed CRISPR-Cas9 as bacteria's adaptive immune response for fighting off invading viruses. Later, CRISPR RNAs (crRNAs) transcribed from the CRISPR sequence were demonstrated to guide the CRISPR-Cas9 complex to the complementary sequence in the viral genome to allow the Cas9 nuclease enzyme to destroy the viral DNA sequences in the bacteria.

Detailed CRISPR-Cas9-mediated bacteria antiviral adaptive immunity has three steps: first, viral DNA is inserted into the CRISPR array as a spacer sequence in the bacteria; second, the crRNA precursor is transcribed and further processed to mature crRNAs, and each individual crRNA contains repeat sequences and viral invader-targeting spacers; third, crRNA guides the CRISPR-Cas9 complex to bind to complementary viral DNA sequences to cleave the viral nucleic acid (Table1.1) (Doudna and Charpentier, 2014).

Over the years, CRISPR-Cas9's functionality has been illustrated step by step. Arginine motifs in the C-terminal domain of Cas9 interact with the protospacer adjacent motif (PAM) DNA sequence, which usually appears as NGG (N could be any nucleic acid). Following an initial DNA helix unwinding event, R-loop forms and expands towards the distal protospacer, and an RNA-DNA hybrid forms at the end. The sgRNA has a sequence that is 20 nucleotides long, an sgRNA-DNA binding following the Watson-Crick base pairing. Mismatches are sometimes tolerable within

three base pairs (bp) starting from the distal protospacer. Evidence suggests a mismatch is not tolerated from the PAM site.

The Cas9 protein has two domains, RuvC and HNH. Each domain is responsible for nicking a single strand. Therefore, CRISPR-Cas9 results in DSBs. Mutations in either domain allow Cas9 to transform into a single strand nickase, thus providing more flexibility for base editing (Doudna and Charpentier, 2014). There are two general natural repair mechanisms – HDR and Nonhomologous End Joining (NHEJ). For gene KOs, large deletion, including critical exons, to generate sequence frame shift is favoured. This requires a pair of CRISPR-Cas9 complexes at each end of the deleting sequences to generate two DSBs. The DSBs are then repaired through NHEJ, which is error-prone. In contrast, precise base editing relies on HDR to repair the CRISPR-Cas9-caused DNA damage. Often, short single strand Oligo Nucleotides (ssODN) contain homology arms, and the desired replacement sequence is delivered together with the CRISPR-Cas9 Ribonucleoprotein complex (RNP). Therefore, when the DSBs occur close enough to the engineered position, the cell can use the ssODN as a template to repair the DSB, and the new sequence can be integrated into the host cell genome (Hendriks et al., 2016d).

One of the ultimate goals of stem cell research is to use patient-specific pluripotent stem cell-derived organs for acute organ failure. Sheep are ideal hosts to grow human organs through blastocyst complementation. Prof Hiromitsu Nakauchi's team used CRISPR-Cas9 technology-generated bi-allelic Pancreatic and Duodenal Homeobox 1 (*PDX1*) gene-mutated oocytes. These embryos were then transferred to a recipient ewe. A four-month-old foetus was collected and subjected to histological immunofluorescence analysis, and evidence suggested Islets of Langerhans were missing, as well as PDX1 and insulin protein. This study suggests that CRISPR-Cas9 technology can be used to 'empty' the sheep's organs by deleting essential organogenesis genes and opening the therapeutic possibility of 'replacing' patient-specific organs in sheep.

A more recent state-of-the-art CRISPR-Cas9 genome editing tool is the 'prime editor', developed by Dr Davide R Liu's team. It allows efficient targeted insertions, deletions, few by-products and all 12 types of point mutations without requiring DSBs or donor DNA templates from human cells. The prime editor is made of Cas9 nickase, prime editing guide RNA (pegRNA) and reverse

transcriptase (RT) domain. The biggest advantage of the prime editor compared to HDR-mediated base editing is that prime editing does not require DSBs. PegRNA's three prime ends have primer binding sites used to help the complex find a target, and pegRNA's RT template helps to reverse transcribe edited RNA to DNA at the target site (the detailed mechanism was included in the manuscript) (Anzalone et al., 2019).

Future CRISPR-Cas9 research needs to focus on binding site availability, as currently, the complex binding is restricted to PAM sequences in the genome. Although PAM sites appear every eight base pairs on average in the human genome, certain nucleotide substitutions are still difficult to operate (Cong et al., 2013). Moreover, a mismatch between guide RNA (gRNA) and targeted DNA sequences could occur. This requires more careful gRNA design as well as off-target checks post-CRISPR-Cas9 edit.

**Table 1.1 Brief summary of major genetic engineering tools**

| Genetic engineering tools | Mechanism of action | Advantage | Disadvantage |
|---|---|---|---|
| ZNF | ZNF domains have relatively small protein motifs that contain multiple finger-like protrusions that make contacts with DNA of interest; then, FOKI domain makes the cut | Highly specific DNA target due to the dimer structure | Time-consuming and context-dependent |
| TALEN | TALEN's DNA binding site is designed to match the DNA of interest in the genome, and its FOKI domain makes the cut | Able to target anywhere in the genome in theory | Time-consuming to engineer TALEN protein |
| CRISPR-Cas9 | gRNA guide the Cas9 enzyme + tracer complex to target DNA and make the cut | Easy to manufacture, highly efficient and allows multiple genetic mutations to be made | Only targets PAM-avaliable sites |

### 1.1.7 Aims and objectives for this chapter

The overall goal for this chapter was to create relevant human iPSC disease models to study the *TJP2* and *ABCB4* mutations' effects on hepatic bile flow. Therefore, in collaboration with DefiniGEN, we have generated one iPSC line with *TJP2* mutations, one iPSC line with *ABCB4* mutations by using CRISPR-Cas9 and reprogrammed one patient fibroblast into iPSC with *TJP2* mutation. More specifically, post-transfected iPSCs need to be sorted into single cells by Fluorescence-Activated Cell Sorting (FACS), and the resulting colonies will be subjected to sanger sequencing to confirm the KO genotype. Patient fibroblasts were transfected with nonintegrating vectors containing five reprogramming factors. Successfully reprogrammed iPSCs were handpicked and sanger sequenced. The upcoming sections illustrate the following:

1. How the problem of FACS single-cell surviving iPSC was solved and how the homozygous gene KO protocol was established.
2. How DefiniGEN reprogrammed patient iPSCs.
3. How the genotypes for both of the established disease iPSC lines were confirmed.

## 1.2 Methods

### 1.2.1 Designing gRNA to KO *TJP2* and *ABCB4* in healthy human iPSCs by using CRISPR-Cas9 technology

A. CRISPR gRNA design and processing

The project's first aim was to generate a homozygous *TJP2* KO iPSC clone. The first step was to find the suitable gRNA targeting regions in *TJP2*. Two gRNAs were designed to flank each side of *TJP2* exon5. The gRNAs make DSBs that contain the deleted region's 749 base pairs according to a bioinformatics prediction. The aim is to generate a new premature stop codon and a frameshift after the indels are repaired by the cell's intrinsic mechanism. Then, the cell will transcribe some of the new *TJP2* DNA. However, this will trigger the non-sense-mediated mRNA degradation (NMD). Therefore, the mRNA theoretically does not get translated. As previously demonstrated, patient ZO2-mutated liver tissue immunostaining does not pick up any ZO2 protein staining. Hence, the goal is to make an iPSC-based disease model that matches the endpoint phenotypes of the patient's hepatocytes.

B. Pre-CRISPR bioinformatics

gRNAs have been chosen based on bioinformatics analysis of *TJP2* and human *ABCB4* in ENSEMBLE UK. Protein coding transcripts and the Havana sequences are the priority transcripts to focus on. The *TJP2* exon5 has been chosen, as well as *ABCB4* exon4, as they are both the first long-coding exons shared by four Havana transcripts. After deleting *TJP2* exon5, and *ABCB4* exon4, a new reading frame will be generated. Based on the chosen targeting exons, appropriate gRNAs with minimal predicated off-target effects can be chosen from the sanger CRISPR website.

## 1.2.2 RNP complex transfection and iPSC cell culturing

First, crRNA and trans-activating crRNA (tracrRNA) are re-suspended and mixed in the final duplex concentration of 100 μM. Second, the mixture is heated at 95°C for five min then allowed to cool at room temperature. The crRNA:tracrRNA duplex and Cas9 nuclease NLS are mixed, then incubated at room temperature for 10 to 20 min to allow the RNP to form. During the 20-min incubation time, WT iPSCs (which had already been incubated with Rho-associated protein kinase inhibitor (ROCKi) for one hour) were dissociated. iPS cells were lifted from incubation with dissociating enzyme TrypLE for five min. Five ml were added to the iPS cell medium then added to the cells, which were pipetted up and down repeatedly to achieve monodisperse cell suspension. The iPS cells were briefly centrifuged at room temperature to form a pellet of cells. 3.2 million WT iPSCs were prepared for three independent electroporation experiments. iPS cells were re-suspended in 295 μl Endocochlear Potential (EP) buffer and mixed with 25μl RNP. 100 μL of the mixture was dispensed into three cuvettes separately. Then, the electroporation condition was set according to Table 1.2. After transferring the RNP and iPSC mixture into the cuvette, the NEPA21 impedance value should be in the range of 30–55 Ω to carry out the electroporation. Once electroporation is finished, a sufficient amount (e.g. 300 μl) of Essential 8 media (E8) with 10 μM ROCKi was added to the cuvette. All cells and mediums were then transferred from the cuvette to the vitronectin-coated plates. Electroporated cells were cultured in a $CO_2$ incubator overnight. The following day, transfected iPSCs were cultured in E8 and required daily medium change.

**Table 1.2 electroporation conditions for iPSCs**

| | Poring Pulse | | | | | | Transfer Pulse | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | V | Length (ms) | Interval (ms) | No. | D. Rate (%) | Polarity | V | Length (ms) | Interval (ms) | No. | D. Rate (%) | Polarity |
| *1* | **125** | **5** | 50 | 2 | 10 | + | 20 | 50 | 50 | 5 | 40 | +/- |

## 1.2.3 FACS single-cell post-transfected iPSCs and PCR-based screening

A. Post-transfection maintaining and screening methods

The post-transfected iPSCs mainly consisted of transfected cells and nontransfected clones. Within the transfected cells, there were successful and unsuccessful homozygous *TJP2* KO clones. Single-cell culturing was required to generate successful KO clones. There were two conventional ways to generate single-cell clone-serial dilution and FACS. Practically, it was very difficult to perform serial dilution due to the multicolony formation when picking the colonies from the agar plate. However, the FACS machine could accurately sort a single cell into each well.

B. FACS Standard operating procedure (SOP)

The sorting cell culture plate was prepared the day before the FACS experiment began. The coating solution was replaced with fresh E8 + Revita + gentamycin + Penicillin Streptomycin. On the day of FACS, transfected iPSCs were incubated with 10 uM ROCKi one hour before FACS. During this incubation, we started to remove the coating solution in the 96 well plates and replaced it with E8 + Revita + Penicillin according to the manufacturer's recommendation. After one hour of incubation with ROCKi, the cells were washed three times with Phosphate-Buffered Saline (PBS) and treated with TrypLE to lift. After counting the cells, the collection tube was centrifuged, and the supernatant was removed. The cell pellet was re-suspended with 1 ml FACS buffer (PBS + 3%

BSA + 0.1 mMEDTA). 4′,6-diamidino-2-phenylindole (DAPI) (10 μM) was added to 900 μl FACS buffer iPSC suspension, and the remaining 100 μl was left unstained. The Aria fusion FACS machine sorted the pool of cells into single cells based on the Forward Scatter (FSC) and the Side Scatter (SSC) value of the cell. The 96 well plates were centrifuged after single-cell sorting at 100 g for 3 min. Post-FACS, the 96 well plates were cultured in a normal oxygen incubator overnight.

C. Troubleshooting FACS experiments

The first technical difficulty was that single-cell iPSCs rarely survive in the 96 well plates post-FACS, resulting in insufficient single-cell colonies to screen. I designed a few experiments to improve the single-cell recovery rate from the 96 well plates.

It is known that ROCKi could help ES single-cell colony survival. A Japanese group demonstrated that single-cells treated with ROCKi improved colony formation from 1% to 27%. In light of this finding, multiple ROCKi concentrations were optimised ranging from 1/1,000, 1/500 and 1/50 for iPSC culture post-FACS. The results demonstrated that 10 μM (1/500) culturing for five successive days gave the best performance; 19 wells out of 96 wells survived from this plate (Table1.3). This was a 20-fold increase in survival compared to the control plate. In addition, Revita has been used to improve single-cell iPSC survival. Therefore, I also designed experiments to test how Revita, combined with various coating conditions, would affect single-cell iPSCs post-FACS. Surprisingly, there was a 40% single-cell iPSC survival rate post-FACS when iPSCs were cultured with Revita supplements (Table1.3). Therefore, Revita supplements set the foundation for all the following single-cell iPSC FACS.

**Table 1.3 Post-FACS iPSC culture optimisation results**

| Conditions | Surviving single cells (Total 95 wells) |
|---|---|
| Revita + Mat | 38/95 |
| Revita + Vit | 28/95 |
| Ri (1/1,000) + Vit | 15/95 |
| Ri (1/1500) + Vit | 10/95 |
| Ri (1/500) + Vit | 19/95 |
| Vit control | 1/95 |
| Mat control | 1/95 |

D. DNA extraction from surviving colonies and genotyping

The growth of the cell colony could be observed by Leica phase-contrast microscope after nine days of culturing with E8 and Revita. Cells were detached from the plate using a gentle cell dissociation reagent. The dissociation reagent was removed, flushed three times with E8, then transferred directly to 24 well plates. When cells reached confluence, the same process was repeated to pass the cells onto 12 well plates. Then, once the cells reached confluence in the 12 well plates, 1/3 of the cells from each well were used for genotyping, and the other 2/3 of the cells frozen for cryopreservation. Cryostor was used to cryopreserve the cells in liquid nitrogen. In order to find a single-cell colony's genotype, 50 µl of quick extract solution (Epicentre) was added. The mixture was incubated for 6 min at 65°C, then vortexed for 15 secs. Finally, the mixture was incubated for 2 min at 98°C to generate ready-to-use single-cell colony DNA. The screening Polymerase Chain Reaction (PCR) primers were designed to flank both sides of ZO2 exon5 CRISPR KO region, which gave the PCR a product of 1.5 kb. The 25 µl PCR mixture contained 10 µM forward (FWD) and 0.5 µl reversed (REV) primer, 2 µl of template DNA, 12.5 µl of (Taq 2X Master Mix NEB) and 9.5 µl of Deuterium Depleted Water (DDW). The PCR machine was set according to the condition in (Table 1.4).

**Table 1.4 Single-cell colony genotyping PCR conditions**

| | | |
|---|---|---|
| Initial denaturation | 95°C | 30 seconds |
| 40 cycles | Denaturation 95°C<br>Annealing 54°C<br>Elongation 68°C | 30 seconds<br>30 seconds<br>90 seconds |
| Final extension | 68°C | 5 min |
| Hold | 10°C | |

The PCR products were stained with 10x gel loading dye and loaded onto a 2% agarose gel, imaged with a BIO-RAD DNA gel-viewing machine. The KO band could be cut using a scalpel under UV light conditions manually. The gel fragments were suspended by buffer QG and incubated at 50°C for 10 min. Then, 200 µl isopropanol was added to the sample, mixed and placed in a QIAquick spin column in a provided 2 ml collection tube, into a vacuum for 1 min or applied to the manifold of the vacuum until all the samples had passed through the column. The column was then washed three times and QIAquick column was placed into a clean 1.5 ml microcentrifuge tube. Finally, to elute the DNA, 20 µl buffer EB was added to the centre of the QIAquick spinning column membrane and centrifuged the column for 1 min. DNA concentration was measured by NanoDrop machine (Thermofisher Scientific).

E. Trioctylphosphine oxide (TOPO) cloning + colony picking

One of the aims of this project was to generate iPSCs that mimic patients' mutations as closely as possible. The goal was to generate a ZO2 homozygous KO in the WT iPSCs. Once the PCR results had confirmed that there was a double allele KO, the KO band was cut, and the DNA from the band gel was isolated. The product was then used in a TOPO cloning experiment to determine the individual allele genotype. First, the agar solution was heated up by microwave until completely dissolved. When the agar solution cooled to a normal temperature, ampicillin (1/1,000) was added to the agar solution. This agar solution was used to coat a bacteria-growing dish. It was stored at room temperature for one hour and then placed in a 37˚C incubator to remove any moisture. The transformation mixture was prepared as 1 µl of salt solution with a final concentration of 200 mM

NaCl, 10 mM MgCl2, 4 µl of the PCR products from the KO clone and 1 µl of the TOPO vector. This mixture was left at room temperature for 5 min to allow the TOPO vector to incorporate the DNA fragment. Then, 2 µl of the mixture were added to 10 competent TOPO cells and incubated 30 min on ice. After, the incubation started to heat shock the cells for 30 seconds at 42˚C; 250 µl Super Optimal broth with Catabolite repression (S.O.C) media was required. Then, the tube was vibrated at a speed of 200 rpm, 27˚C for one hour on an orbital shaker. Finally, 50 µl from this mixture were spread on a pre-warmed agar plate with a glue spreader. This plate was incubated at 37˚C in an incubator overnight. The next day, eight or nine suitably sized colonies were picked from the agar plate and transferred into luria-bertani (LB) broth bacteria growing media with an ampicillin concentration (1/1,000). Finally, this tube was placed on an orbital shaker with 180 rpm at 37˚C overnight.

F. Mini prep

A successful TOPO cloning solution had a cloudy colour. Those tubes were then selected for mini prep experiments to extract the amplified TOPO vector's DNA from the cells. First, the cloudy TOPO-cloned solution was transferred to a new Eppendorf. Then, the tubes were centrifuged at 6,800 rcf for 3 min, and the supernatant was removed. After that, the pellet was re-suspended with 250 µl of buffer P1 and transferred to a microcentrifuge tube. Next, 250 µl of buffer P2 were added and mixed with the solution. N3 solution was added to the tube and centrifuged for 10 min at 13,000 rpm per min. The supernatant was removed from the tube and transferred into a QIAprep spin column. The spin column was further centrifuged for 10 min to allow the DNA to bind to the membrane. The column was washed three times with PE buffer, and EB buffer was used to elute the DNA. The DNA was then sent to Source Bioscience (a sequencing company) for sanger sequencing.

### 1.2.4 Reprogramming patients' PFIC fibroblasts with *TJP2* mutation to iPSCs

Human specimens in this study were obtained under the study protocol, conforming to the ethical guidelines of the 1975 Declaration of Helsinki, and were approved by the Institutional Review Board of King's College Hospital. About 8-mm skin punch biopsies or peripheral blood cells were collected from volunteer subjects attending King's College Hospital. Fibroblasts were derived from the donor tissue in cGMP conditions using standardised in-house protocols and expanded in

a standard fibroblast culture medium. Nonintegrating origin of viral replication/Epstein–Barr virus nuclear antigen 1 (*OriP/EBNA1*) vectors from Invitrogen containing five reprogramming factors (*OCT4*, *SOX2*, *LIN28*, *KLF4* and *L-MYC*) were transfected into the patient fibroblasts as previously described (Drozd et al., 2015). The entire reprogramming process was free of feeder cells. Successful human iPSC clones were cultured in E8™ specially formulated for the growth and expansion of human pluripotent stem cells.

### 1.2.5 qPCR pluripotency gene mRNA quantification

Total RNA was extracted from cells by the RNeasy kit (Qiagen) following the manufacturer's instructions. After measuring total RNA concentration, 500 ng of RNA were subjected to reverse transcription reactions. The real-time PCR by the TaqMan probe system (gene expression master mix) and the QuantStudio system (ThermoFisher) quantified mRNA of target genes with specific primers and quantification protocol. After normalised with a housekeeping gene (Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), each gene expression level was described relative to iPSC control HDF.

## 1.3 Results

### 1.3.1 Improved post-FACS single-cell iPSC colony survival

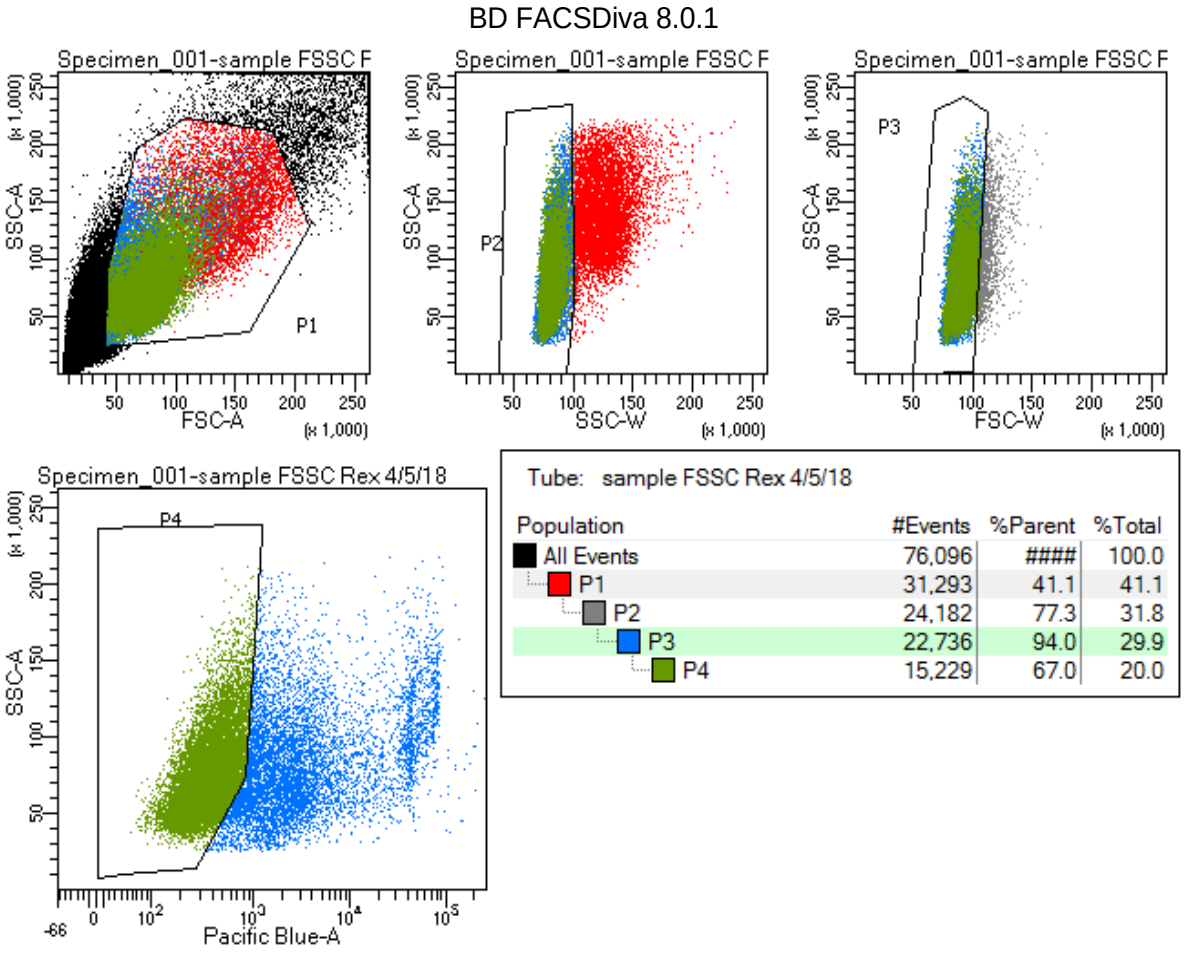
The initial low survival rate of single-cell iPSC colonies post-FACS became a bottleneck for the entire project because it was not possible to make the cholestasis disease model without *TJP2*-mutated iPSCs. To solve this rate-limiting problem, the issue was hypothesised to occur in the following three parts: pre-FACS, during FACS and post-FACS. Thus, I decided to investigate the root cause from each section independently. First, the post-transfected iPSCs were lifted before FACS and counted by the cell counter. The 91.8% viability of iPSCs suggested by the cell counter would be the evidence that cells pre-FACS were alive (data not shown). Second, the iPSCs were

selected based on two conditions: alive and singular. Single cells could be gated out by SSC and FSC parameters. Live cells, however, had to be DAPI negative, as it stained for dead cells. Therefore, after four successive tight gatings, in theory, all cells sorted into the 96 well plates should have been singular and alive (Fig1.5a).

Logically, the root cause of the problem could only be presented during the post-iPSC sorting period. In this section, there were two possibilities causing low single-cell iPSC survival rates. One was that the nozzle of the FACS machine was not suitable for the iPSCs. The other was that post-FACS culturing conditions lead to cell death. In light of this logic, I tried different nozzle sizes for the same FACS machine; there was no clear difference in iPSC survival rate between test and control groups (data not shown). Finally, I troubleshot the FACS iPSC experiment in this section by altering different culturing conditions, as described in the method section. E8 supplemented with Revita or ROCKi led to a significant improvement in iPSC survival and the ultimate success of *TJP2* KO iPSC generation.

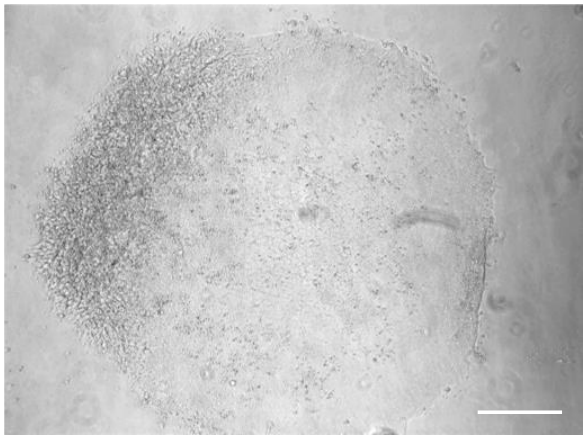
Furthermore, I wished to determine if Revita or ROCKi could impact iPSCs' morphologies and passage abilities. The brightfield microscopy images suggested the post-FACS single-cell iPSCs did not show any morphological differences between the three culturing conditions, and iPSCs from all conditions were able to pass to the new wells (Fig1.5b). The single-cell iPSC FACS SOP established here introduces a novel method to efficiently generate viable and functional single-cell iPSCs, which otherwise would die due to the lack of cell–cell contact.

(A)

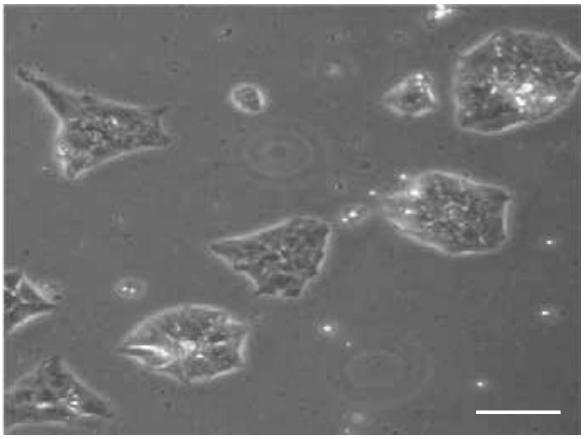

BD FACSDiva 8.0.1
Specimen_001-sample FSSC F
SSC-A
FSC-A
P1
Specimen_001-sample FSSC F
SSC-A
SSC-W
P2
Specimen_001-sample FSSC F
SSC-A
FSC-W
P3
Specimen_001-sample FSSC Rex 4/5/18
SSC-A
Pacific Blue-A
P4
Tube: sample FSSC Rex 4/5/18
Population #Events %Parent %Total
All Events 76,096 #### 100.0
P1 31,293 41.1 41.1
P2 24,182 77.3 31.8
P3 22,736 94.0 29.9
P4 15,229 67.0 20.0

(B)

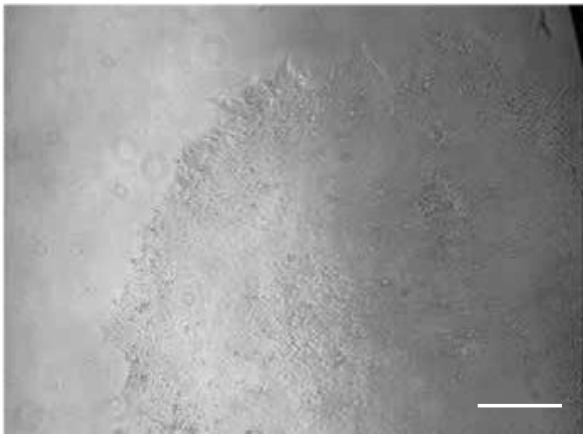
(1) control iPSC



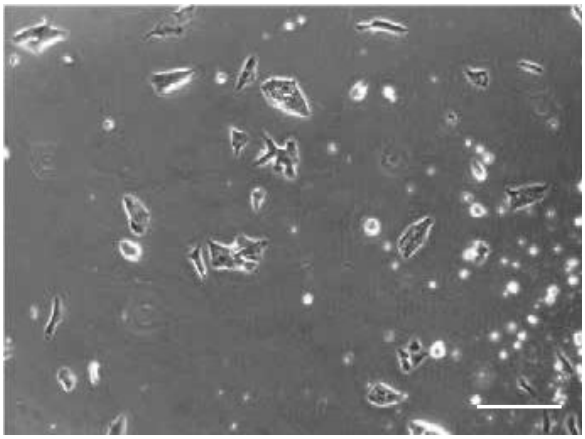
(2) control iPSC passage



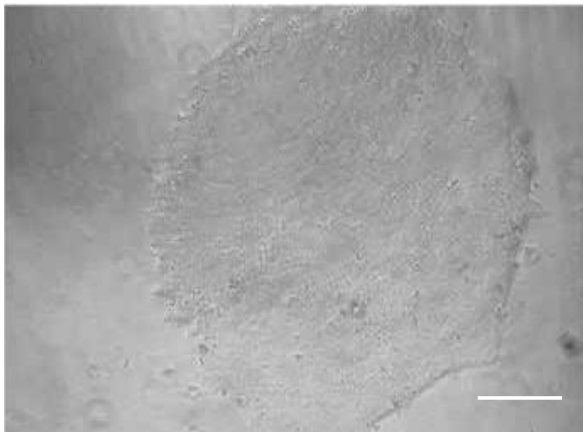
(3) Revita iPSC



(4) Revita iPSC passage



(5) ROCKi iPSC



(6) ROCKi iPSC passage

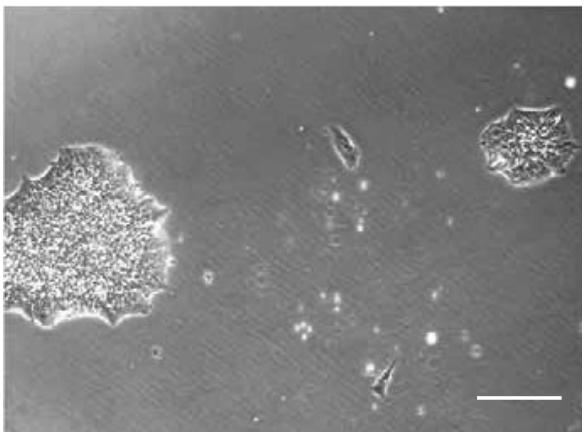


**Figure 1.5 Troubleshooting single-cell iPSC generation** (A) FACS results from four independent gatings. DAPI (Pacific blue) was a marker for dead cells, SSC and FSC were used as the parameters for single-cell selection. Population 4 (P4) was selected as the live and single-cell population used to sort into the plate. (B) Brightfield images of iPSCs in three independent culturing conditions – Control (E8 only), Revita, ROCKi (1/500). Images B1, B3 and B5 show single-cell iPSC colonies of all three conditions appearing round and intact. Images B2, B4 and B6 show post-gentle cell dissociation reagent treatment, the big single iPSC clone could be broken into various smaller colonies. All new iPSC colonies had the round shape and halo ring around the cells. Scale bar, 100 µm.

### 1.3.2 *TJP2* homozygous KO iPSC generation

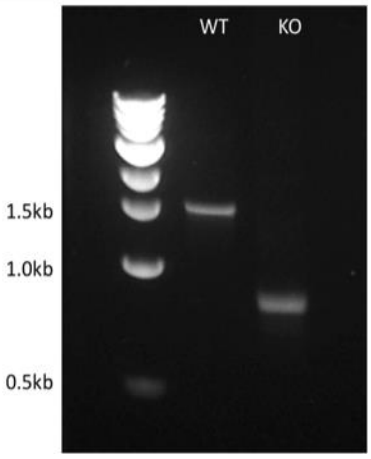
PCR was used to screen for homozygous KO single-cell colonies post-FACS. FWD and REV primers were designed 400 base pairs away from the gRNAs and WT iPSC DNA, which resulted in a PCR product of 1.5 kb (Fig1.6a). By using the same primers to screen for the single-cell derived exon5 KO colony by PCR, one colony showed a clear single band just below 1 kb (Fig1.6b). This potential homozygous KO colony's PCR product was further analysed using a TOPO cloning experiment to find out the individual allele differences of this KO clone. Eight colonies had been picked randomly from the agar plate, which, in theory, represented eight random DNA fragments from the pool of amplified DNA in the KO clone's PCR products.

Allele 1 and 2's TOPO cloning results demonstrated a balanced ratio (4:4) (Fig1.6c). Hence, there was an equal number of allele 1 and allele 2 DNA fragments in the KO clone's amplified PCR products. Sanger sequencing results suggested both alleles had been KO by gRNA2 and gRNA4's RNP complex. Allele 1 had 688 bp deletion, and allele 2 had 691 bp deletion (Fig1.6d). Overall, I generated a *TJP2* exon5 homozygous KO clone using CRISPR-Cas9 technology.

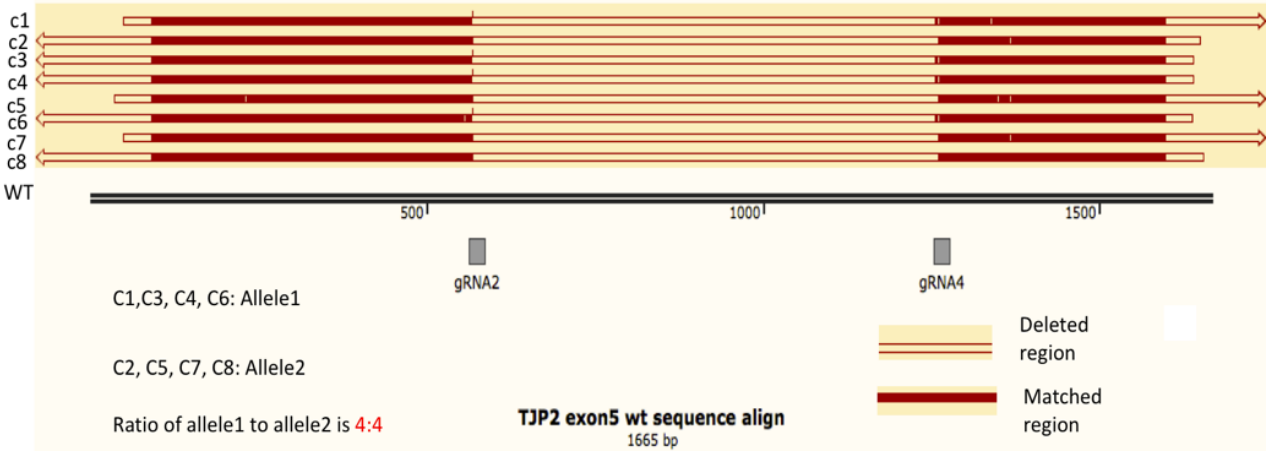
**(A)**


FWD screening primer
TJP2
69,220,200
69,220,250
69,220,300
69,220,350
TJP2
69,220,400
69,220,450
69,220,500
69,220,550
gRNA1
TJP2
69,220,600
69,220,650
69,220,700
The deleted region in SCC
gRNA2
TJP2-202 Exon 5
TJP2
69,220,800
69,220,850
69,220,900
The deleted region in SCC
TJP2-202 Exon 5
TJP2
69,220,950
69,221,000
69,221,050
69,221,100
The deleted region in SCC
TJP2-202 Exon 5
TJP2
69,221,150
69,221,200
69,221,250
gRNA3
gRNA4
The deleted region in SCC
TJP2-202 Exon 5
TJP2
69,221,350
69,221,400
69,221,450
TJP2
69,221,550
69,221,600
69,221,650
REV screening primer
TJP2
69,221,700
69,221,750
69,221,800
69,221,850

(B)


WT
KO
1.5kb
1.0kb
0.5kb

(C)


c1
c2
c3
c4
c5
c6
c7
c8
WT
500
1000
1500
gRNA2
gRNA4
C1,C3, C4, C6: Allele1
C2, C5, C7, C8: Allele2
Ratio of allele1 to allele2 is 4:4
TJP2 exon5 wt sequence align
1665 bp
Deleted region
Matched region

(D)

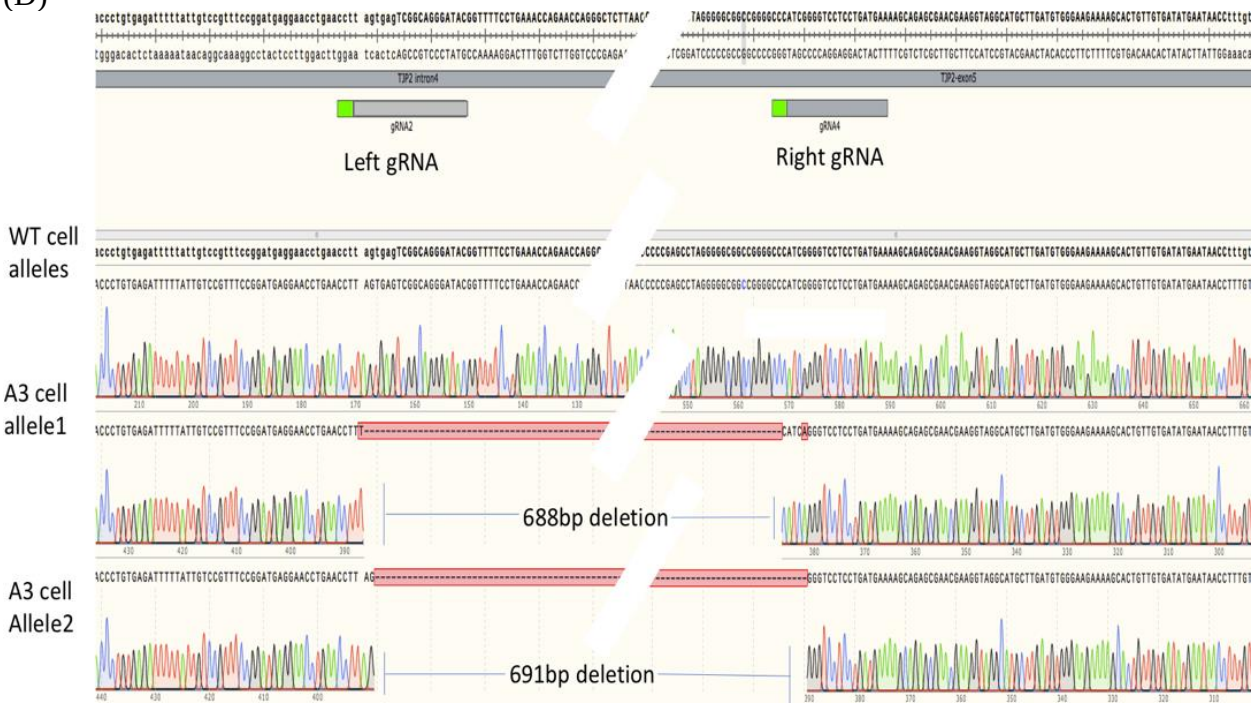



**Figure 1.6 Making *TJP2* KO iPSCs** (A) Schematic illustration of CRISPR *TJP2* KO gRNAs and screening primer bioinformatic design. gRNA 1 and 2 are located in intron 4 of the *TJP2* gene, and gRNA 3 and 4 are located at the end of exon 5 of the *TJP2* gene. (B) Diagrams illustrating iPSC clones' PCR screening results. WT iPSCs result in a 1.5 kb band; KO clones result in a 1.0 kb band. The single band in KO iPSCs indicates the clone is a homozygous KO. (C) Diagrams illustrating C1 to C8 colonies' TOPO cloning results. The mini prep products of C1 to C8 are sent for sanger sequencing, and results are aligned with the WT *TJP2* exon5 sequence. Among those clones, half are allele 1, and the other half are allele 2. (D) Diagrams illustrating sanger sequencing results of iPSC *TJP2*-KO clone genomic DNA. KO clone A3 cell allele 1 has 688 bp deletion, and allele 2 has 691 bp deletion.

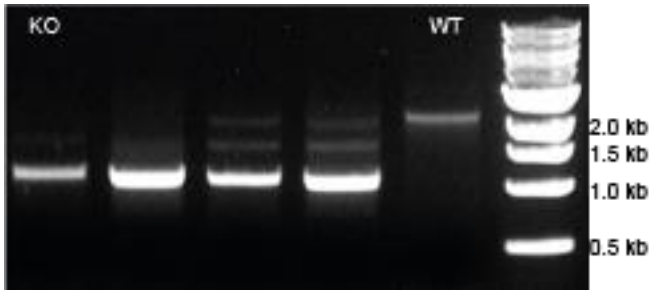
### 1.3.3 *ABCB4* homozygous KO iPSC generation

In addition, I also generated a homozygous human *ABCB4* KO iPSC by using the same approach. FWD and REV primers were designed 280 base pairs away from the gRNAs, and WT iPSC DNA resulted in a PCR product of 2 kb (Fig1.7a). By using the same primers to screen for the single cell-derived exon 4 KO colony by PCR, one colony gave a clear single band at 1 kb (Fig1.7b). This potential homozygous KO colony's PCR product was further analysed by performing a TOPO cloning experiment to find the individual allele differences of this KO clone. Eight colonies were picked randomly from the agar plate, which, in theory, represented eight DNA fragments from the pool of amplified DNA in the KO clone's PCR products. Sanger sequencing results suggested that both alleles had been KO by gRNA1 and gRNA4's RNP complex. Allele 1 had 1090 bp deletion, and allele 2 had 1083 bp deletion (Fig1.7c). Overall, I generated an *ABCB4* exon4 homozygous KO clone using CRISPR-Cas9 technology.

(A)


FWD screening primer
ABCB4
16,450
16,500
16,550
16,600
16,650
16,700
16,750
ABCB4
16,800
16,850
16,900
16,950
17,000
17,050
17,100
PmeI
The deleted region in SCC
gRNA1
gRNA2
ABCB4
17,150
17,200
17,250
17,300
17,350
17,400
17,450
The deleted region in SCC
ABCB4-002 Exon 4
ABCB4
17,500
17,550
17,600
17,650
17,700
17,750
17,800
The deleted region in SCC
gR
ABCB4
17,900
17,950
18,000
18,050
18,100
18,150
18,200
The delet...n in SCC
NA3
gRNA4
ABCB4
18,250
18,300
18,350
18,400
18,450
18,500
18,550
ABCB4
18,600
18,650
18,700
18,750
18,800
18,850
18,900
REV screening primer
ABCB4

(B)




(C)

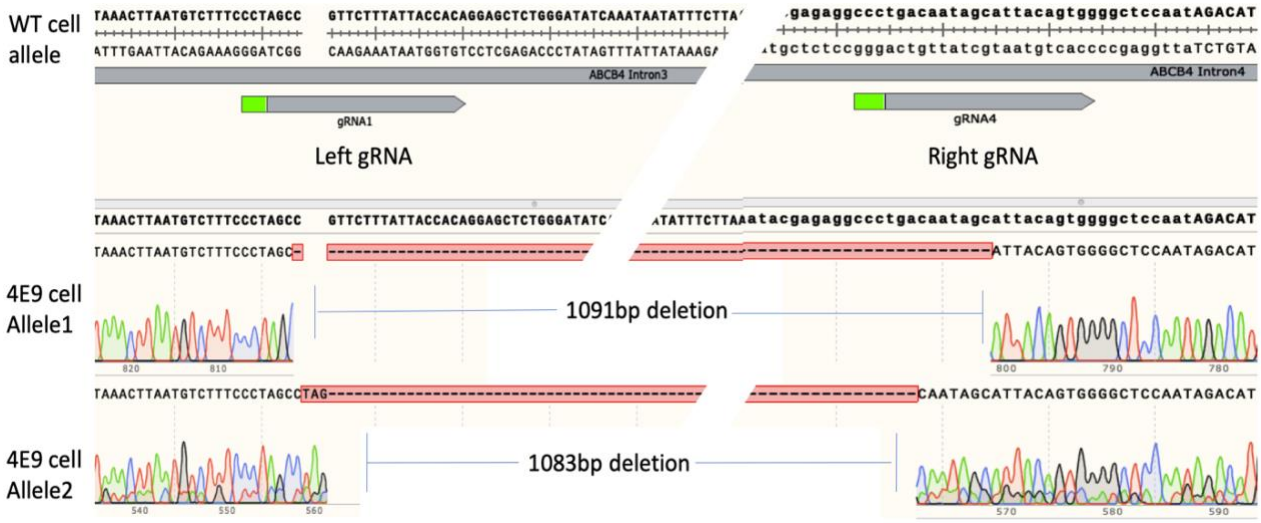
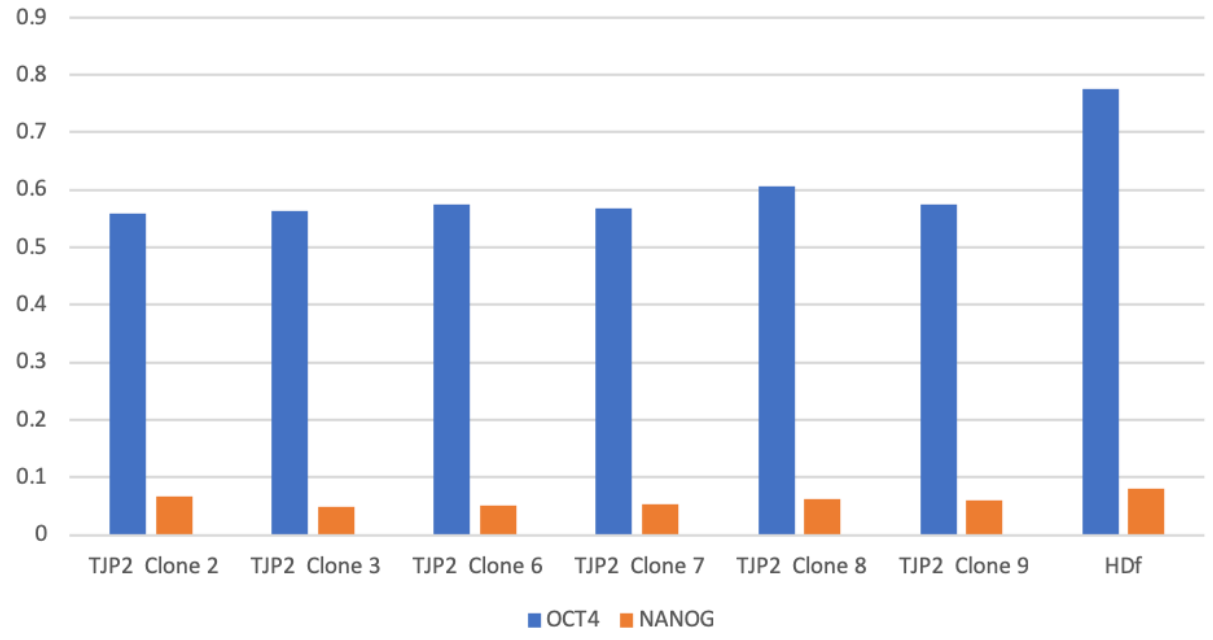



**Figure 1.7 Making *ABCB4* KO iPSCs** (A) Schematic illustration of CRISPR *ABCB4* KO gRNAs' and screening primers' bioinformatic designs. gRNA 1 and 2 are located in intron 3 of the *ABCB4* gene, and gRNA 3 and 4 are located in the middle of intron 5 of the *ABCB4* gene. Therefore, exon 4 is included in the KO region. (B) Diagrams illustrating iPSC clones' PCR screening results. WT iPSCs result in 2.0 kb band; KO clones resulted in 1.0 kb bands. The single bands in KO iPSCs indicate that the clones are homozygous KOs. (C) Diagrams illustrating sanger sequencing results of iPSC$^{ABCB4\text{-}KO}$ clone genomic DNA. KO clone 4E9 cell allele 1 has 1091 bp deletion; allele 2 has 1083 bp deletion.

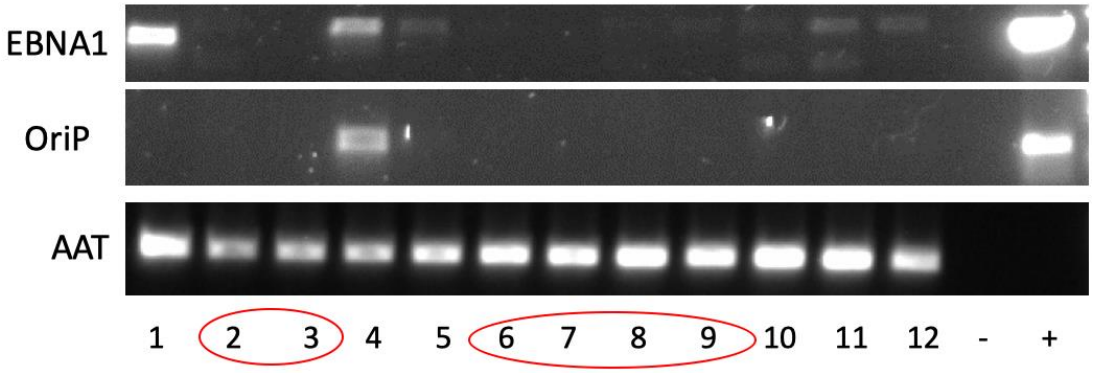
### 1.3.4 Reprogramming patients' secondary PFIC fibroblasts with *TJP2* mutation to iPSCs

DefiniGEN first generated an iPSC line from one patient with truncating *TJP2* mutations by reprogramming their fibroblasts (iPSC$^{PFIC\text{-}patient1}$). Two patient fibroblasts had been collected, but only one fibroblast line was successfully reprogrammed to iPSCs. There were six iPSC clones that had been picked up from the culture plate. All iPSC clones' pluripotency gene expressions were checked by qPCR. All six clones had high expression of *OCT4* and relatively low expression of *NANOG*, which highlighted their pluripotent nature (Fig1.8a) (Radzisheuskaya et al., 2013). Moreover, similar *OCT4* and *NANOG* expressions were also observed in the positive control HDF (a stable iPS cell line) (Fig1.8a). In addition, *oriP/EBNA1* PCR revealed all iPSCs picked were free from *EBNA1* and *OriP* vector sequences; therefore, no external genetic material had been integrated into the genome (Fig1.8b). To confirm the genotypes of generated iPSCs, sanger sequencing was performed on fibroblast$^{PFIC\text{-}patient1}$ and iPSC$^{PFIC\text{-}patient1}$. Sanger sequencing revealed that the homozygous truncating mutation of p.Y261Sfs*50 (c.782delA) in exon 5 of the patient remained after reprogramming (Fig1.8c and Fig1.8d).

(A)

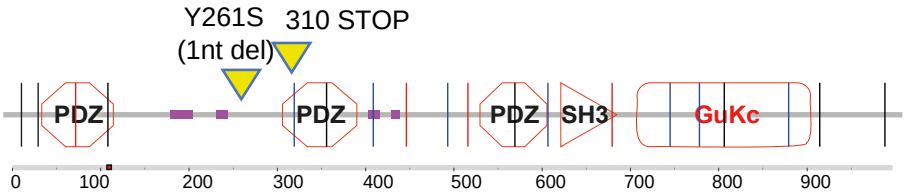


(B)




(C)

Patient 1 & iPSC$^{PFIC\text{-}patient\ 1}$
p.Y261S fs*50




(D)

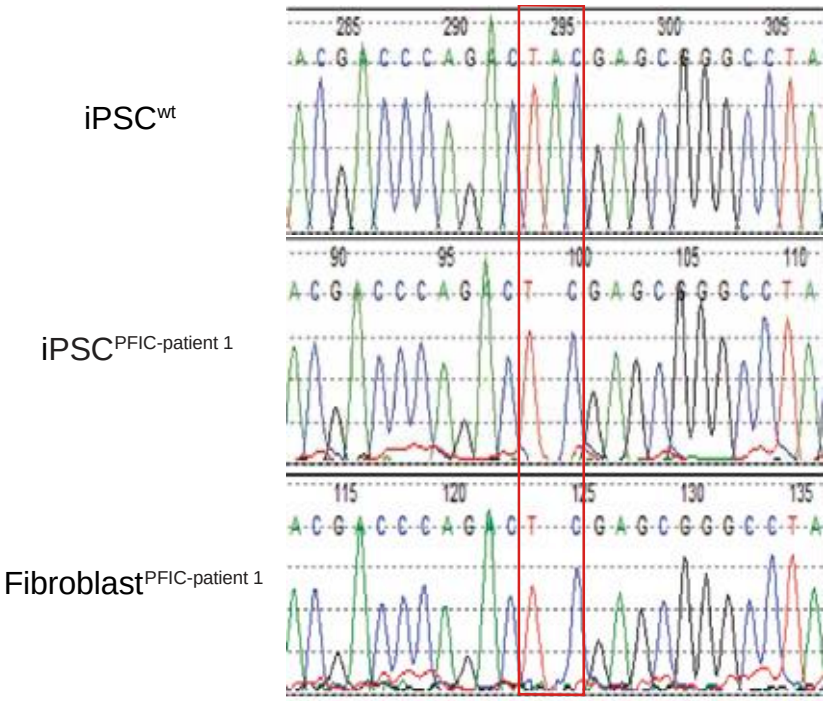


**Figure 1.8 Reprogramming patient fibroblasts with TJP2 mutation to iPSC** (A) Schematic illustration of six reprogrammed iPSC clones' pluripotency markers' qPCR. All six iPSC clones demonstrated similar *OCT4* and *SOX2* expression to positive iPSC control HDF. (B) Diagram illustrating iPSC clones' epi vector clearing results. Clones 2, 3, 6, 7, 8 and 9 were cleared for *EBNA* and *OriP* DNA. However, all iPSC clones above were preserved Alpha-1 antitripsin (*AAT*) DNA. (C) Diagrams illustrating the location of the *TJP2* mutation and protein truncation of *TJP2*. The upper panel shows a mutation and frameshift (c.782delA, p.Y261Sfs*50) in iPSCs derived from patient 1. (D) Schematic illustration of sanger sequencing results of WT iPSCs, $iPSC^{PFIC\text{-}patient\text{-}1}$ and $fibroblast^{PFIC\text{-}patient\text{-}1}$. The red box highlights the mutation from TAC to T-C in $iPSC^{PFIC\text{-}patient\text{-}1}$ and $fibroblast^{PFIC\text{-}patient\text{-}1}$.

## 1.4 Discussion and conclusion

The most technically challenging part of this chapter's work was making a single iPSC survive after FACS. As human iPSCs normally appeared as clusters, the physical separation triggered cell death. Ohgushi et al had shown human ESC dissociation-related apoptosis was caused by ROCK-dependent hyperactivation of actomyosin and efficiently suppressed by the myosin inhibitor Blebbistatin (Ohgushi et al., 2010a). The physical separation between human ESCs caused a loss of E-cadherin-dependent intercellular contact and, hence, actomyosin hyperactivation (Ohgushi et al., 2010b). Given the similarity between human ESCs and iPSCs, similar apoptosis phenotypes appeared after iPSC dissociation. Interestingly, given the mechanism of hPSC dissociation apoptosis as ROCK-dependent, a potent ROCKi, such as Revita, should have reversed the apoptosis phenotype post-single-cell dissociation, and my results confirmed this hypothesis. Having cleared this rate-limiting obstacle from the iPSC CRISPR-Cas9 engineering pipeline, more single-cell iPSC colonies survived, and this increased the pool of cells I could use for PCR screening. Ultimately, this led to the successful generation of *ABCB4* and *TJP2* homozygous KO iPSC lines. There were attempts to generate other cell lines, such as probable phospholipid-transporting ATPase IC, FIC1, encoded by *ATP8B1*, and *ABCB11* KO iPSCs, but due to time limits, I did not manage to complete all the screening for them.

My PhD work set up this working pipeline for CRISPR-Cas9 to engineer iPSCs that could be easily adapted to more precise mutation introduction. The patient fibroblast reprogramming process was not completely successful, given that one patient fibroblast did not reprogram to iPSC.

This reprogramming failure first highlighted the individual differences between patients' fibroblasts. Furthermore, the *oriP/EBNA1* vector was also characterised by low–copy representation in the cells of primates and could be replicated once per cell cycle; this might have caused low reprogramming efficiency (Medvedev et al., 2010). Due to the project's focus on *TJP2*, for the rest of my study, I focused on iPSCs with *TJP2* KO and patients with *TJP2* mutations reprogrammed to iPSCs only. Future experiments should focus on (without being limited to) the generation of iPSCs with PFIC patients' specific mutations and iPSCs with *ABCB11* and *ATP8B1* homozygous KO. A comprehensive iPSC library with PFIC mutations could allow scientists to find *in vitro* disease phenotypic differences between different types of PFIC or the individual mutations within one type of PFIC.

# Chapter 2: Characterising iHEPs and cholestasis disease phenotypes

# 2.1 Introduction

## 2.1.1 cGMP-level iPSCs to hepatocyte-like cell differentiation

End-stage liver diseases, such as liver fibrosis and acute liver injury, exhaust the liver's regeneration capacity, and, without liver transplantation, the diseases will progress and lead to death. Unfortunately, a severe shortage of compatible human liver donors leads to the death of many patients with end-stage liver disease (Washburn and Halff, 2011). Transplanting human primary hepatocytes is the gold standard method of treating end-stage liver disease. However, such treatment requires large cell quantities and has an exceptionally high cost. Therefore, primary hepatocyte transplantation will not become a prevalent treatment option for end-stage liver patients any time soon (Iansante et al., 2018).

Human pluripotent stem cell-derived hepatocytes have been a promising alternative cell source to replace primary hepatocytes in cell therapy. In particular, iPSCs from patients could be used in autologous cell transplantation and, therefore, will not require an immune suppressant. In theory, iPSCs and ESCs are able to expand without limit; thus, they can meet cell therapy demands.

There are a large number of published cytokine-based differentiation protocols, all of which attempt to mimic in vivo endoderm development and endoderm hepatic specification followed by hepatic cell maturation. Important early studies from animal cap cells reveal activin and beta fibroblast growth factors (bFGF) are responsible for mesodermal and endodermal cell induction (Jones et al., 1993). Furthermore, human ESCs and iPSCs are able to differentiate to endoderm cells with the help of activin-A, as the immune fluorescence results suggest cells express endoderm markers, such as SRY-Box Transcription Factor (SOX17*)* and fork–head box protein A2 (FOXA2) (Cai et al., 2007; Rashid et al., 2010b). Other protocols include cytokines, such as fibroblasts growth factor 2 (FGF2), Bone Morphogenetic Protein4 (BMP4) and CHIR, to induce mesoderm cells from pluripotent stem cells (Blackford et al., 2019b). In addition, the presence of Phosphoinositide 3 kinase inhibitor in the culture medium is sufficient to promote activin A-induced endoderm development in human ESCs (McLean et al., 2007). To further induce endoderm cells to hepatocytes, human ESC data suggests FGF and bone morphogenetic protein

(BMP) signalling pathways are needed; their inhibition disrupts hepatocyte differentiation (Mfopou et al., 2010).

From initial mouse studies, we know oncostatin M (*OSM*) is expressed highest during mid-gestation and decreases during the post-natal period, whereas hepatocytes growth factors (*HGF*) is only expressed after birth. In the same study, western blot analysis suggests foetal mouse hepatocytes treated with either OSM or HGF increase *ALB* expression compared to control cells. OSM is shown to boost the hepatic function via signal transducer and activator of transcription 3 (*STAT3*), whereas HGF contributes through a different signalling pathway (Kamiya et al., 2001). In a different mouse study, HGF was mediated by tyrosine-protein kinase Met (c-MET) tyrosine kinase receptors, and *HGF* homozygous KO mice showed incomplete liver development, which caused their deaths in utero (Schmidt et al., 1995). In our lab, OSM and HGF are both supplemented in the basal media with other reagents to gain mature and functional hepatocyte-like cells from pluripotent stem cells (Blackford et al., 2019b).

Less prevalent hepatic differentiation methods require genetic or epigenetic modification. For example, Takayama et al overexpressed *HNF4 alpha* in hepatoblasts through transduction. This led to higher *ALB* and *A1AT* expression compared to the control. In contrast, the expression levels of cholangiocyte markers, such as Cytokeratin-7 (*CK7*), decreased; therefore, *HNF4* overexpression promoted more efficient hepatic differentiation (Takayama et al., 2012). Furthermore, Zhou et al successfully induced functional hepatocytes from mouse ESCs by using Histone Deacetylases (HDAC) inhibitor sodium butyrate, which stopped the ESC cycle, and, consequently, cells could commit to hepatocytes (Zhou et al., 2007).

### 2.1.2 Establishment and regulation of hepatocyte polarity

PFIC secondary to *TJP2* mutation is an inherited disorder affecting tight junction, and, hence, the structural polarity of hepatocytes; this, consequently, leads to cholestasis. In addition, mutations in *ABCB11* can lead to the functional defect of ABC transporter BSEP, hence disturbing the hepatocyte's polarity. Therefore, both structural and functional components contribute to the final hepatocyte polarity phenotype.

Hepatocytes have a unique polarisation arrangement in which two or three hepatocytes' apical membranes contribute to form a tube-like structure in the bile canaliculi. This arrangement allows hepatocytes to have a basolateral membrane facing the sinusoid side with blood and an apical side with bile acid (Fig2.1a). The canaliculi are functionally sealed by tight junction proteins to act like a 'fence' which separates the proteins from the basolateral side to the apical side and also act like 'gates' that control substances coming in and going out of the canaliculi. How this unique polarity is established and regulated is largely unknown in human hepatocytes. Recently, inhibition of Liver kinase B1 (LKB1) or 5' AMP-activated protein kinase (AMPK) in primary sandwich-cultured rat hepatocytes revealed LKB1-activated AMPK was essential for canaliculi structure formation and maintenance (Fu et al., 2010).

A follow-up study used the same experimental setting and found that bile acid could stimulate canaliculi formation through the cAMP-Epac-MEK-Rap1-LKB1-AMPK pathway (Fu et al., 2011). Liver-specific deletion of LKB1 in mice caused defective bile canaliculi formation and accumulation of bile acid in the liver. This was because BSEP became trapped inside the hepatocytes rather than localised to the canalicular membrane (Woods et al., 2011). Another follow-up study cultured the primary hepatocytes from LKB1 KO mice in a collagen sandwich and used live-cell imaging, fluorescence recovery after photobleaching (FRAP), particle tracking and biochemical assays to explain how LKB1 was essential for BSEP trafficking to the canalicular membrane. In the same study, cyclic Adenosine Monophosphate (cAMP) acted through a Protein Kinase A (PKA)-mediated pathway to accelerate BSEP trafficking to the canalicular membrane (Homolya et al., 2014). Indeed, a successful activation of AMPK would lead to greater cell polarisation, which specifically appears in cells as increased canalicular trafficking, tight junction formation and cytoskeleton organisation. Notably, such positive effects on cell polarisation were also seen when the cAMP–PKA pathway was activated. In addition, AMPK activation decreased anabolic processes, such as lipid and protein synthesis as well as growth, whereas it increased catabolism (Fig2.1b).

Although the full mechanism of hepatocyte polarity is not completely known, there are a few key elements that could influence it. First, liver Extracellular Matrix (ECM) is the 3D scaffold made from macromolecules and minerals that gives structural and biochemical support to surrounding

hepatocytes. During mouse liver development and regeneration, laminin is present in the ECM. However, laminin is absent in the mature liver ECM. This lack of basal lamina allows macromolecules to travel from sinusoid blood to the space of Disse through endothelial cell fenestra. Indeed, liver fibrosis patient samples suggest excessive deposition of ECM to the space of Disse changes hepatocyte polarity and causes hepatocyte malfunction (Fig2.1c) (Schuppan et al, 1990). Second, hepatocytes are connected through tight, anchoring gap junctions between the cells.

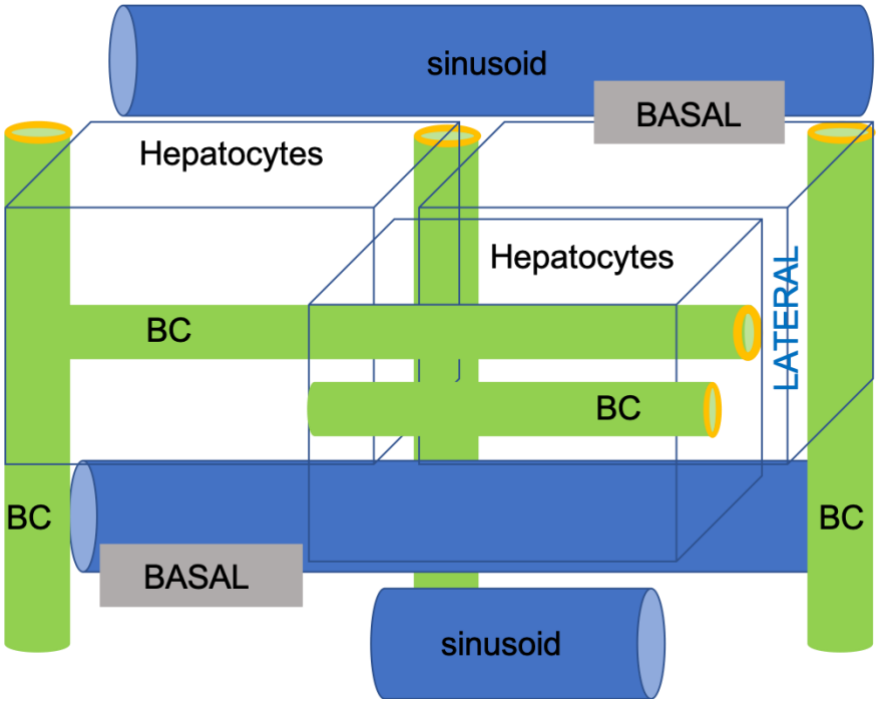
Tight junction is known for preventing leakage of transported solutes and water. More importantly, it seals the paracellular pathway. There are three types of proteins presented in the tight junction complex: claudins, occludins and TJPs. which connect in a specially arranged manner. In patients with PFIC with a *TJP2* mutation, scientists observed the absence of ZO2, which led to the failure of localisation of CLDN1 and tight junction distortion. This is an example of a mutation in tight junction protein that led to a change in hepatocytes polarity and caused abnormal bile acid secretion (Fig2.1c) (Sambrotta et al., 2014a).

Third, hepatocytes use different intracellular protein trafficking pathways to sort polytopic and single membrane-spanning proteins. A study of rat hepatocytes revealed that the immature proteins of polytopic ABC transporters would be labelled in the Transgolgi Network (TGN), then directly sorted to the canalicular membrane, or enter the recycling endosome (RE) pool first then shift to the canalicular membrane. Inversely, a single membrane-spanning protein is first transported to the hepatocyte basolateral membrane, then transcytosed to the hepatocyte apical membrane (Fig2.1c) (Kipp et al., 2001). Wilson disease patients suffer from Cu accumulation causing hepatocellular death. Wilson disease patients have mutations the *ATP7B* gene frequently caused by protein products not being able to travel to the canalicular membrane, hence suffering from Cu accumulation, which causes hepatocellular death.

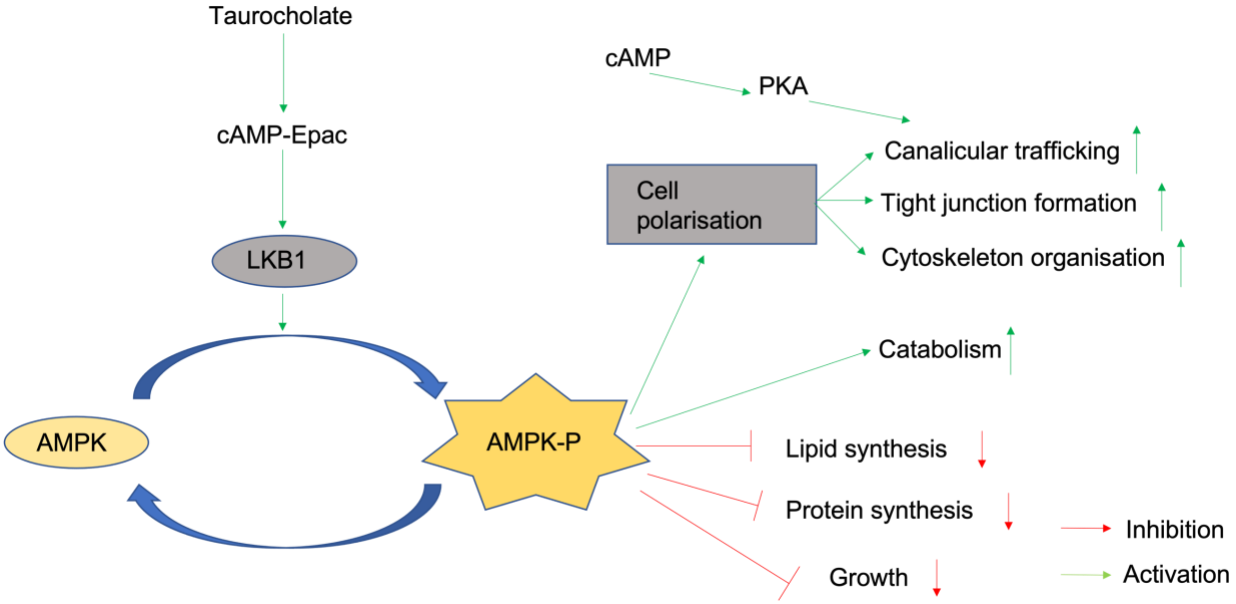
Polishchuk et al illustrate how *ATP7B* travels to the canalicular membrane in HepG2 cells. In response to elevated Cu in hepatocytes, *ATP7B* moves from Golgi to lysosomes. *ATP7B* storage lysosomes stand by, waiting for the next Cu surge to move towards the canalicular membrane; this requires interaction with P62 protein. Finally, Cu ions are released to the bile canaliculi via

lysosomal exocytosis (Polishchuk et al., 2014). Fourth, cytoskeletal microfilament and microtubular networks indirectly affect hepatocyte polarity by changing ABC transporter trafficking in cells (Fig2.1c). Wakabayashi et al demonstrate how in the WIF-B9 cell, BSEP is transported from the TGN to the canalicular membrane along microtubules (Wakabayashi et al., 2004). Last, AMPK is known as a necessary element in cell metabolism that is required for maintaining energy homeostasis (Fig2.1c). LKB1 activates AMPK to suppress growth and proliferation when cellular energy is low. Previously, rat hepatocyte sandwich-cultured evidence suggested taurocholate activates the cAMP-Epac-MEK-Rap1-LKB1-AMPK pathway to help canaliculi network formation. Therefore, logically, hepatocyte polarity would be energy dependent, but the mechanism was unclear (Gissen and Arias, 2015a). Overall, almost all reported inherited genetic cholestasis has mutations in genes that make important contributions to hepatocyte polarity. Therefore, more mechanistic studies on normal and abnormal hepatocyte polarity development are required to understand how the loss of hepatocyte polarity leads to cholestasis.

(A)


sinusoid
BASAL
Hepatocytes
Hepatocytes
LATERAL
BC
BC
BC
BC
BASAL
sinusoid

(B)


Taurocholate
cAMP-Epac
LKB1
AMPK
AMPK-P
cAMP
PKA
Cell polarisation
Canalicular trafficking
Tight junction formation
Cytoskeleton organisation
Catabolism
Lipid synthesis
Protein synthesis
Growth
Inhibition
Activation

(C)

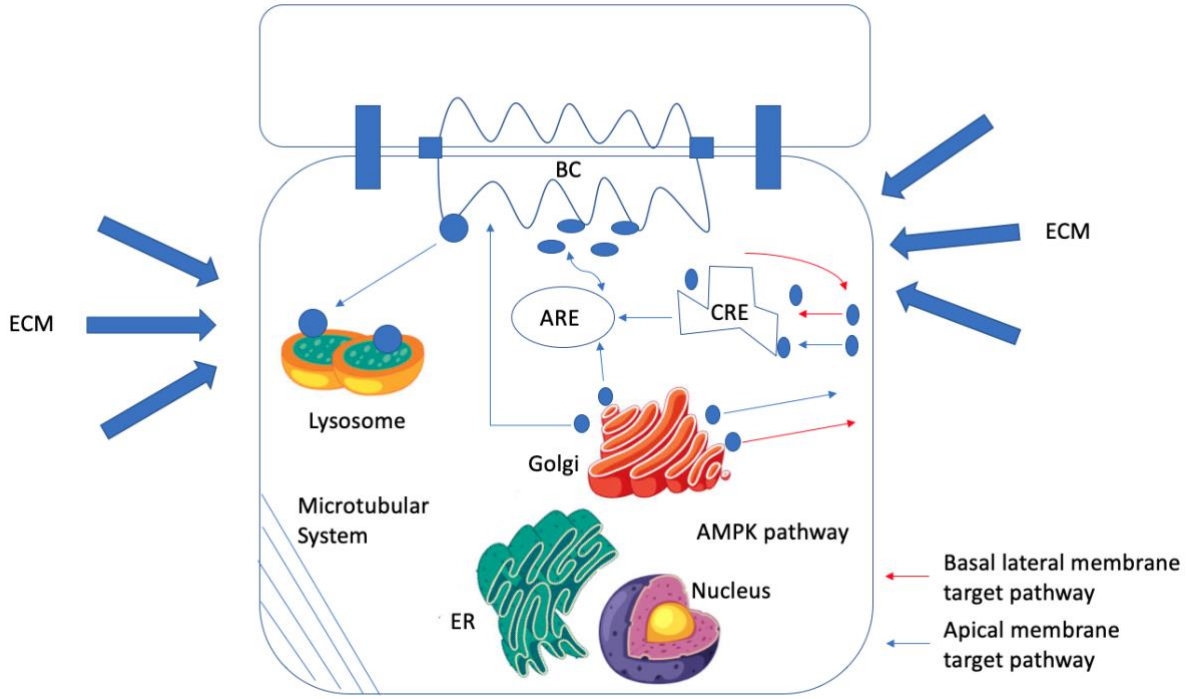



**Figure 2.1 Epithelial hepatocyte polarity** (A) Schematic illustration of connected hepatocytes forming bile canaliculi (green) between the apical membranes of hepatocytes. The canaliculi lumen is sealed by the surrounding tight junction (yellow). Hepatocytes have two basal domains that face the adjacent sinusoids. One hepatocyte forms canaliculi lumen and can connect with three adjacent bile canaliculi. Figure was adapted from (Gissen and Arias, 2015b). (B) Signalling pathway in hepatocyte polarity. Taurocholate stimulates microtubular-dependent trafficking by activating the cAMP-Epac pathway; this switches on LKB1, which then activates AMPK. There are three main effects of AMPK activation. First, it enhances cell polarisation by stimulating canalicular trafficking, tight junction formation and cytoskeletal organisation. Second, it enhances catabolism in hepatocytes. Third, it inhibits anabolic activities, such as lipid and protein synthesis as well as growth. cAMP activation leads to increased intracellular trafficking via a PKA-dependent pathway. This process is independent from AMPK. Figure was adapted from (Gissen and Arias, 2015b). (C) Summary of key elements influencing hepatocyte polarisation. Extrinsic factors, such as ECM, play a key role in hepatocyte polarity development, as they interfere with the signalling pathway in hepatocytes related to polarity development. Absence of tight junction proteins can lead to a change in hepatocyte polarity, as this alters the bile canaliculi structure and interferes with related signalling pathways. Intracellular protein trafficking can affect hepatocyte polarity. Apical membrane proteins travel to the canaliculi membrane via organelles, such as the apical recycling endosome (ARE). Basolateral membrane proteins travel to basolateral membranes via organelles such as the common recycling endosome and Golgi. Details of intracellular protein trafficking

mechanisms are largely unknown. Canalicular membrane ABC transporters traffic Golgi to apical membranes along microtubules; therefore, changes in microtubules also affect hepatocyte polarity. AMPK controls energy metabolism and, hence, affects hepatocyte polarisation, but the mechanism is unclear. Figure was adapted from (Gissen and Arias, 2015b).

### 2.1.3 PFIC with *TJP2* mutation's potential mechanisms

It is still unclear how *TJP2* mutations lead to cholestasis in patients and why different mutations cause different disease onsets for different patients. In order to design effective therapeutic interventions for PFIC with *TJP2* mutation, scientists need a better understanding of the disease's molecular mechanisms. From reported epithelial cell literature, ZO2 is shown as the master regulator of gene expression, cell proliferation, cytoarchitecture and cell size. There are multiple pathways associated with ZO2, depending on the function ZO2 serves and the environment it is in. It is distributed accordingly at the tight junction, cytoplasm and nucleus (Lechuga et al., 2010). Here, I summarise ZO2's reported functions and its relevant molecular pathways. MDCK cells transfected with cyclin D1 gene and ZO2 separately were used to study ZO2's effect on cell proliferation. The combined results from gene reporter assay, siRNA, real-time PCR, and DNA synthesis assay revealed *TJP2* overexpression downregulated cyclin D1 expression, hence inhibiting cell proliferation. From the same study, ChIP and HDAC inhibitor experiments further illustrated ZO2 interactions with *C-MYC*. The complex then binds to the Enhancer Box (E-BOX) of cyclin D1 to recruit HDAC in order to prevent the E-BOX from opening. This results in cyclin D1 transcription repression (Lechuga et al., 2010). This is an example of ZO2 regulating cell proliferation by inhibiting cyclin D1 gene expression.

ZO2 also plays an important role in regulating cell apoptosis. Duplication of WT *TJP2* gene in humans has led to progressive nonsyndromic hearing loss. From the same study, Reverse Transcription Polymerase Chain Reaction (RT-PCR) data suggests an increase in BCL2-like protein 11 gene, *BCL2L11*, expression in lymphoblast cells of affected individuals. BCL2L11 had been previously reported to promote apoptosis via the intrinsic mitochondrial pathway. Overall, duplication of *TJP2* leads to increased pro-apoptotic gene expression and increased susceptibility to inner ear cell apoptosis (Walsh et al., 2010).

Another factor that can contribute to inner ear hair cell loss is the presence of the mutated deubiquitinating enzyme Ubiquitin Specific Peptidase 53 (USP53). Scientists observed progressive hearing loss and outer hair cell (OHC) dysfunction in mambo mice with a missense mutation in the *USP53* gene. Immunofluoresence data revealed USP53 localised to the tight junction in cochlear epithelial cells and formed a complex with ZO1 and ZO2 proteins. Furthermore, mambo mice's EP decreased by almost 17% compared to the control mice. This was possibly caused by tight junction barrier function and increased paracellular flux of K+ ions. Indeed, ex vivo culture of organs of Corti from P7 homozygous mambo mice suggested a significant loss of OHC compared to their heterozygous mambo counterparts. Overall, mutation of USP53 destablises ZO2 and, hence, alters the EP by leaking K+ ions through the impaired tight junction from endolymph to perilymph. This disruption of cochlear homeostasis results in OHC apoptosis and hearing loss in homozygous mambo mice (Kazmierczak et al., 2015).

ZO2 also regulates the cytoarchitecture of epithelial monolayers. Scientists knocked down ZO2 in MDCK cells and observed enlarged intercellular space and lower cell-to-substratum attachment. In the same study, knocking down ZO2-activated small GTPase (Ras), Ras homolog family member A (RhoA) and ROCK led to an increase in vinculin and stress fibres in MDCK cells. Furthermore, ZO2 knocked down increased cell cycle regulator cell division control protein 42 homolog (CDC42)'s activity as well as mitotic spindle disorientation in MDCK cells. MDCK cells had been known for their directional migration. However, by knocking down ZO2, actin polymerisation destabilisers' Rac1 and cofolin's activity increased; hence, scientists observed random cell migration in MDCK cells. Overall, the absence of ZO2 in MDCK cells leads to changes in cytoarchitecture, activating the Rho/ROCK pathway to alter intracellular actin networks or cell migration patterns (Raya-Sandino et al., 2017).

ZO2 has a PDZ domain that can interact with its counterpart in the Yes-Associated Protein (YAP)/Transcriptional activator with PDZ-binding domain (TAZ) complex and regulates its activity. It has been demonstrated that YAP2's localisation is dependent on ZO2 in human embryonic kidney (HEK293) cells using co-immunostaining. When the NLS of ZO2 was mutated, YAP2's nucleus stain was significantly decreased in the cells. Furthermore, the regulatory role of ZO2 is dependent on cell type. In MDCK cells, ZO2 prevents YAP2 cells' proliferative role,

whereas in HEK293 cells, ZO2–YAP2 forms a complex to promote cell detachment and apoptosis (Oka et al., 2010). In addition, TAZ also interacts with ZO2 in the MDCK cells via the PDZ domain. This was demonstrated by the protein co-localisation in immunostaining. TAZ was fused with gal4, which stimulated the luciferase activity. Interestingly, there was a large decrease in gal4-dependent luciferase activity when the expression of ZO2 increased (Remue et al., 2010). This study also attempted to use shRNA to knock down the ZO2 and expected to reverse its inhibitory effects. However, the luciferase activity did not change. Overall, ZO2 can bind YAP/TAZ and act as a shuttling protein to move between the cytoplasm and nucleus. Its role in regulating YAP/TAZ-mediated gene transcription remains elusive.

### 2.1.4 Currently existing cholestatic disease models

Human primary hepatocyte is a good type of *in vitro* cell models to study liver physiology. However, conventional 2D hepatocyte cultures rapidly lose many liver-specific functions, cell polarity and structure, including bile canaliculi, and redistribute canalicular membrane proteins. Prolonged culture of fresh human primary hepatocytes decreases the cell's viability. Early studies describe sandwich culture hepatocytes (SCH) systems, in which rat or human primary hepatocytes are cultivated between two layers of ECM scaffolds, usually made of collagen type I or GeltrexTM/MatrigelTM extracted from murine tumours (Dunn et al., 1989; Gross-Steinmeyer et al., 2005) (Fig2.2a). All studies suggest rat and human hepatocytes cultured in sandwiches developed intact 'chicken-wire' bile canalicular networks and maintained expression of polarity proteins enriched in the apical and basolateral membranes. In particular, bile canaliculi in rat SCHs quickly developed, between 24 and 48 hours, followed by more uniform and homogenous networks through days three to seven. The bile acid excretion also reached a maximum around this time. Such observation was also seen in human iHEPs in our study. Thus, all evidence suggests SCHs are a user-friendly and functionally competent system for mid- to long-term cholestatic disease and liver physiology studies (Swift et al., 2010).

Imagawa et al demonstrate PFIC2 patient iHEPs in Matrigel sandwiches were defective in biliary excretion (Imagawa et al., 2017). They used the conventional biliary excretory index (BEI) to measure bile acid transportation from hepatocytes to the canaliculi. The initial protocol involved culturing SCHs in HANKs balanced salt solution (HBSS) cacium2+ (Ca2+) or HBSS Ca2+ free

medium for 10 min. Ca2+ is crucial for canaliculi integrity. Therefore, any taurocholate transports from hepatocyte to canaliculi will leak away under such conditions. All cells were then incubated with a standard HBSS medium to stop the destruction of canaliculi. After three more HBSS rinses, SCH samples would be lysed and quantified by mass spectrometry. The bile transportation could be calculated by subtracting taurocholate accumulated in the cells from taurocholate accumulated in the cell and canaliculi. However, such protocol needed to be cautiously performed, as leaving SCHs in HBSS Ca2+ medium for too long would lead to canaliculi reform, causing inaccurate readouts (Swift et al., 2010).

The SCH culturing system's biggest disadvantage is that its fluid system is static. Future SCH cultures can be based on transwells or microfluidic devices to incorporate sinusoidal blood flow in one direction and exterioris bile flow from the opposite direction. This setup is similar to in vivo human liver physiology and is likely to improve the accuracy of the liver disease model. In fact, the human liver is made up of two main types of cells: parenchymal and nonparenchymal cells. The functional hepatocytes make up 65% of the total liver cells; nonparenchymal cells comprise the remaining 35% of total cells. The nonparenchymal cells can be further divided into sinusoidal endothelial cells, kupffer cells, stellate cells and hepatic natural killer (NK) cells. Thus, it is highly important to incorporate the relevant nonparenchymal cells along with the hepatocytes together in cholestatic liver disease models (Vekemans and Braet, 2005) (Fig2.2a). There are two main reasons for this: first, incorporating nonparenchymal cells enhances the hepatocyte's functions; second, when studying immune-mediated DILI, the presence of immune cells is indispensable for accurate disease modelling (Asai et al., 2017; Bonzo et al., 2015).

In contrast to randomly distributed co-cultures, micropatterned co-cultures (MPCCs) allow precise modulation of Neural Progenitor Cells (NPCs) without significantly affecting homotypic interactions between hepatocytes, which are important for maintaining cell polarity and stability (Fig2.2a). Khetani et al made a MPCC in which human hepatocytes were organised into many islands surrounded by supportive stromal cells. The MPCC hepatocytes maintained in vivo-like morphology and displayed functional canaliculi. BSEP inhibitor troglitazone was removed from the market due to its severe cholestatic nature. In this study, even 7.5 uM of trigolitozon was able

to cause the BEI to decrease, suggesting the sensitivity of MPCC in predicting DILI (Khetani and Bhatia, 2008).

Another commonly used DILI disease model is hepatocyte spheroids. In these systems, spheroids are obtained from random self-aggregation of hepatocytes through spontaneous cellular self-assembly mediated by cell-released ECM (Fig2.2a). Non-attachment to the substrata and oxygen supply was found to be critical for the spread of spheroids and the maintenance of functional characteristics, respectively. This setup allows sufficient cell–cell interaction and preserves core hepatocyte function and polarity. Most importantly, hepatocyte spheroids express key apical membrane transporters, such as BSEP and MRP2, and are competent in bile acid transportation (Hendriks et al., 2016a). Hendriks et al made spheroids separately from primary hepatocytes and HepaRG cells (Hendriks et al., 2016b). The spheroids were exposed to compounds with cholestatic liability and bile acid (Fig2.2a). The Cholestatic Index (CIX) used in the study was defined as the ratio between the EC50 value from co-exposure to a compound bile acid (BA) mixture and the EC50 value from exposure to the compound alone. After 14 days of compound repeated exposure to the spheroids from primary hepatocytes (PHH), all cholestatic compounds had lower CIXs compared to the noncholestatic compounds. Furthermore, this study also illustrated that when spheroids were given the cholestatic drug Chlorpromazine (CPZ), they had cholestatic patterns, which appeared as an accumulation of fluorescently labelled bile acid (Hendriks et al., 2016a).

One of the key advantages of using hepatic spheroids is that it allows long-term DILI injury study. As Bell et al. show, PHH spheroids could be stably maintained in cultures for at least five weeks (Bell et al., 2016). More specifically, 3D hepatic spheriods show higher sensitivity in predicting DILI drugs compared to their monolayer counterparts (Takayama et al., 2013). Despite all the advantages, spheroids were still lacking sufficient cell-to-ECM interactions and often suffered from cell necrosis at their centres.

Although hepatic spheroids made from scaffolds have cell–ECM interactions, the cells are not embedded in biomaterials, ultimately unable to precisely modulate the biophysical and biochemical environment. These challenges can be mitigated with FDA-approved natural (alginate- and cellulose-based) or synthetic (polyethylene glycol (PEG)-based) biomaterials that

serve as engineered polymer scaffolds to facilitate intercellular contact, spheroidal assembly and spread while shortening the time required for spheroid formation (Ng et al., 2018).

In our lab, we engineered inverted colloid crystals (ICCs), whose 3D environment could resemble the extracellular niche of the immature human liver (Fig2.2a). The manufacture involved laying regular-sized polystyrene beads into a square Polydimethylsiloxane (PDMS) mould to create a crystal lattice. Then, the lattice is infiltrated by poly glycidyl methacrylate PGEDA. With the help of UV light, PEG starts to polymerise, and the lattice is finally removed by adding tetrahydrofuran. The ICC scaffold coated with collagen-1 allows hepatocytes to attach and create sufficient cell–cell and cell–ECM interaction.

There were two phases for hepatocytes to settle in ICCs, iHEPs or foetal hepatocytes: first, they underwent cell attachment (phase1), then organoid formation (phase2). Moreover, iHEPs formed organoid cultures in ICCs that not only mimicked the human liver in morphology but also resembled its transcriptomic and protein expression profiles. Consequently, the 3D liver microenvironment means iHEPs in ICCs outperform iHEPs in 2D in many hepatic functional tests, such as ALB production and CYP3A4 activity. Most importantly, ICCs preserved their bile acid transporting capacities. Hence, when CLF was added to the culture, there was intracellular CLF accumulation. When cholestatic drug Troglitazone was added with CLF, there was a significant drop in intracellular CLF accumulation. Additionally, hiPSC-hepatocytes like cells (HLCs) could integrate, vascularise and function following in vivo implantation into livers of immune-deficient mice (Ng et al., 2018). Collectively, this study reveals iHEPs in ICCs' disease modelling and cell therapy potential.

In a different study, Vernetti et al generated hepatic organoids from pluripotent stem cells in a stepwise low/medium-throughput manner (Fig2.2a) (Vernetti et al., 2016a). Immunofluorescence data suggested there were two cell types – hepatocytes and cholangiocytes – produced at the end of the differentiation. The spheroids expressed key apical membrane transporters and tight junction proteins, and due to cholangiocyte's presence, biliary cyst structures also formed in the organoids. After the addition of CLF into the hepatic organoids, real-time live imaging suggested bile acids flowed from the hepatocytes' canaliculi to biliary cysts. In addition, incubation of carboxy-

dichlorofluorescein fluorescein (CDF) and Troglitazone led to a strong buildup of cytoplasmic CDF in the cells of the organoids, and a prolonged incubation of the cholestatic drug caused a clear loss of bile canaliculi (Ramli et al., 2020). This study provided scientists with the opportunity to study cholestatic injury not only to hepatocytes but also to cholangiocyte-formed structures.

Despite all the advantages spheroids have in resembling the liver's in vivo structure, they still lack a few key physiological characteristics. Microfluidic devices, however, have advantages such as controlled temporal and spatial regulation of cell arrangement and mechanical stimulation, e.g. fluidic flow and shear stress, as well as pH and temperature maintenance, nutrient and oxygen exchange and waste product removal. This flow is imperative for the formation of oxygen/nutrient/hormone gradients necessary for a sustainable hepatic function (Gough et al., 2020) (Fig2.2a). The current human biomimetic liver microphysiology (HBL-MPS) includes multiple cell types, such as primary hepatocytes, liver sinusoidal endothelial cells (LSEC) and hepatic stellate cells. There are two ways to develop HBL-MPS platforms to resemble the physiological environment of the human liver acinus. The traditional way involves the careful design and manual placement of multiple cells into the HB; the advanced way involved 3D bioprinting of cells and matrices to mimic the organisation of cells in the liver. Once the cells are seeded into the HBL-MPS device, the fluid currents can flow in the 'upper hepatic channel' and lower 'vascular channel' separately but in an opposite direction, mimicking the blood–bile countercurrent flow.

There are three key characteristics of the current HBL-MPS that makes it an excellent cholestatic disease model: first, the system can combine, exclude and/or engineer cell types of interest as well as fine-tune the physical and biochemical environment in a given study; second, the system is suited to testing small molecules or other therapeutic agents; third, the HBL-MPS can be disassembled into smaller parts to allow specific cell types and supernatants. Future HBL-MPS can start to incorporate patient iPSC-derived liver cells for patient-specific cholestasis disease modelling (Gough et al., 2020). The ultimate goal is to build a microphysiology system database (MPS-DB) to collect, analyse and form a conclusion based on data generated from the DILI disease model (Fig2.2b).

Vernetti et al designed the sequential self-assembly liver model (SQL-SAL) as a proof of concept for the idea. Four distinct liver cells were included in SQL-SAL. There was a media/drug influx entry site on one side and a media efflux exit site that directly connected to machines to measure biochemical potential of hydrogen (pH), oxygen and metabolism of the media. The SQL-SAL chip was also able to carry out high content imaging-based bile transportation assay, and the images were analysed by software. Both the high content imaging analysis results and media biochemical readouts could be stored in an MPS-DB. Combining the data from other databases, such as Drugbank or Pubchem, with the SQL-SAL data, a smart algorithm could be built over time to continue building a better disease model as long as there is new information fed into the system (Vernetti et al., 2016b) (Vernetti et al., 2016a).

The upgraded version of the microfluidic device was named ‘organs on a chip’ and is composed of multiple cell types from multiple organs (Fig2.2a). This state-of-the-art artificial disease model allows us to understand the effect of drug metabolism on the liver and other tissue types. Such a model poses a huge engineering challenge, requiring a ‘communal’ media fit for all cell types. From existing knowledge, terminally differentiated cells mediums have specific growth factors to maintain the cells’ somatic stage morphologies and functions (Blackford et al., 2019b). Nevertheless, scientists designed a human iPSC-derived four-organ chip, in which liver, brain, intestine and kidney cells were derived from a single iPSC. Terminally differentiated cells under static 2D cultures were characterised by relevant markers’ immune fluorescent staining before they were put into the chip. A common growth factor-depleted medium was used in the surrogate blood circuit during co-culture, and all cells were shown as viable during the 14-day period. RNA sequencing was performed multiple times during the cultures, the results strongly indicating that long-term multicell co-culture had a profound effect on the RNA expression profile. However, this study did not perform real-time functional analysis for each cell type to ensure the function was preserved during the 14-day cultures (Ramme et al., 2019).

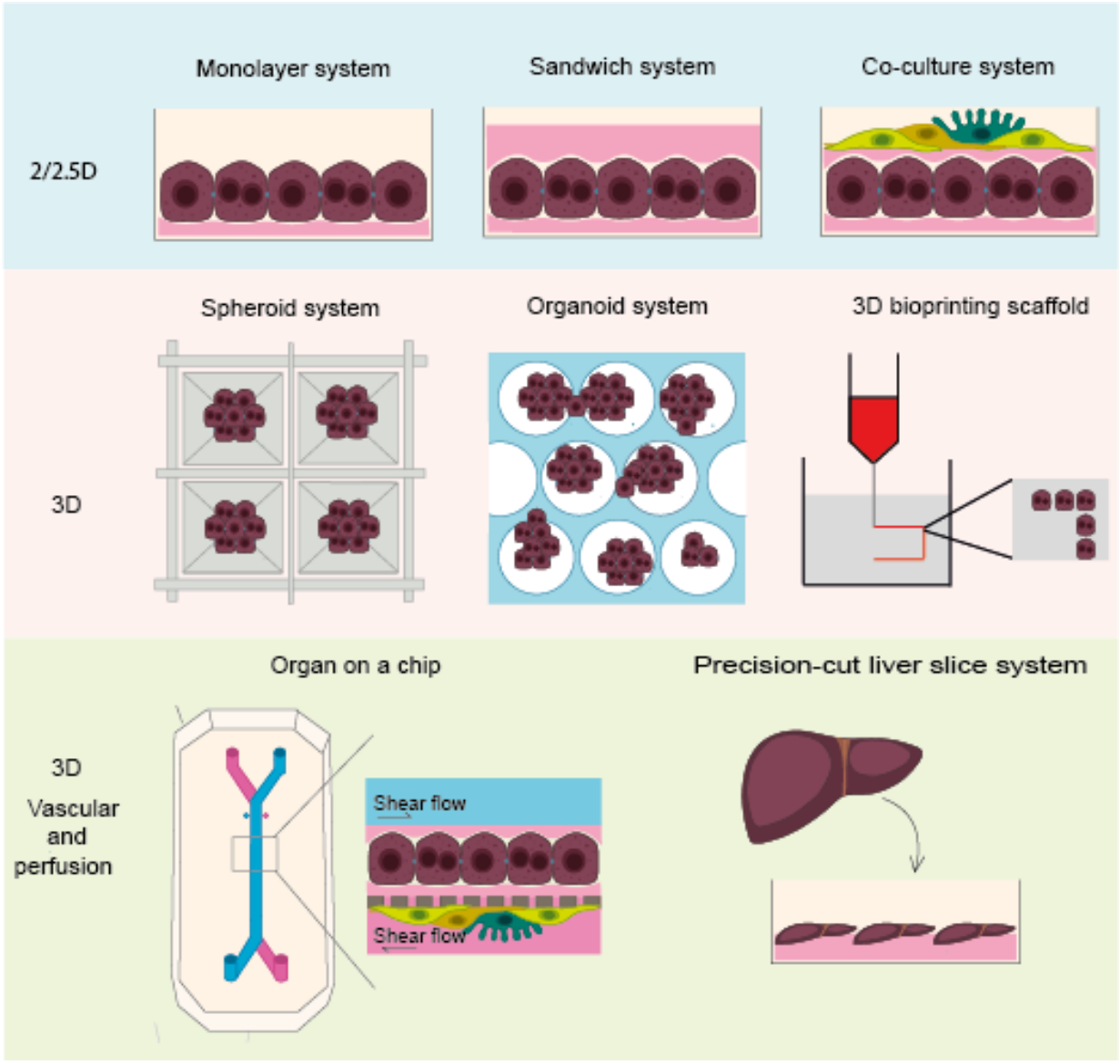
Viravaidya et al described microscale cell culture analogue (UCCA) composed of two chambers used to culture liver and lung cells. The other two compartments were ‘fat’ and ‘other tissue’, which had no cells but mimicked the distribution of fluid in rapidly and slowly perfused tissues. Hepatoxcin naphthalene was converted to its metabolite in the UCCA liver chamber. This

metabolite then circulated to the lung chamber, causing lung cell damage by depleting glutathione (GSH) (Viravaidya et al., 2004). This example demonstrates that liver metabolites' toxicity to the lung could be recapitulated in the UCCA disease model, hence, increasing the chances for similar diseases, such as A1ATD.

The best cholestasis model to date is precision-cut liver (PCL) slice, in which all liver cells are present as well as the intact liver ECM (Fig2.2a). Hence, hepatocytes from PCL preserve better function and structure than their counterparts from all other culture models (Starokozhko et al., 2017). When PCL is exposed to three well-known cholestatic drugs in the presence of bile acid in a dose-dependent manner, there is a clear increase in bile acid inside the hepatocytes quantified by Liquid Chromatography/Mass Spectrometry (LC/MS). In addition, cholestatic drugs also cause a decrease in sodium/taurocholate transporting polypeptide (*NTCP*) gene expression; therefore, fewer bile acids can get into the cells. This suggests that PCL mimicks the patient cholestasis phenotype (Starokozhko et al., 2017). However, the key disadvantage is that cells in PCL do not survive beyond two days; therefore, PCL currently does not support long-term DILI study.

Overall, as scientists increase the complexity of the disease model, it starts to lose the throughput. More importantly, cholestasis diseases have different disease pathologies. Therefore, there is no single disease model 'fit for all', but there exists a 'fit for purpose' customised cholestasis disease model.

(A)


Monolayer system
Sandwich system
Co-culture system
2/2.5D
Spheroid system
Organoid system
3D bioprinting scaffold
3D
Organ on a chip
Precision-cut liver slice system
3D
Vascular
and
perfusion
Shear flow
Shear flow

(B)

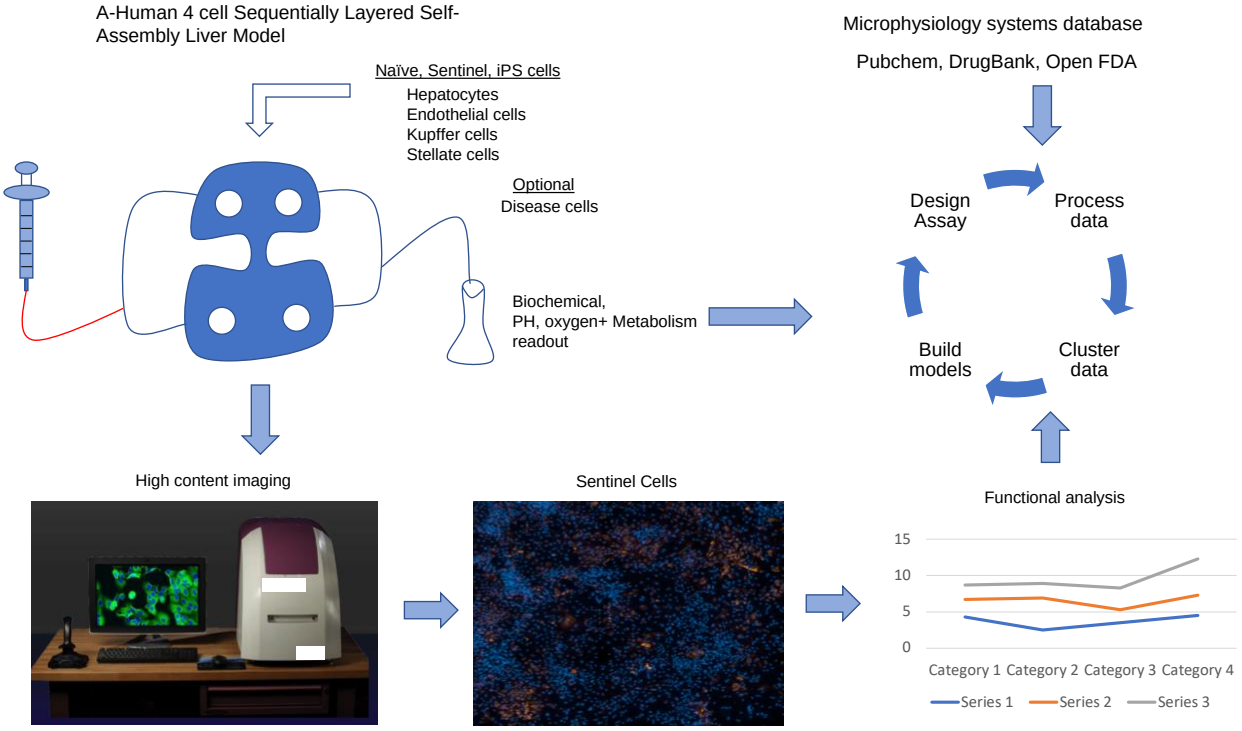



**Figure 2.2 Summary of existing and advanced in vitro cholestasis disease models** (A) Schematic illustration of cholestasis models with varying complexity. 2/2.5D models include a monolayer system, a sandwich culture and a co-culture system. More complex 3D models include spheroids, organoid systems and bioprinting scaffolds. The most complex system involves 3D models with vascular and perfusion systems; examples include organs on a chip and PCL slices. (B) Diagram illustrating an overview of an advanced human liver microphysiology platform (HLMP) for studying cholestasis. The HLMP is composed of four-cell SQL-SAL, which has a microfluidic device integrated into the system. Biochemical data are collected from biosensors in the system or from the extracted culture media, which are later subjected to external biochemical assays. However, high content imaging readouts of live fluorescent bile flow functional assays can be collected from the operetta. All data will be stored in an MPS-DB, which is further connected with external databases, such as DrugBank and PubChem. All information will be used to improve existing cholestasis disease models to gain a better understanding of the disease. Figure adapted from (Vernetti et al., 2016c)

### 2.1.5 Aims and objectives for this chapter

The previous chapter summarised the manufacture of a toolbox of iPSC PFIC disease models. However, until the iHEPs demonstrated patient-relevant structural and functional disease phenotypes, they were not quantified as disease models. Therefore, this chapter aimed to integrate the cutting-edge low/medium-throughput imaging assay with existing sandwich cultured iHEPs in order to generate a disease model which reproducibly delivers accurate patient disease phenotypes in a low/medium-throughput manner. The following sections illustrated the following three things:

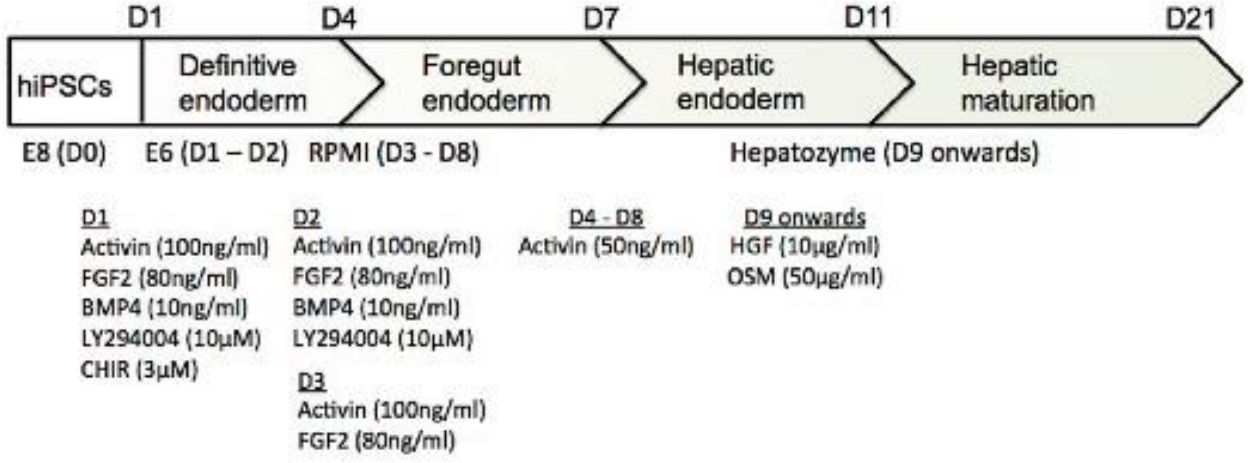
1. How I generated a functional and polarised iHEP in the Matrigel sandwich system.
2. How I established a low/medium-throughput functional assay to model PFIC with *TJP2* mutation.
3. How I quantified the disease phenotype I generated from this disease model.

## 2.2 Methods

### 2.2.1 iPSC maintenance, differentiation and Matrigel sandwich making

Matrigel sandwich-based iHEPs are derived according to (Blackford et al., 2019a). Human iPSCs are passed using gentle cell dissociation reagents (stem cell technologies), then transferred onto plates pre-coated with gelatine. Cells are then grown in three different culture media: Essential 6 media (E6), RPMI and Hepatozyme supplemented with seven different differentiation factors, i.e. CHIR99021 (3 μM; Stemgent), Ly294002 (10 μM; Calbiochem), Activin (100 ng/ml; R&D Systems), FGF2 (40 ng/ml; R&D Systems), BMP4 (10 ng/ml; R&D Systems), HGF (20 μg/ml; Peprotech) and Oncostatin-M (10 μg/ml; R&D Systems), following the order shown in (Fig2.3a). Day 1 to day 4 was a stage of development from hiPSC to definitive endoderm. From day 4 to day 7, the cells developed into foregut endoderm. Then, from day 7 to day 11, the hepatic endoderm cells started to mature, and from day 11 onwards, mature hepatocytes appeared. For cells to establish polarity in the Matrigel sandwich, on day 21 of hepatic differentiation, 6% Matrigel (Corning) was overlaid on top of iHEPs cultured in a 96-well plate for one week.

(A)


D1
D4
D7
D11
D21
hiPSCs
Definitive endoderm
Foregut endoderm
Hepatic endoderm
Hepatic maturation
E8 (D0)
E6 (D1 – D2)
RPMI (D3 - D8)
Hepatozyme (D9 onwards)
D1
Activin (100ng/ml)
FGF2 (80ng/ml)
BMP4 (10ng/ml)
LY294004 (10μM)
CHIR (3μM)
D2
Activin (100ng/ml)
FGF2 (80ng/ml)
BMP4 (10ng/ml)
LY294004 (10μM)
D3
Activin (100ng/ml)
FGF2 (80ng/ml)
D4 - D8
Activin (50ng/ml)
D9 onwards
HGF (10μg/ml)
OSM (50μg/ml)

(B)

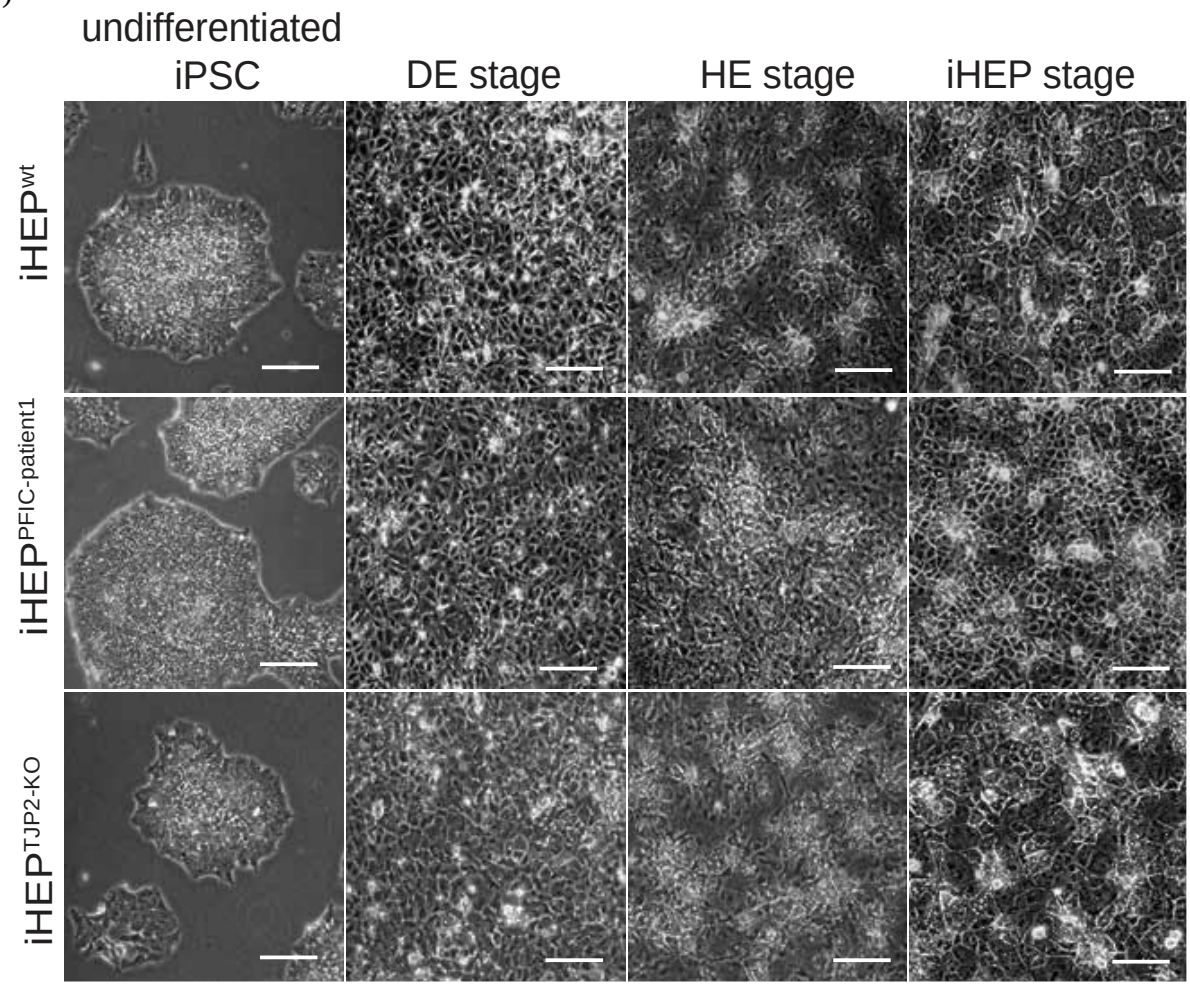
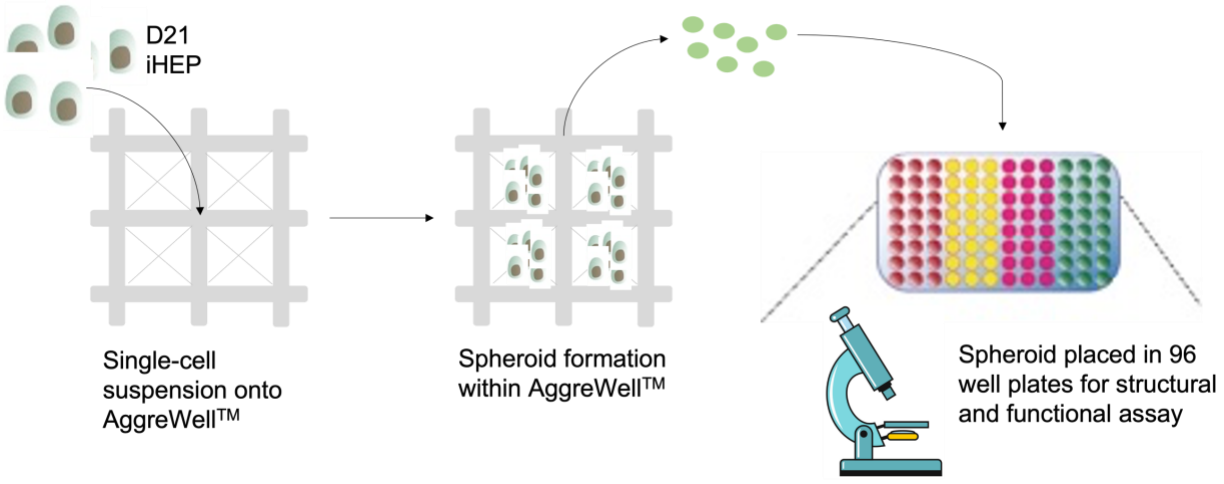

undifferentiated iPSC
DE stage
HE stage
iHEP stage
iHEPwt
iHEPPFIC-patient1
iHEPTJP2-KO

**Figure 2.3 A summary of iPSCs to hepatocyte-like cell differentiation** (A) Schematic illustration of iPSCs to iHEP differentiation. Day 1's E6 culture contained Activin a, FGF2, BMP4, LY294004 and CHIR, Day 2's E6 culture included Activin, FGF2, BMP4 and LY294004. From day 3 to day 7, the culture media switched to RPMI+B27. Day 3's growth factors included Activin and FGF2. Day 4 to day 8's Activin dose was halved. From day 9 onwards, the culture media switched to hepatozymes supplemented with HGF and OSM. (B) Brightfield microscopy images revealed the morphological changes of each iPSC line at each stage of differentiation. Scale bar, 100µm.

## 2.2.2 iHEP spheroid making

Once iHEPs were differentiated on day 21, TrypLE 1x (Thermo Fisher Scientific) was used to generate a single-cell suspension of iHEPs. $0.3 \times 10^6$ cells were seeded per well of a 24-well Aggrewell-400. Mature day 34 iHEP spheroids were subjected to functional and structural assays on 96 well plates (Fig2.4a) (STEMCELL Technologies). Aggrewell plates were prepared as recommended by the supplier. Centrifugation at 200 g for 3 min was carried out to deposit cells into the microwells of the plate (Fig2.4b).

(A)



(B)

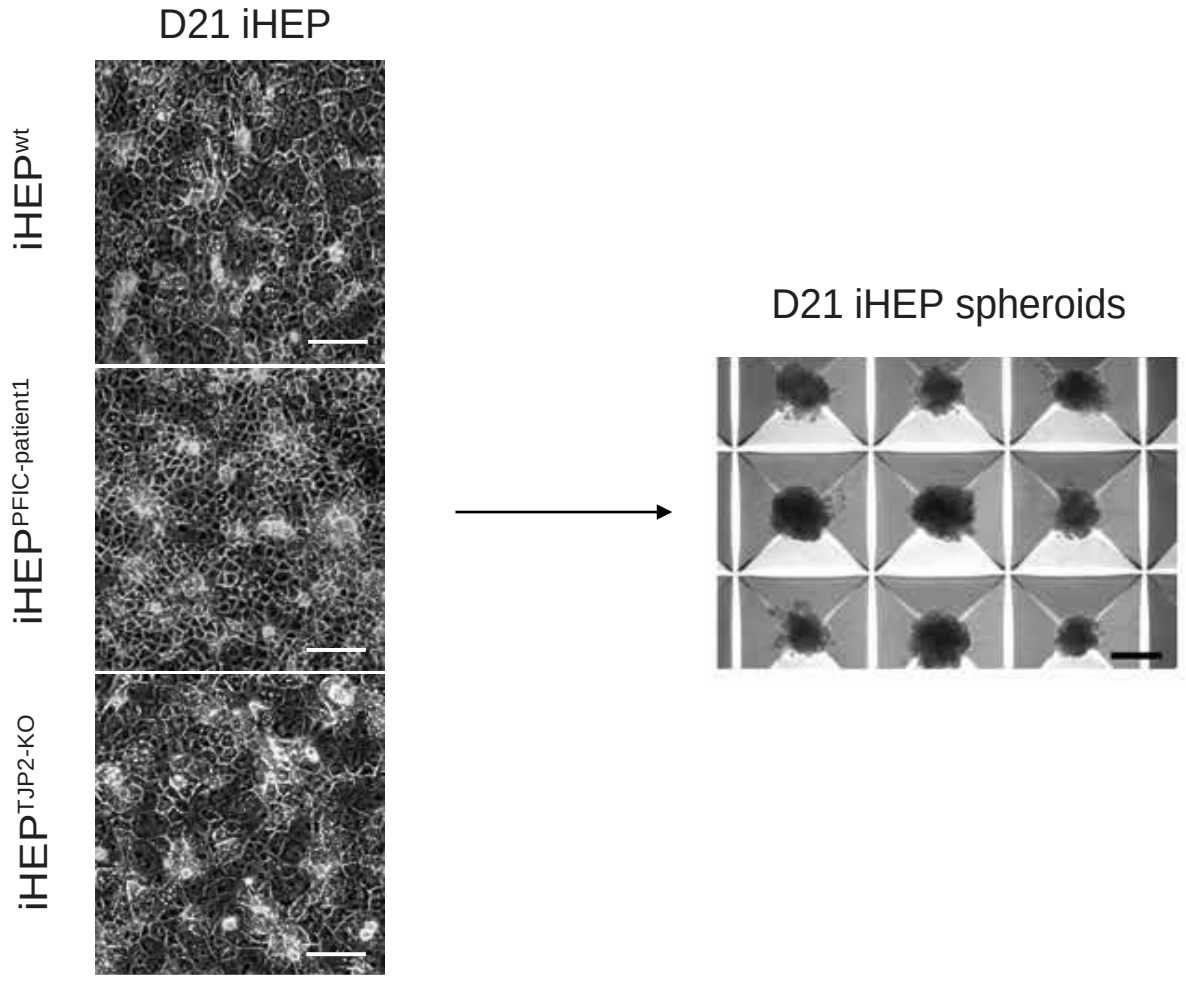



**Figure 2.4 Summary of making iHEP spheroid** (A) Schematic illustration of day 21 iHEPs from three independent iPSC lines lifted and placed onto Aggrewell-400$^{TM}$. After further maturation in the spheroid, day 34 iHEP spheroid were placed in a 96 well plate for functional and structural assessments. (B) Brightfield confocal images suggest day 21 iHEPs from three independent iPSC lines cultured in 2D condition had mature hepatocyte morphology. After iHEPs were placed into Aggrewell-400$^{TM}$ , the plates were centrifuged and brightfield confocal images revealed the spheroid had similar 'ball'-like structures. Scale bar, 100µm.

### 2.2.3 Isolation of human hepatocytes from human foetal liver

All human tissues were collected with informed consent following ethical and institutional guidelines. Foetal livers were obtained from the Human Developmental Biology Resource of University College London. Human foetal tissue was dissociated using Collagenase XI (sigma) enzymatic dissociation for 25 min at 37 °C with agitation. Foetal liver media was made up of 10% fetal bovine serum FBS (sigma), 1/100 pen/strep (sigma), 1/1,000 insulin (gibco),1/1,000 Dexamethasone (Bio-Techne) in dulbecco's modified eagle medium (DMEM). Isolated foetal hepatocytes were placed on 1/30 Matrigel-coated plates; the following day, after five washes with Dulbecco's Modified Eagle Medium, 6% Matrigel (diluted in foetal liver media) were loaded on top to form the Matrigel sandwich. Foetal hepatocytes were allowed to grow to day 6 to let the canaliculi develop in the platform.

### 2.2.4 Immunofluorescence

The iPSCs differentiated to hepatocyte-like cells to day 21, then were transferred onto collagen-coated plates and continued differentiating until day 30. Cells were washed with PBS three, times then fixed with 4% paraformaldehyde (PFA) for 10 min. After three washes with PBS, a blocking solution (Bovine Serum Albumin (BSA)1%+ Donkey serum (DS)3%+triton x 0.1%) was added and incubated for 20 min at room temperature. During blocking incubation, a primary antibody in PBS (1/200) was prepared, then the primary antibody solution was added to the well and incubated overnight. The next day, cells were washed three times with PBS. A secondary antibody was prepared in PBS with a dilution of 1 in 500. The mixture was incubated for 40 min at room temperature in darkness followed by three washes with PBS. Finally, DAPI (NUC BLUE fixed-cell life technology) was prepared in PBS, then imaged with Leica SP8 microscopy.

### 2.2.5 qPCR pluripotency gene mRNA quantification

Total RNA was extracted from cells using the RNeasy kit (Qiagen) following the manufacturer's instructions. After measuring total RNA concentration, 500 ng of RNA was subjected to RT reactions. The real-time PCR by the TaqMan probe system (gene expression master mix) and the QuantStudio system (ThermoFisher) quantified the mRNA of target genes, with specific primers and quantification protocol. After being normalised with a housekeeping gene (*RPL13A*), each

gene expression level was described relative to iHEP$^{wt}$ or baseline controls. All primers used were summarised in supplementary table S1.

### 2.2.6 CLS operetta based CDFDA functional assay

To evaluate the bile acid transport capacity and morphology of bile canaliculi in iHEP Matrigel sandwiches, CDFDA was supplemented into cells for 20 min. Accumulation of a fluorescent tracer (a metabolite of CDFDA) in bile canaliculi was then captured by CLS high content confocal imaging microscope (Operetta CSL) and analysed using ImageJ software. The images were first processed using global thresholding to filter out the noise and saturated signals. Next, size exclusion filtering was applied to remove imaging artefacts and to select biological relevant signals to determine the area, circularity, Feret's diameter, aspect ratio, roundness, and solidarity of the accumulated fluorescent tracer in bile canaliculi.

### 2.2.7 Leica confocal microscope CLF functional assay

To evaluate the bile acid transport capacity and morphology of bile canaliculi in iHEP spheroids, 5uM BSEP-specific substrate Cholyl-L-lysyl-fluorescein (CLF) Corning was supplemented into cells for 20 min. The fluorescent-labelled bile acid analogue CLF in bile canaliculi was then captured by Leica SP8 microscopy and analysed using ImageJ software. All images were taken according to the same imaging conditions for standardised comparison across three cell lines. The images were first processed using global thresholding to filter out the noise and saturated signals. Next, size exclusion filtering was applied to remove imaging artefacts and to select biologically relevant signals to determine the intensity of the 488 channel (CLF accumulated tracer in bile canaliculi).

### 2.2.8 Dead iHEP Matrigel sandwich quantification analysis

By definition, propidium iodide (PI) only stains dead cell DNA in the nucleus but not cytoplasm or elsewhere. I decided to create an imaging analysis pipeline on ImageJ to quantify the real dead cell signals in my cultures. To summarise, the process had four steps: 1) the PI signal was adjusted to only allow signals to appear on the nuclei; 2) PI+DAPI overlap images were used to select overlapping regions, which were defined as the regions of interest (ROI); 3) DAPI-only images

were then separated, and intensity threshold was used to remove the DAPI count from the ROI. The remaining analysed DAPI particles were the live-cell count; 4) intensity threshold was used to include all DAPI in the ROIs; analysed particles in this region were the dead-cell count. Finally, I calculated the percentages of cell death by dividing the number of dead cells by the number of total cells or the number of dead cells by the number of live cells. Calculations were performed across three cell lines with three independent repeats (three separate wells).

## 2.3 Results

### 2.3.1 Stepwise iHEP differentiation brightfield images

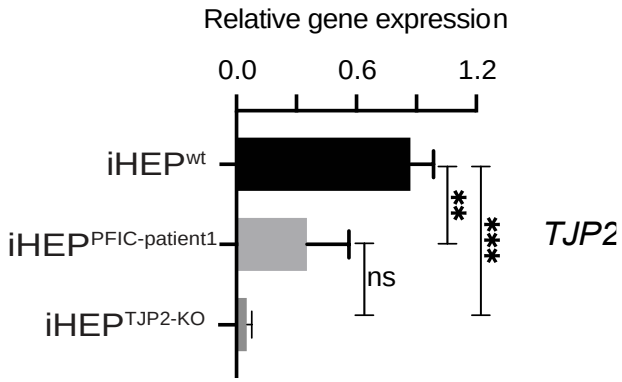
To study the impact of *TJP2* mutation on hepatocytes, I induced hepatic differentiation from all iPSC lines for the study of cholestasis, using the previously developed protocols that mimic the stages of human liver development (Blackford et al., 2019a). This three-step protocol followed the key stages of in vivo hepatocyte development. The first step involved turning hiPSCs into definitive endoderm cells. During the three iPSC lines' differentiations, the clear borders signifying stem cell colonies disappeared gradually, with the peripheral cells starting to spread out and migrate sporadically by day 1 post-differentiation. After massive cell expansion and proliferation, a confluent monolayer was formed by day 4. The culture medium kept the cells in a definitive endoderm state until day 9 when fully supplemented hepatozyme media was added. Brightfield images revealed that cells acquired dynamic morphology, mostly elongated. By day 14 of the differentiation, cells reached the hepatic endoderm stage, which acquired a cuboidal shape. Last, cells became hapatocyte-like, appearing polyhedral, on day 21 (Fig2.3b).

### 2.3.2 Mature hepatic gene qPCR and hepatic protein-immune fluorescence

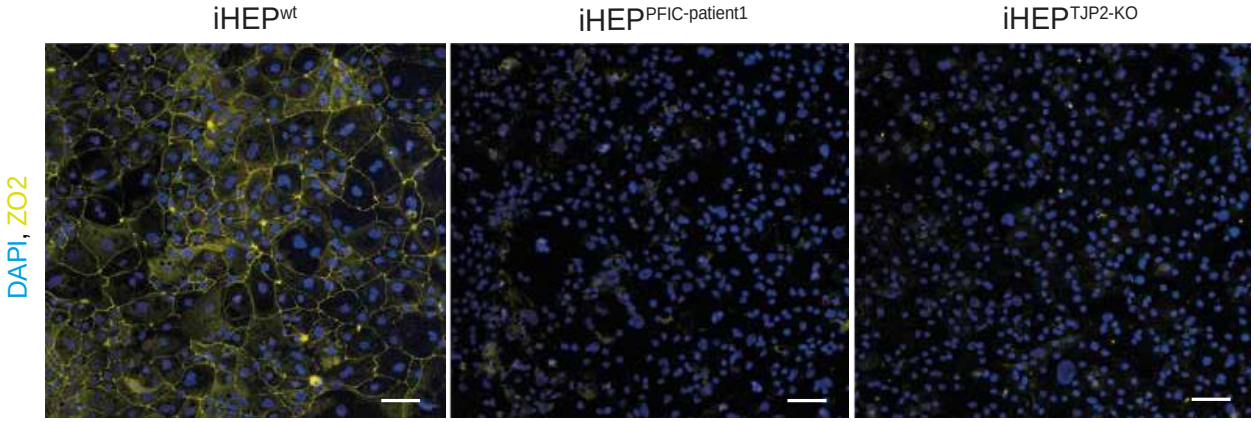
To validate *TJP2*'s gene expression and protein ZO2's translation from the new iHEP disease lines, *TJP2* gene expression of the iHEPs was compared by RT-PCR. $iHEP^{PFIC\text{-}patient\text{-}1}$ and $iHEP^{TJP2\text{-}KO}$ showed significantly lower *TJP2* expression compared to $iHEP^{wt}$ (Fig2.5a). Furthermore, 2D immunofluorescence data suggested ZO2 was expressed on the edge of day 34 $iHEP^{wt}$. In contrast, this expression was lost on day 34 in $iHEP^{PFIC\text{-}patient1}$ and $iHEP^{TJP2\text{-}KO}$

(Fig2.5b). iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ did not have any ZO2 protein expression, which was also observed in PFIC secondary to *TJP2* mutation patients' liver. Next, I checked the expression of the hepatocyte-specific markers across three day 34 iHEPs cultured in a Matrigel sandwich; all of them homogenously expressed *HNF4A* and *ALB* (Fig2.5c). Furthermore, the key hepatic genes' transcription profiles of iHEP were comparable among the iHEP cells, shown by assessing the hepatic-specific mRNA expression of *ALB*, asialoglycoprotein receptor 2 (*ASGR2*), *SERPINA1* and cytochrome 2E1 (*CYP2E1*); no significant differences were observed among the three iHEPs (Fig2.5d). However, gene expression results from foetal liver hepatocytes cultured in Matrigel sandwiches were higher than all iHEPs, suggesting iHEPs used in this study were relatively immature (Fig2.5d). Collectively, my data confirms that *TJP2* truncating mutations induced ZO2 deficiency in the induced hepatocytes, and all iPSC lines are capable of differentiating into iHEPs with comparable expression levels of hepatic differentiation markers

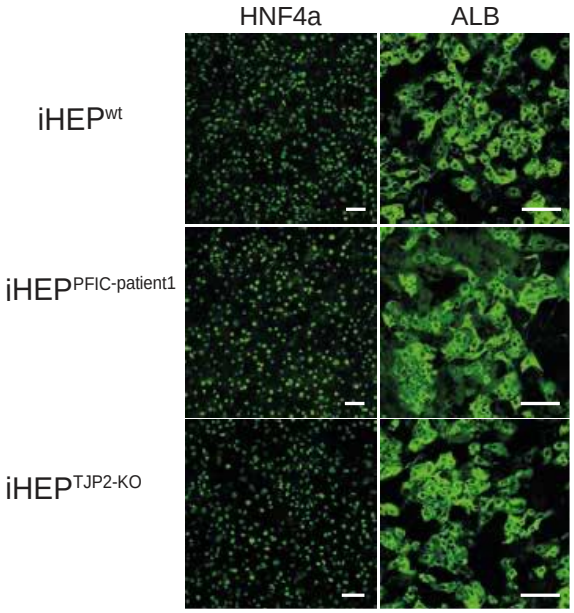
(A)




(B)



(C)


HNF4a
ALB
iHEPwt
iHEPPFIC-patient1
iHEPTJP2-KO

(D)

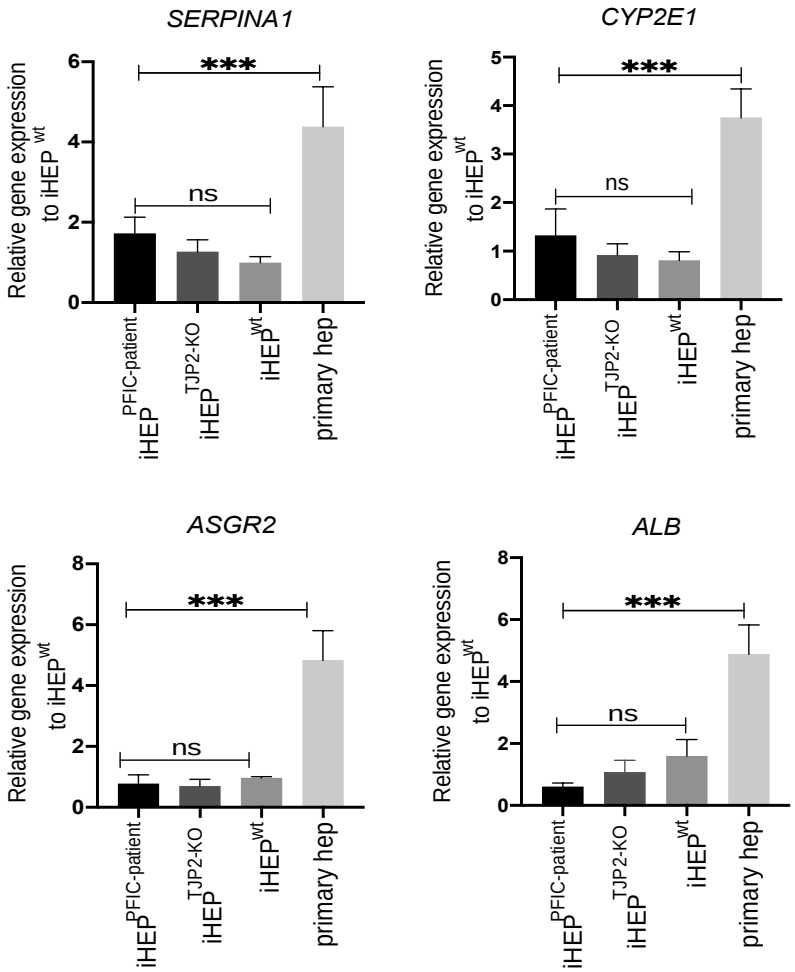

SERPINA1
CYP2E1
ASGR2
ALB
Relative gene expression to iHEPwt
***
ns
iHEPPFIC-patient
iHEPTJP2-KO
iHEPwt
primary hep

**Figure 2.5 Hepatocyte-like cell post differentiation quality analysis** (A) Quantitative gene expression analysis by RT-qPCR reveals that the *TJP2* is significantly downregulated in hepatocyte-like cells differentiated from $iPSC^{PFIC\text{-}patient1}$ and $iPSC^{TJP2\text{-}KO}$ compared to $iHEP^{wt}$. Mean± standard deviation (SD), n = 3, unpaired student t-test, **p < 0.01, ***p < 0.001, ns nonsignificant. (B) This diagram reveals ZO2 protein is absent in the lines with *TJP2* mutation ($iHEP^{TJP2\text{-}KO}$ and $iHEP^{PFIC\text{-}patient1}$), while the protein is preserved in $iHEP^{wt}$. Scale bar, 100 µm. (C) This diagram reveals immunofluorescent confocal images detecting the signature markers of hepatocytes (HNF4a and ALB), revealing a similar pattern to hepatic maturation in $iHEP^{wt}$, $iHEP^{PFIC\text{-}patient1}$ and $iHEP^{TJP2\text{-}KO}$ in the Matrigel sandwich system. Scale bar, 100 µm. (D) RT-qPCR gene expressions of four key hepatic genes (*ALB*, *ASGR2*, *SERPINA1* and *CYP2E1*) are comparable between $iHEP^{wt}$, $iHEP^{PFIC\text{-}patient1}$ and $iHEP^{TJP2\text{-}KO}$, with no significant differences detected by the one-way unpaired analysis of variance (ANOVA) analysis. Mean±SD, n = 3.

### 2.3.3 Hepatic polarity development in Matrigel sandwich

Proof of concept to demonstrate $iHEP^{wt}$ and foetal hepatocytes in the Matrigel sandwich culturing condition was the formation of elongated canaliculi-like structures, satisfying the three rules for bile canaliculi characterisation: first, condensation of actin (phalloidin staining) marked the presence of canaliculi; second, MRP2 was localised on top of the condensed actin; third, canaliculi were formed between two cells marked by DAPI, which stained for nucleus (Fig2.6). Day 34 $iHEP^{TJP2\text{-}KO}$ and $iHEP^{PFIC\text{-}patient1}$ formed a circle-like canaliculi-like structure that also satisfies the three characters of an *in vitro* canaliculi. To summarise, immunal fluoresence staining from all three cell lines' turned into polarised hepatocyte-like cells, and $iHEP^{wt}$ canaliculi were particularly similar to foetal liver cells' canaliculi. One structural difference between healthy hepatocytes ($iHEP^{wt}$, foetal iHEP) and the two diseased iHEPs ($iHEP^{TJP2\text{-}KO}$ and $iHEP^{PFIC\text{-}patient1}$) was the change from an elongated to a circular shape; such a circular canaliculi structure was also seen in rat hepatocytes with AMPK mutation cultured in Matrigel sandwich (Fu et al., 2010).

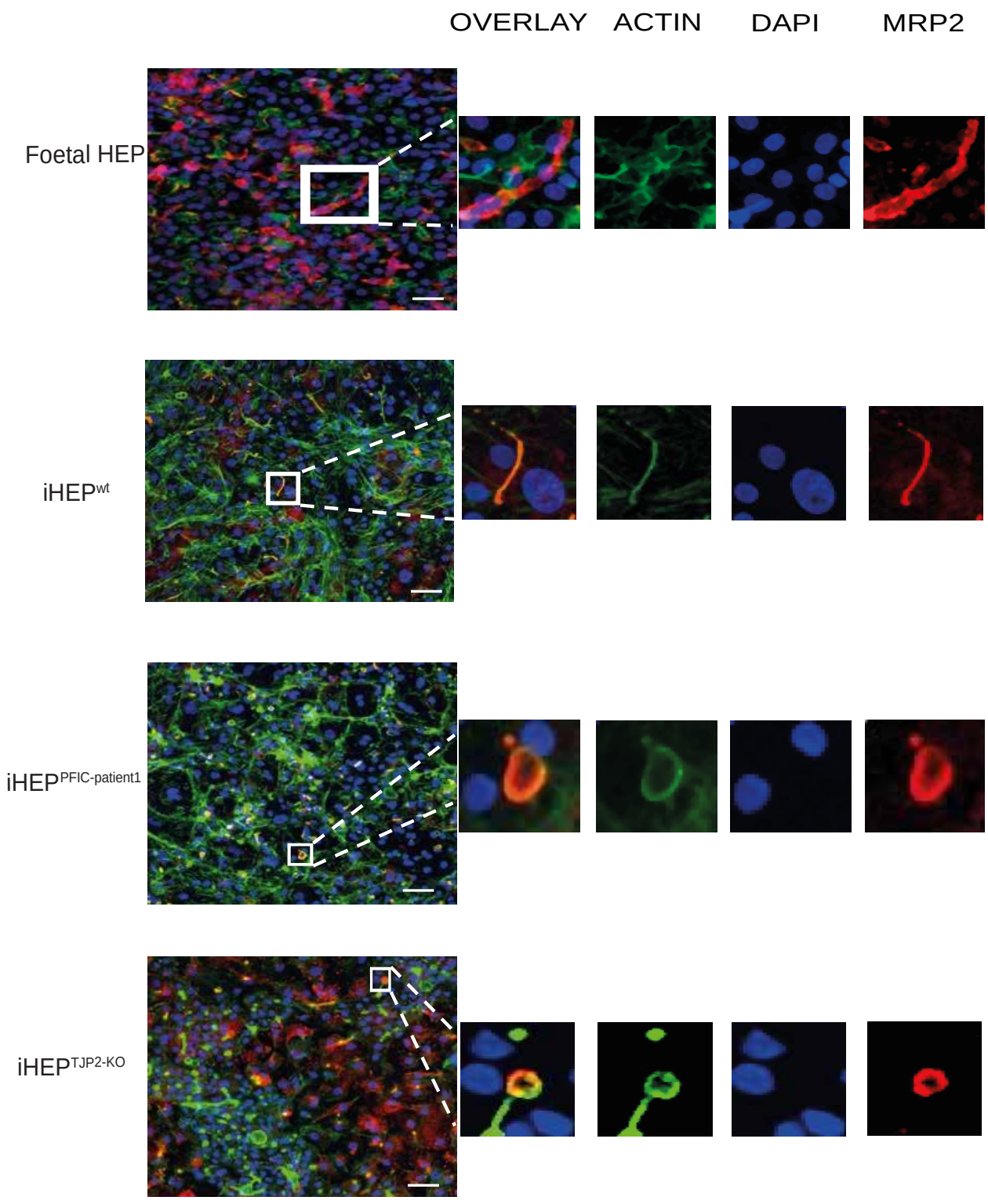
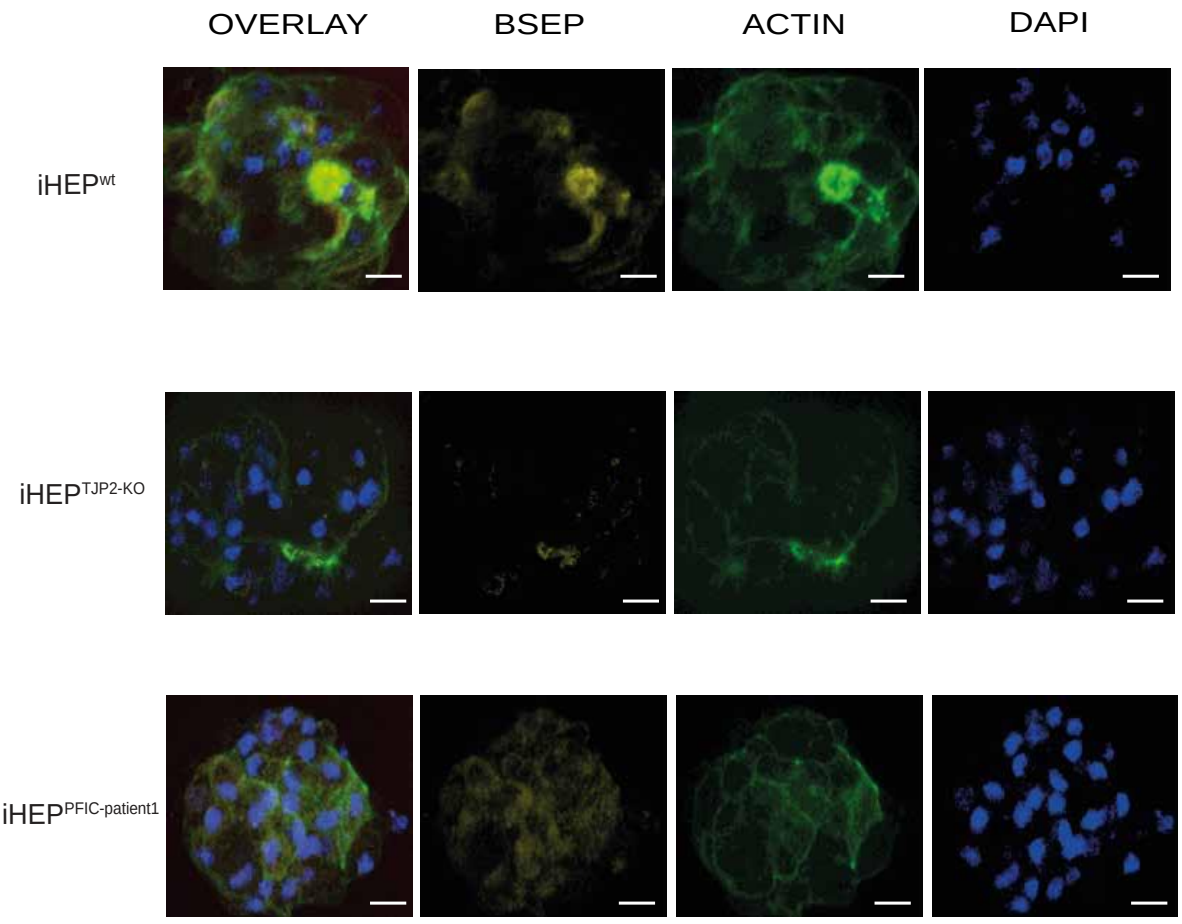

OVERLAY
ACTIN
DAPI
MRP2
Foetal HEP
iHEP^wt
iHEP^PFIC-patient1
iHEP^TJP2-KO

**Figure 2.6 iHEPs in Matrigel sandwich developed canaliculi-like structures** This diagram illustrates iHEPs from foetal liver cells, and $iHEP^{wt}$ in Matrigel sandwich forming chicken-wire-like canaliculi-like structures. Foetal hepatocytes form a canaliculi-like structure that is wider and more elongated compared to its $iHEP^{wt}$ counterparts. However, the canaliculi-like structures formed by $iHEP^{PFIC\text{-}patient1}$ and $iHEP^{TJP2\text{-}KO}$ share circular structures. All four iHEPs in the Matrigel sandwich form canaliculi-like structures marked by condensation of actin (green) with MRP2 (red) staining on top. In addition, a canaliculi-like structure is located between two nuclei (marked by DAPI blue staining). Scale bar, 100 µm.

## 2.3.4 Hepatic polarity and maturity of iHEP spheroids

The structural assessment data for iHEPs that formed spheroids were not straightforward to generate. Immune fluoresence images revealed apical membrane transporter BSEP co-localised with actin in $iHEP^{wt}$-formed spheroids, whereas such strong co-localisations were not seen in $iHEP^{PFIC\text{-}patient1}$ or $iHEP^{TJP2\text{-}KO}$-formed spheroids (Fig2.7a). In addition, although Z stack images were taken for all spheroids, DAPI stains were only partially captured. Even though BSEP appeared in $iHEP^{wt}$-formed spheroids, the numbers were far from adequate compared to existing hepatic spheroids (Hendriks et al., 2016c). In fact, only very few structures satisfied all three conditions mentioned previously which define canaliculi. This could be due to iHEPs' immature phenotype. Indeed, when I stained the cells with three key hepatic markers – Albumin, HNF4A, Cytokeratin18 (CK18) – to determine the cell maturity, I observed very little expression of the key hepatic markers, which indicated the spheroids' immaturity across all the iHEPs (Fig2.7b). Due to time limits, foetal liver hepatocyte spheroid data were not available for comparison

(A)


OVERLAY
BSEP
ACTIN
DAPI
iHEPwt
iHEPTJP2-KO
iHEPPFIC-patient1

(B)

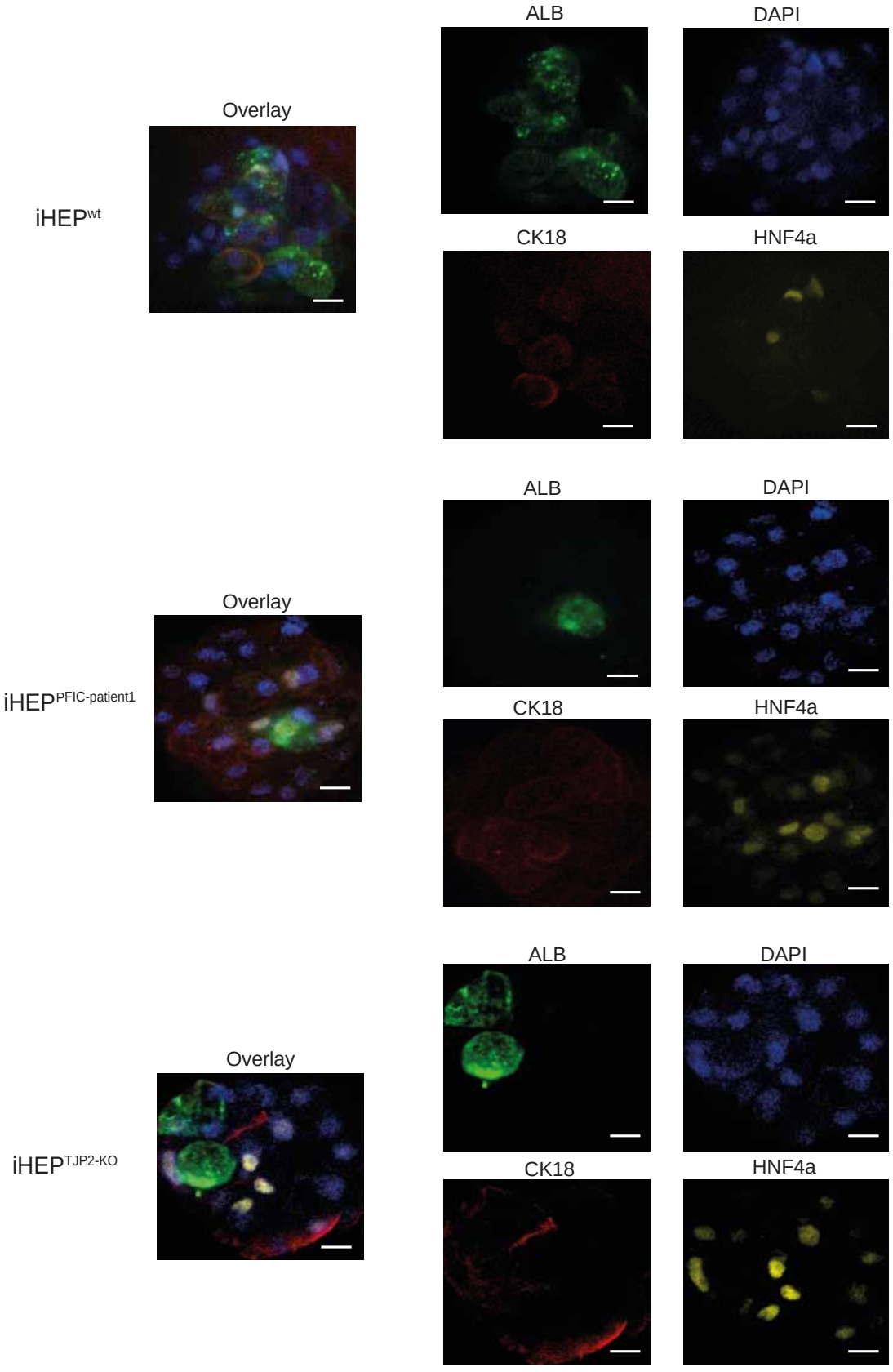
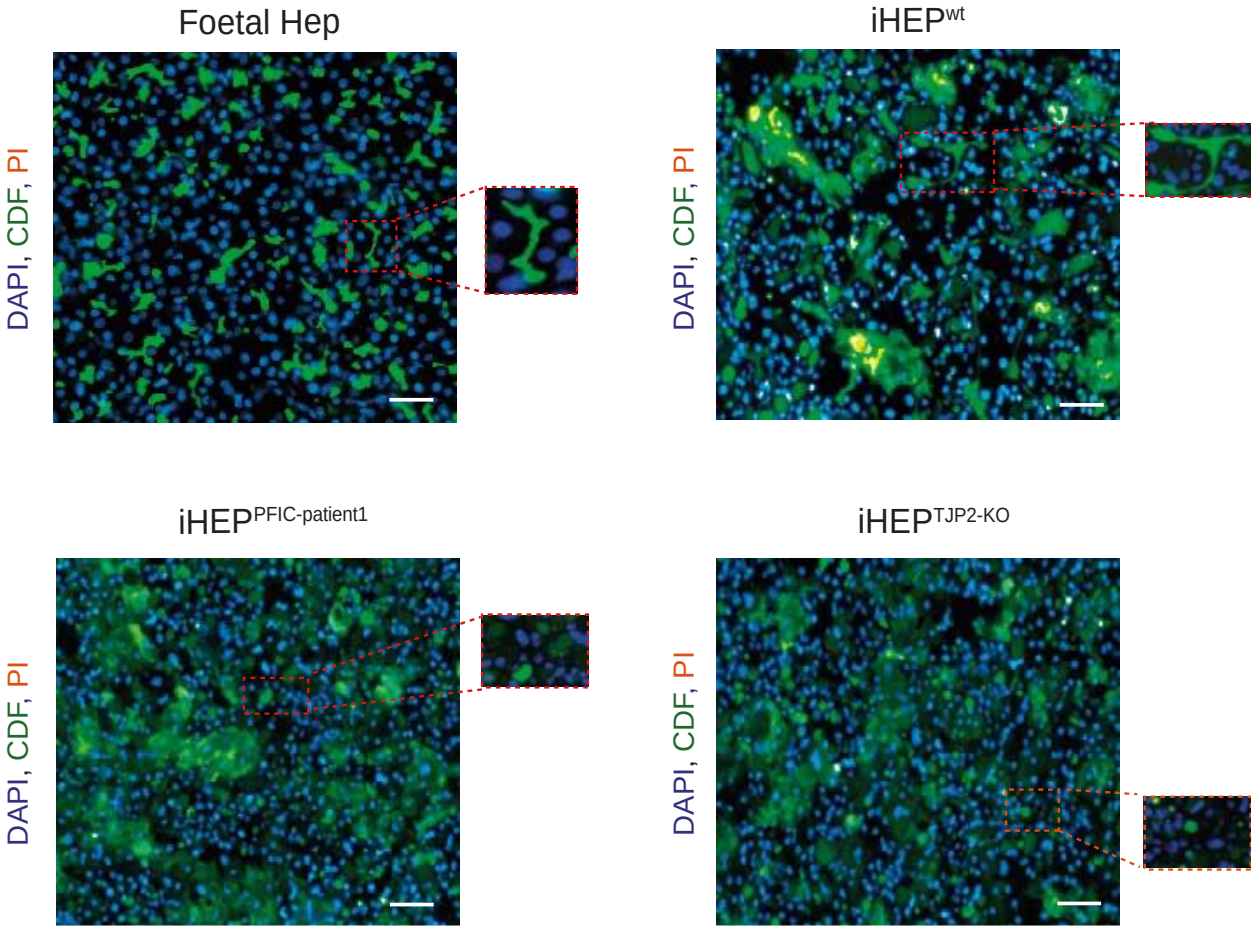

ALB
DAPI
Overlay
iHEP^wt
CK18
HNF4a
ALB
DAPI
Overlay
iHEP^PFIC-patient1
CK18
HNF4a
ALB
DAPI
Overlay
iHEP^TJP2-KO
CK18
HNF4a

**Figure 2.7 iHEP spheroids developed canaliculi-like structures (**A) This diagram illustrates iHEPwt cultures in spheroid-formed circular canaliculi-like structures, which are marked by BSEP and actin co-localisation. In $iHEP^{PFIC\text{-}patient1}$ and $iHEP^{TJP2\text{-}KO}$ spheroids, no such canaliculi-like structures appeared. Scale bar, 100 µm. (B) Selected hepatic markers, immunofluorescence images, of iHEP spheroids. All hepatic markers appeared low across the three iHEPs. More specifically, $iHEP^{wt}$ and $iHEP^{PFIC\text{-}patient1}$ appeared to have some albumin expression but very little HNF4A and CK18 expression. $iHEP^{TJP2\text{-}KO}$ appeared to have some strong local albumin, HNF4A and CK18 expression. Scale bar, 100 µm.

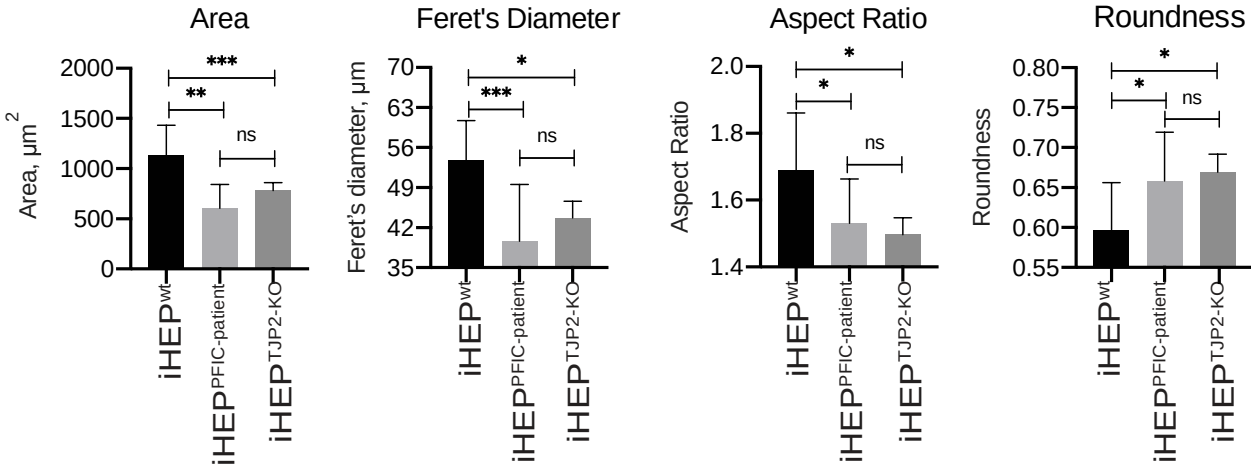
### 2.3.5 Healthy and diseased iHEPs' CDFDA functional phenotype readout

To determine whether these disrupted canaliculi structures compromised the bile acid transport in *TJP2*-deficient hepatocytes, I used CDFDA, a fluorescent tracer of a hepatic apical transporter (MRP2), to visualise the extension of bile canaliculi in Matrigel-sandwich cultured iHEPs. In $iHEP^{wt}$ and foetal hepatocytes, the active transport and accumulation of the fluorescent tracer were detected along the border of adjacent cells and formed elongating networks of bile canaliculi, suggesting the presence of functional MRP2 in the apical membrane (Fig2.8a). The morphology of the bile canaliculi traced by accumulation was significantly different in *TJP2*-deficient iHEP. The bile canalicular network resembling a chicken-wire-like structure in the healthy hepatocytes ($iHEP^{wt}$ and Foetal iHEP) was lost in the diseased hepatocytes ($iHEP^{PFIC\text{-}patient1}$ and $iHEP^{TJP2\text{-}KO}$). Instead, the small and isolated spherical structures were revealed, as shown in the inlets. I then tried to use MRP2/BSEP-specific inhibitor sitaxsentan sodium (SS) to reverse the 'chicken-wire'-like canaliculi phenotype from $iHEP^{wt}$ and foetal iHEP cultured in the Matrigel sandwich system, because in theory, this active 'CDF' transportation should be stopped once MRP2/BSEP proteins were inhibited. Indeed, the 'chicken-wire'-like canaliculi structures massively decreased when SS was added to the $iHEP^{wt}$ and foetal iHEP culture (Sup Fig1A). Another point worth mentioning is that foetal hepatocytes in the Matrigel sandwich culture were far more mature than their $iHEP^{wt}$ counterpart. Therefore, when quantifying the bile canaliculi total area and canaliculi percentage areas between the two cell types, foetal hepatocytes outperformed $iHEP^{wt}$. In addition, within each of the quantifying parameters, there were more significant differences observed between the SS-treated and nontreated conditions in foetal hepatocytes than $iHEP^{wt}$ (Sup Fig1B).

**(A)**


Foetal Hep
iHEPwt
iHEPPFIC-patient1
iHEPTJP2-KO
DAPI, CDF, PI

(B)


Area
Area, µm2
2000
1500
1000
500
0
***
**
ns
Feret's Diameter
Feret's diameter, µm
70
63
56
49
42
35
*
***
ns
Aspect Ratio
2.0
1.8
1.6
1.4
*
*
ns
Roundness
0.80
0.75
0.70
0.65
0.60
0.55
*
*
ns
iHEPwt
iHEPPFIC-patient
iHEPTJP2-KO

(C)

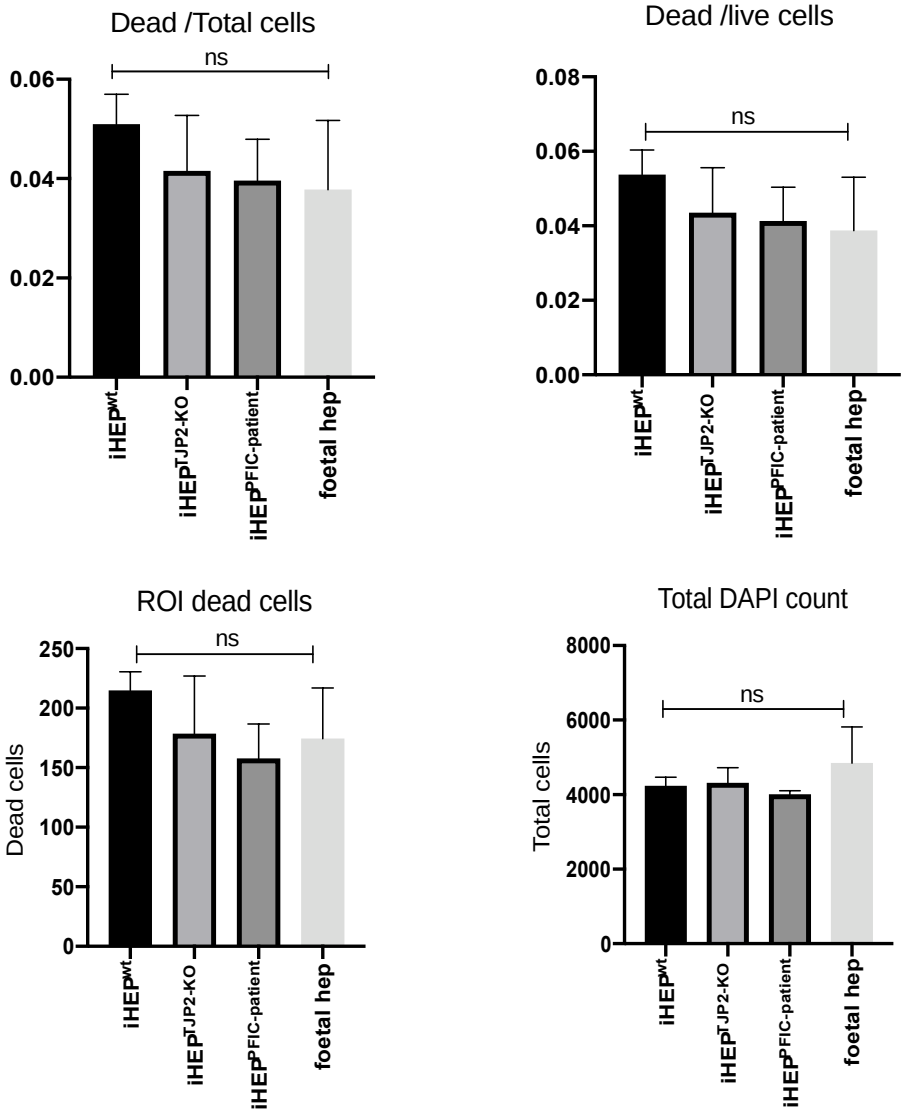



**Figure 2.8 Imaging and quantification of Matrigel sandwich iHEPs' canaliculi-like structures** (A) Representative confocal micrographs of iHEP$^{wt}$, iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ were taken by a low/medium-throughput imaging system, Operetta CLSTM, which shows the transport and accumulation of a fluorescent tracer (metabolites of CDFDA, green), representing the function and morphology of bile canaliculi of cells cultured in the Matrigel sandwich system. Dead cells were stained by PI (yellow), and nuclei were stained by DAPI (blue) to assess the viability of the cells. iHEP$^{wt}$ displayed a chicken-wire-like network of bile canaliculi, while iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ displayed small spherical bile canalicular structures or droplets of accumulating fluorescent tracer. Scale bar, 100 µm. (B) Morphological quantification of bile canalicular network and structure observed in CDFDA assays using the ImageJ processing algorithm. The pattern of fluorescent tracers in iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ show significantly smaller and rounder structures compared to those in iHEP$^{wt}$. Mean±SD, n = 9, unpaired student t-test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ and ns: nonsignificant. (C) PI quantification of dead cells

in all used images from this study. Quantification used the ImageJ processing algorithm. One-way unpaired ANOVA statistical analysis was performed. Mean±SD, n = 3. iHEP$^{wt}$, iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ showed no significant difference in total cell counts, dead/total cells, dead/live cell ratios or ROI dead cell count.

### 2.3.6 Healthy and diseased canaliculi quantification and viability assessments of iHEPs in Matrigel sandwich

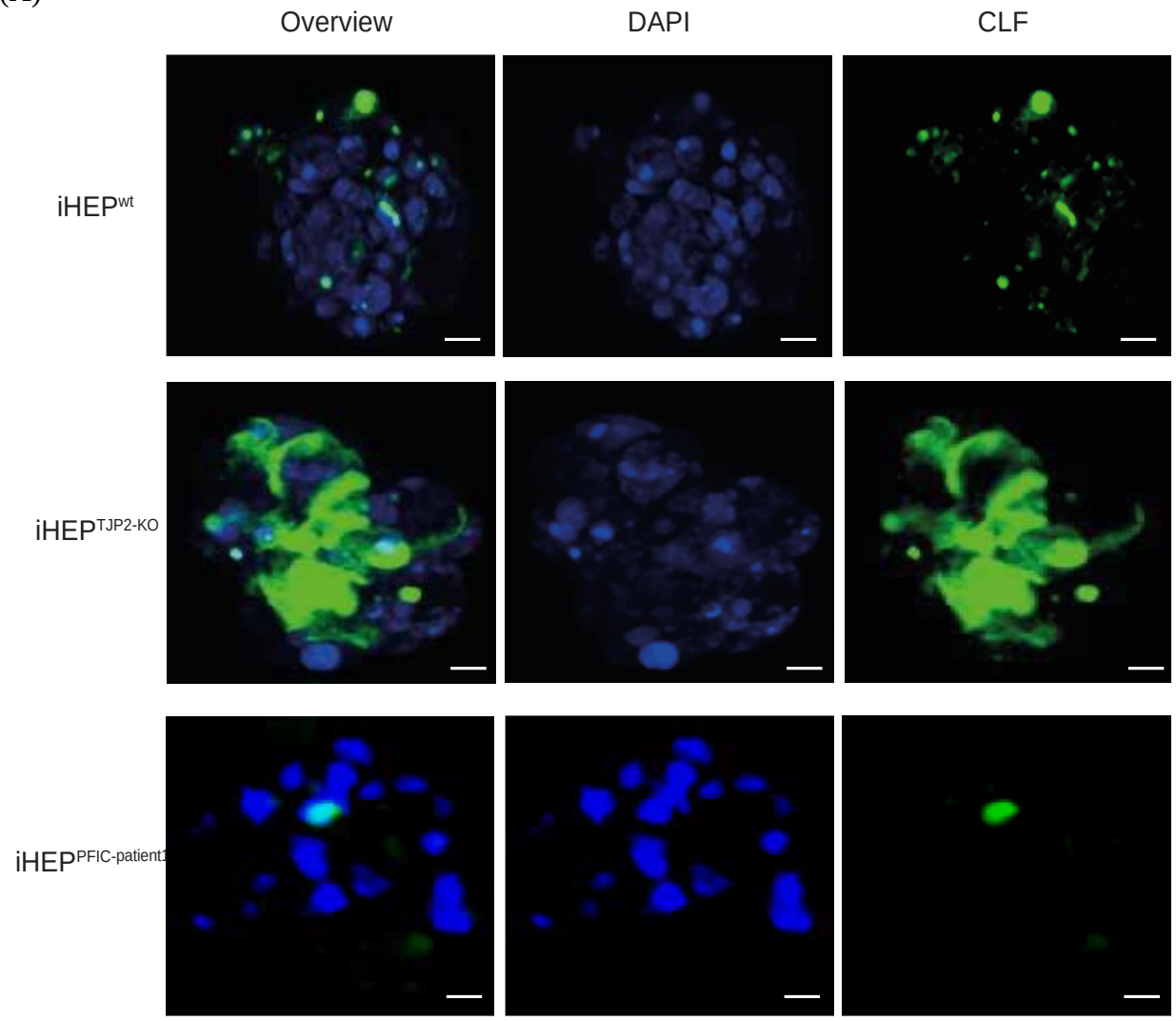
To quantify the bile canalicular formation by morphometric analysis, I developed an image-processing assay to evaluate the transport and accumulation of the tracer captured by a low/medium-throughput confocal microscopy system. The image analysis revealed that both iHEP$^{TJP2\text{-}KO}$ and iHEP$^{PFIC\text{-}patient1}$ lacked the interconnected network formed by the elongated bile canaliculi in iHEP$^{wt}$; instead, they contained isolated spherical bodies of bile canaliculi. The morphometric analysis of bile canaliculi in ZO2-deficient iHEPs revealed that they were smaller (Area), branched less (Feret's diameter), less elongated (Aspect ratio) and more bloated in shape (roundness) (Fig2.8b). These small bile-containing spheres might be the inclusion structures with apical membranes, found by TEM, representing the defective bile canaliculi of ZO2-deficient iHEPs due to the loss of cytoskeletal support needed to maintain the structure (Rao and Samak, 2013). These parameters were quantified by the imaging algorithm to enable us to determine the level of bile canalicular disruption caused by the ZO2 deficiency. Collectively, iHEP$^{TJP2\text{-}KO}$ and iHEP$^{PFIC\text{-}patient1}$ showed the lack of ZO2 expression and significant structural distortion on the cellular polarity and the bile canalicular formation compared to the healthy iHEP$^{wt}$. This coincides with the disrupted localisation of functional transporters, suggesting that ZO2-deficient iHEPs recapitulate the pathology occurring in the liver of PFIC patients with *TJP2* mutation. In addition, dead iHEPs were selected according to the ROIs described in the method section. The total cell counts for all wells were consistent across four cell lines, and three viability assessment (dead/total cells, dead/live cells, ROI dead cells) results demonstrated that there were very few dead cells present in the culture. More importantly, the dead cell numbers were consistent for all three iHEPs (Fig2.8c).

### 2.3.7 iHEP spheroids' CLF-functional phenotype readout and assessments

One of the commonly used industrial bile acids for *in vitro* cholestasis models is CLF. This is a fluorescein-labelled bile acid whose biological behaviour closely resembles that of naturally occurring cholyl glycine. CLF is picked up by the basolateral bile salt transporters, NTCP and, to a lesser extent, by the Organic-Anion-Transporting Polypeptides (OATPs). CLF can then be transported across the canalicular membrane by the BSEP and accumulate in the bile canalicular space. Measuring accumulation of CLF in bile canaliculi transported from hepatocytes using fluorescent microscopy partially mimics measuring human bile transportation. Thus, it is likely to yield a cholestatic phenotype in diseased iHEPs.

In $iHEP^{wt}$, the active transport and accumulation of the fluorescent tracer was detected between cells. However, this structure did not form elongating networks of bile canaliculi (Fig2.9a). Instead, intracellular accumulation of the fluorescent tracer was detected in $iHEP^{TJP2\text{-}KO}$ spheroids. This was highly similar to PFIC with TJP2 deficiency patients' liver disease phenotype (Fig2.9a) (Sambrotta et al., 2014b). Following the observations from $iHEP^{TJP2\text{-}KO}$ spheroids, I wondered if $iHEP^{PFIC\text{-}patient1}$ spheroid would have the same CLF intracellular accumulation phenotype. To my surprise, there was no such large CLF retention observed in the $iHEP^{PFIC\text{-}patient1}$ spheroid (Fig2.9a). I then quantified the CLF (488 channel) intensity from each image by using ImageJ software. The image analysis revealed that $iHEP^{TJP2\text{-}KO}$ had significantly higher CLF accumulation compared to $iHEP^{wt}$ and $iHEP^{PFIC\text{-}patient1}$ spheroids. More specifically, the CLF stainings in $iHEP^{TJP2\text{-}KO}$ spheroids were bigger (average size), more prevalent (CLF % area) and more widespread (total CLF area) compared to $iHEP^{wt}$ and $iHEP^{PFIC\text{-}patient1}$ spheroids (Fig2.9b). Collectively, CLF stainings suggested there are few canaliculi-like structures formed in the $iHEP^{wt}$ and $iHEP^{PFIC\text{-}patient1}$ spheroids, and there are severe CLF accumulations in $iHEP^{TJP2\text{-}KO}$.

(A)



(B)

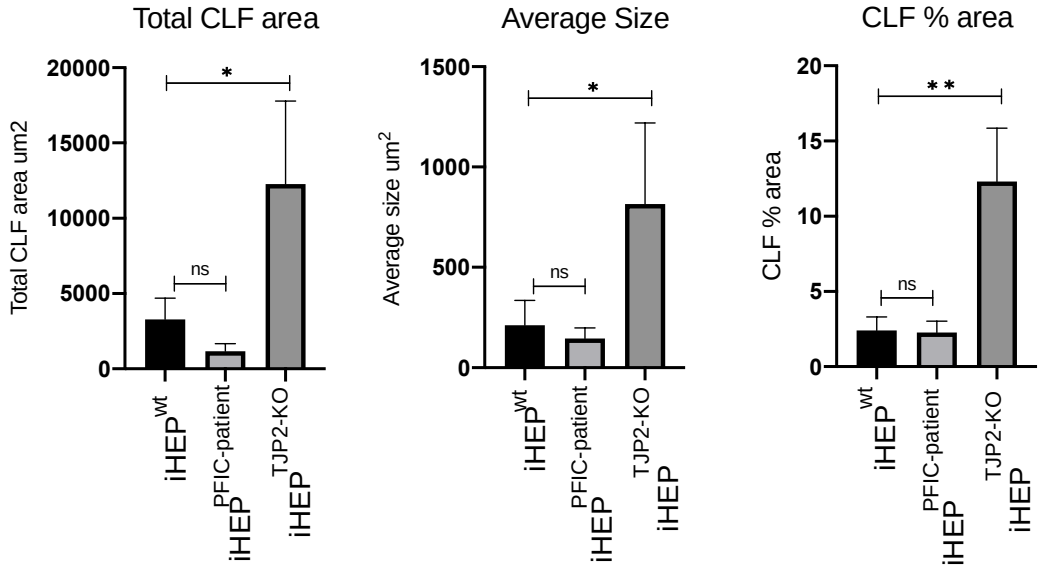



**Figure 2.9 Imaging and quantification of spheroid iHEPs' canaliculi-like structures** (A) Representative confocal micrographs of iHEP$^{wt}$, iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ were taken by a high content imaging system, Leica SP8, to show the transport and accumulation of a fluorescent tracer (metabolites of CLF, green) representing the function and morphology of bile canaliculi of cells cultured in a spheroid system. Nuclei are stained by DAPI (blue) to mark the position of nuclei. iHEP$^{wt}$ and iHEP$^{PFIC\text{-}patient1}$ displayed very few canaliculi-like structures and little CLF accumulation, while iHEP$^{TJP2\text{-}KO}$ displayed severe fluorescent tracer accumulation. Scale bar, 100 µm. (B) CLF-staining quantification of bile canalicular network and structure observed in CLF assays using the ImageJ processing algorithm. The pattern of fluorescent tracers in iHEP$^{TJP2\text{-}KO}$ showed significantly higher values in 'total CLF area', 'average size' and 'CLF % area' compared to those in iHEP$^{wt}$ and iHEP$^{PFIC\text{-}patient1}$. Mean±SD, n = 3, one-way unpaired ANOVA, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ and ns: nonsignificant.

## 2.4 Discussion and conclusion

In this chapter, I tried to create two independent PFIC with *TJP2* mutation disease models which could be used for drug screening and mechanistic study. There were two criteria vital for the effective use of hiPSCs and their derivatives for disease modelling. These were 1) the adequate exhibition of the required properties of differentiation and maturation that were comparable to the corresponding adult cells and 2) the ability to accurately recapitulate the main characteristics of the disease itself (Hannoun et al., 2016). First, I differentiated iPSC$^{wt}$, iPSC$^{TJP2\text{-}KO}$ and iPSC$^{PFIC\text{-}patient1}$ to hepatocyte-like cells independently. The cells were relatively immature compared to foetal liver cells due to iHEP's lower hepatic gene expression. Luckily, the PFIC with *TJP2* mutation patients were all relatively young; therefore, the immature hepatic phenotypes of iHEPs used in this study might have been at an advantage in modelling the disease (Sambrotta et al., 2014b). However, improvements in hepatocyte maturity would help scientists to better mimic the human in vivo liver environment. Despite the foetal liver cells forming canaliculi in the Matrigel sandwich, they shared similar morphology with their iHEP$^{wt}$ counterparts; the canaliculi distribution was still different between the two cell types.

Other than cells, the hepatic niche was also important to establish an efficient disease model. As I mentioned in the introduction, the biggest problem for the sandwich-cultured hepatocyte model was the lack of countercurrent media flow and the absence of niche cells. In addition, plane Matrigel sandwich-cultured cells limiting multidimesional cell–cell contact and lacking regional oxygen gradient, nutrients and temperature control, all restricted iHEPs' potentials to become fully mature. Spheroid-based cholestasis disease models improved the cell–cell contact in 3D, but it had disadvantages, such as centre necrosis and lack of ECM interactions. Recent technologies, such as organs on a chip and 3D bioprinting, could help to overcome these problems, and therefore, future iPSC differentiation based on such devices could yield more mature hepatocytes (Esch et al., 2014; Messner et al., 2013). Second, iHEP$^{TJP2\text{-}KO}$ and iHEP$^{PFIC\text{-}patient1}$ in Matrigel sandwich culture successfully recapitulated PFIC with *TJP2* mutation patients' defective canaliculi structural phenotype, and more importantly, the healthy 'chicken-wire' phenotypes from iHEP$^{wt}$ and foetal hepatocytes were reversed when inhibitors were added to the culture. This suggests that the sandwich system was competent in mimicking human bile transportation physiology partially;

also, it could recapitulate some defective canaliculi phenotypes from PFIC4 patients' derived iHEPs.

Although the key patient functional disease phenotype intrahepatic bile acid accumulation had not been observed from the diseased iHEPs in the Matrigel system, the change of canaliculi structure at least indicated the difference between the three iHEPs in the modelling system. From the spheroid model results, we observed very few canaliculi-like structures formed in the $iHEP^{wt}$ and $iHEP^{PFIC\text{-}patient1}$ spheroids, as marked by little CLF staining. However, there was severe CLF accumulation observed in $iHEP^{TJP2\text{-}KO}$ spheroids compared to their healthy counterparts. The intracellular accumulation of bile acid was the key pathological phenotype in PFIC4 patients. The little CLF staining in $iHEP^{wt}$ and $iHEP^{PFIC\text{-}patient1}$ spheroids could be explained by lack of OATP basolateral membrane transporters or bile canaliculi leakage. OATP protein-immune fluorescent staining and leakage assays could be used to test this hypothesis. In addition, the CLF accumulation in $iHEP^{TJP2\text{-}KO}$ spheroids was likely due to a lack of BSEP expression or function on hepatocyte apical membranes. BSEP immune fluorescent staining in $iHEP^{TJP2\text{-}KO}$ spheroids will further help explain the mechanism behind this CLF accumulation phenotype. Future experiments should also include adding BSEP inhibitors to the $iHEP^{wt}$ spheroid system to find if CLF accumulation could be observed in $iHEP^{wt}$ due to defective BSEP transportation. Most importantly, future drug screening experiments should include currently existing drugs/compounds from a pharmaceutical company's library to test both modelling assays' efficiency and quality when screening for potential drugs to treat PFIC4 patients.

The main functions of the tight junction at the bile canaliculi were to form a bile-blood barrier and to separate bile from the blood (Jou et al., 1998). It has been postulated that the bile-blood barrier in the liver of patients with *TJP2* deficiency is compromised, thus leading to progressive cholestasis (Sambrotta et al., 2014b). Further analysis, such as transepithelial electrical resistance (TEER) and trichloroacetic acid (TCA) leakage assay can test the canalicular barrier functions of diseased and normal iHEPs in future studies. There were some canaliculi quantification analysis readout variations within each iHEP. This could be a natural representation of the heterogeneity of canaliculi shapes formed around the iHEPs. This heterogeneity was a real limit for low/medium-throughput drug screening image quantification, as it required a clear difference between healthy

and diseased lines (ideally, the difference should be three standard deviations). There are two ways to solve this canaliculi heterogeneity problem within each cell line: the first way is to work on the biology of canaliculi formation and produce more mature hepatocytes; the second way is to work on the technology. This is where software, such as phenologic, can be used to capture the canaliculi more accurately by 'training' the computer to pick up the correct shape according to human's order until the software can do what the human eye did or even better.

Human iPSCs derived hepatocyte has been used to study genetic liver diseases such as wilson's disease (WD) involving hepatocyte polarity (Overeem et al., 2019). Consider their iHEP functional assay as a platform , the iHEPs in matrigel sandwich assay platform demostrated hepatic markers like albumin and HNF4a. *In vivo*-like branching canalicular network flanked by tight junctions was also developed in this platform. Moreover, a polarized distribution of bile canalicular membrane proteins at the apical surface domain were established. Bile canalicular efflux pump substrates were able to transport to the canaliculi lumen from the hepatocytes. Finally the platform display regulated polarized trafficking as exemplified by the copper-stimulated redistribution of ATP7B to the canalicular domain (Overeem et al., 2019). However this is an end point assay, with no living iHEPs, and it was not integrated to low/medium-throughput imaging machine like Operetta. More importantly, their biological endpoint analysis only included relative fluorescence intensity of ATP7B and ABCC2 co-localization. More biological endpoints were needed to differentiate between healthy and WD patients' iHEPs' characteristics, scientists could then use the most clear biological endpoints for drug efficacy testing. All drawbacks will limit their ability to become to top-notch future personalised preclinical in vitro model systems for the evaluation of drug efficacy for WD patients.

Functional assays and the whole drug discovery ecosystem start to move towards "industry 4.0" where everything will be automised, digitized, and eventually become intelligent (Arden et al., 2021). It is my responsibility as a functional assay builder start to design the assays in a real-time, living cells, patient specific, low/medium-throughput, atomised (assay+ analysis), digitized manner, hence paves the road for future personalised drug efficacy screening platform's success. The platform I generated from this chapter's work already posseses some of these key characters, for example the platform is based on living iHEPs and be able to monitor the bile transportation

process in real-time. More importantly, it is compatible with the powerful low/medium-throughput imaging machine CLS oppereta and hence all biological end points can be analysed by machine learning software such as phenologic. Therefore, my platform is easier to transform to match the "industry 4.0" standard, but remain challenges such as the low maturity of iHEPs and lack of blood and bile flow.

Moreover, the ultimate goal of building functional assay is to study the mechanism of the disease and eventrually yield novel theraputic agents for PFIC patients. I quote the founder of Vertex pharmaceutical Joshua Boger's words: "the cell is like a busy running city, there are things going on in every corner of the city, the only way to understand more is to send reporters to these places and ideally report in a live format, the more information one has, the better the drug one will design" (https://cystic-fibrosis.com/living/breath-from-salt-review). Complex diseases such as PFIC or drug induced liver injury (DILI) are difficult to study due to their complxed nature and causes. By having a low/medium throughput real-time patient specific *in vitro* platform with clear disease biological end points, scientists will be able to access the more information from the diseased cells with higher quality, and hence increase the chance of discovering novel theraputic targets.

Overall, the iHEP Matrigel sandwich imaging-based low/medium-throughput platform provides new oppertunities for PFIC with *TJP2* mutation drug screening and mechanism study, which has translational values for industry and academia.

# Chapter 3: Using *in silico* analysis to link DILI SNPs with CREs of bile salt-metabolising and transporting genes

# 3.1 Introduction

## 3.1.1 GWAS and DILI

Genetic variants from coding and noncoding regions' contributions to human diseases have long been considered important causing factors (Lincke et al., 1991; Maurano et al., 2012). However, it was not until sequencing costs decreased in the early 2000s that researchers started to use sequencing-based approaches in disease diagnosis and to systematically link genetic sequence variations with certain disease phenotypes. For example, PFIC with *TJP2* mutation was discovered by using targeted sequencing and whole-exome sequencing (WES) analyses of cholestasis patients (Sambrotta et al., 2014c). In addition, more complex diseases, such as DILI, normally have multiple variants across many different genes. Therefore, it was the initiative of GWAS to collect all the variants which are associated with certain complex diseases. For DILI, there were some mutations reported from coding regions and more mutations from the noncoding regions (Koido et al., 2020). Previously, when scientists knew little about the noncoding regions, they were classified as 'junk DNA'. With improved understanding and technology development, more functional DNA elements, such as CREs were identified in noncoding regions, which regulate gene expression. Indeed, Maurano et al suggest that over half of the common disease-associated variants are concentrated in CREs that are active in disease-relevant tissues (Maurano et al., 2012).

There are two major challenges in noncoding variant human disease research. First, gene expression regulation by noncoding variants is highly complicated. Variants located in CREs are able to interact with multiple gene regulatory elements such as promoters and enhancers in proximal or distal regions. Second, even when the noncoding variants' gene regulatory mechanism is revealed, there are few relevant *in vitro* functional assays to help establish the link between the genetic variants and the patients' disease phenotype. Thus far, most noncoding DILI genetic risk variants have not been functionally interpreted nor have their disease-causing mechanisms been resolved. Therefore, it is this study's purpose to use bioinformatic tools and publicly available databases to systematically identify DILI risk variants located in CREs in order to further understand how DILI variants contribute to the disease phenotype.

Before GWAS became popular, genetic variants used to be identified in case-control candidate gene association studies. This method relied on the selection of candidate genes based on existing mechanistic knowledge of DILI, e.g. those involved in metabolic, immunological and mitochondrial pathways (Chalasani and Björnsson, 2010). For example, null mutations in the metabolic genes involved in phase I bioactivation of drugs, e.g. *CYP2E1* mutant C2/C2 allele, has been associated with reduced liver toxicity, notably through reduced *CYP2E1* activity and, hence, less accumulation of hepatotoxins, e.g. in the case of isoniazid (Huang et al., 2003). While the candidate gene association studies approach could successfully identify variants in the pathways known as DILI pathogenesis, it runs the risk of ignoring genetic variants of unknown mechanisms (Daly and Day, 2001). However, GWAS uses a low/medium-throughput NGS-based approach to identify variants associated with certain diseases, which allows GWAS to examine the patients' genome more comprehensively and faster compared to a sanger sequencing-based approach. The most well-studied GWAS identified that DILI-relevant risk variants were in human leukocyte antigen (*HLA*) class I and II alleles that encoded major histocompatibility complex-1 (*MHC*-I) and II molecules, respectively (Raúl et al., 2004). There are over 15 clinically used drugs that show increased susceptibility to DILI in the presence of specific *HLA* haplotypes. For example, the major histocompatibility complex, class2, DR Beta1 (*HLA-DRB1*)∗15:01 class II haplotype, common in European populations is a well-established risk factor for increased susceptibility to amoxicillin-clavulanic hepatotoxicity (Stephens et al., 2013). In addition, DILI risk variants have also been identified near other immune-relevant genes and near genes related to drug metabolism, but the mechanisms behind the association of specific risk variants of DILI remain largely unknown.

### 3.1.2 ENCODE consortium

Approximately 2% of the human genome encodes proteins; the remaining 98% of the human genome's function remains relatively unknown. The goal of ENCODE is to build a comprehensive

list of functional elements in the human genome, including elements that act at the protein and RNA levels and regulatory elements that control cells and circumstances in which a gene is active (Abascal et al., 2020a). The ENCODE portal (https://www.encodeproject.org/) includes a large collection of datasets (for example, gene expression enhancers, promoters, open chromatin and histone marks) collected across over 100 human tissues and cell types. These datasets have been further integrated to identify active CREs in those tissues and cell types.

Chromatin is a complex of DNA and protein which is used to package long DNA molecules into a more dense structure. As chromatin condensation proceeds, the 3D spatial organisation of chromatin is established, in which highly self-interacting regions 0.1–1 Mb in size is termed Topological Associating Domains (TADs) (Shanta et al., 2020). Few technologies can be used to investigate chromatin. ChIP-seq can be used to identify histone marks, which are small molecules, usually a methyl (Me) or acetyl (Ac) group, that covalently bond to the amino acid units of histone tails. The histone marks are associated with DNA regulatory regions. For example, histone3 and lysine27 acetylation (H3K27ac) is often associated with active enhancer, which is defined as a sequence of DNA that functions to enhance transcription (Creyghton et al., 2010). In addition, DNase I hypersensitive sites sequencing (DNAse-seq) technology can be used to determine chromatin accessibility. This information can be used to support ChIP-seq identified enhancers, as enhancers tend to associate with chromatin opening regions (Shaliu et al., 2018). More advanced technologies, such as paired-end tagging (ChIA-PET) and low/medium-throughput chromosome conformation capture (Hi-C) chromosome conformation capture can be used to investigate chromatin interactions within the 3D chromatin structure and RNA binding sites in the genome.

These two revolutionary tools can be used to uncover the interplay between different CREs or between CREs and gene coding regions, which are highly essential for understanding the mechanism of gene regulations in health and disease (Glenn et al., 2006). Furthermore, low/medium-throughput RNA-sequencing technology is also included in the ENCODE project to help scientists understand the transcriptome of investigated samples. Finally, ENCODE also investigates DNA Methylation patterns by using whole-genome bisulfite sequencing (WGBS, BS-seq). In mammals, DNA methelytion occurs almost exclusively in CG (or CpG)-rich sequence name CpG island. More importantly, in humans, about 70% of promoters located near the

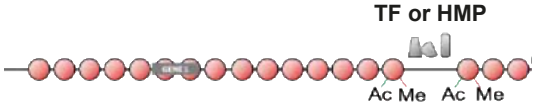
transcription start site of a gene (proximal promoters) contain a CpG island (Serge et al., 2006). Therefore, uncovering mammalian tissues' DNA methylomes can further help scientists understand the mechanism behind gene transcription (Kamel and Giorgio, 2004). Variants identified from GWAS tend to be enriched in enhancers and much more tissue specific than promoters and genes in general (IIakya et al., 2021). In my study, I focused on liver-related cell types, such as HepG2, ESC-derived hepatocytes as well as published data on primary liver cells from the ENCODE portal.

There have been four major data release phases so far. Phase one of the project looked at 1% of the human genome in a few human cell lines (Ewan et al., 2007). First, scientists identified the protein-coding regions and DNA regions that regulate gene transcription (CREs) (Fig3.1a). However, the methods used at phase 1 of the ENCODE project were relatively low-throughput. Therefore, only 1% of the genome was studied. Phase two of the project extended the search for these functional elements to the whole genome in more human cell lines (Fig3.1b). More importantly, high-throughput sequence-based readouts (for example, RNA-seq and ChIP-seq) were used in phase two to compare to the array-based methods used in phase one. Recently completed ENCODE phase three mapped global 3D DNA functional element interactions and revealed RNA-binding regions (Fig3.1c). The samples included in phase three also moved from human and mouse cell lines to primary tissues. The results of all of these experiments are publically available at the ENCODE portal (http://www.encodeproject.org). The latest phase, phase four, of the ENCODE project focused on the functional impacts of the identified CREs: high-throughput CRISPRa/i screens with clear functional biological endpoint assay fit this purpose well (Fig3.1d). The samples' diversity would also increase and start to include assays to measure condition-specific activation or repression of transcriptional control elements. Ideally, scientists will include multiple stages of the cells during development (for example, there were four distinct stages when pluripotent stem cells turned into hepatocytes), including even cell lines at multiple passages, as the current ChIP-seq results from ENCODE 4 did not represent the real-time epigenomic modifications along stem cell to somatic cell differentiaition (Abascal et al., 2020b).

**(A) Phase 1**

Microarray based arrays mapped regions in 1% of human genome of few human cell lines

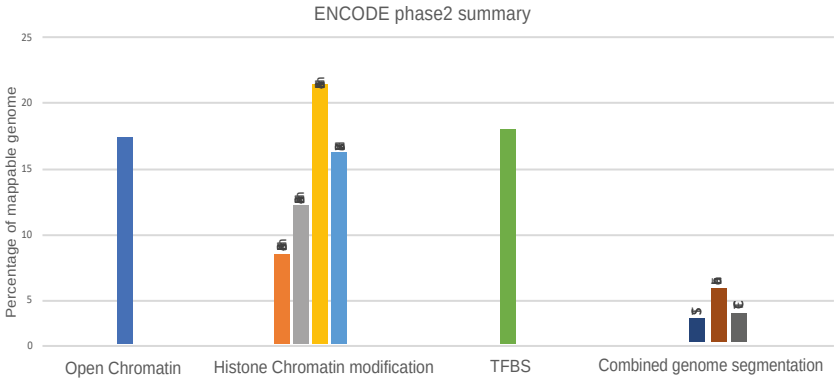
- •Transcribed regions
- •Open chromatin
- •TF or HMP binding sites




**(B) Phase 2**

Sequencing-based technologies mapped the whole human genome and transcriptome of a vaiety of mammalian cell lines

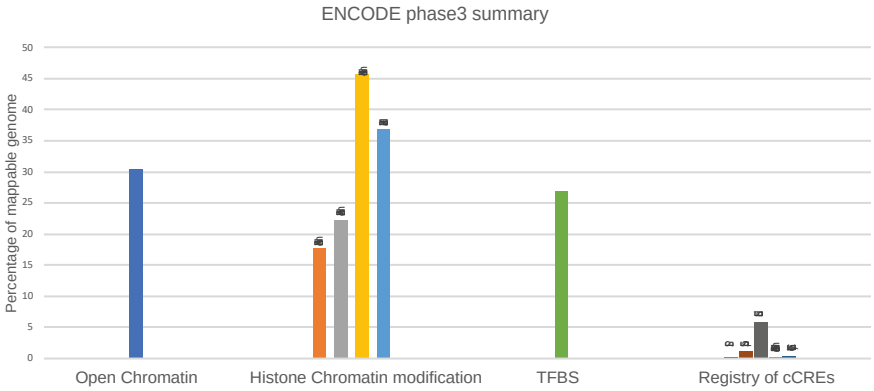
- •TF or HMP binding sites (CHIP-seq)
- •Transcribed regions (RNA- seq)




**(C) Phase 3**

Expand ENCODE2 production and add new assays to reveal RNA binding and 3D organisa-tion of chromatin of mammalian cell lines and primary tissue

- •chromatin interaction analysis (ChIA-PET and Hi-C)
- • Registry of cCREs




**(D) Phase 4 (ongoing)**

Expand ENCODE3 production and add new assays to give DNA functional elements' functional interpretation

- • DNA functional elements' assocaition with diseases
- • DNA functional elements' assocaition with normal human

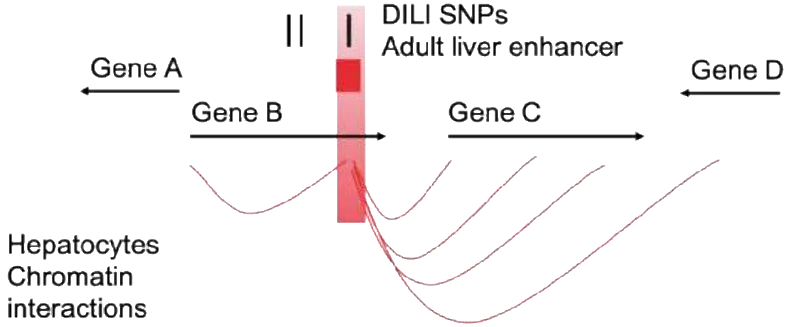


**Figure 3.1 Summary of ENCODE project** (A) Schematic illustration of ENCODE phase 1, transcription factor (TF) or histone-modifying protein (HMP) binding site normally has open chromatin and histone residue modification (Histones are red circles in the figure). (B) Diagrams illustrating ENCODE phase 2 include open chromatin regions, four histone modification markers, transcription factor binding sites (TFBS) and combined genome segmentation, such as transcription start site (TSS), enhancers and CCCTC-binding factor (CTCF) binding sites. The overall percentage of the mappable genome is low. Figure adapted from (Abascal et al., 2020b)**.** (C) Diagram illustrating ENCODE phase3, all original phase2 assays and a novel registry of candidate CREs (cCREs). The overall percentage of the mapped genome is higher than phase 2. Figure adapted from (Abascal et al., 2020b)**.** (D) One of the goals for ENCODE phase 4 was to determine how disease-associated SNPs affect gene regulation. This graph shows the DILI variant (black stripe in highlighted area) generated from *in silico* analysis residing in the adult liver enhancer (red square in highlighted area) near gene B, which regulated more distal genes' expression.

### 3.1.3 CRISPR activation and interference technology

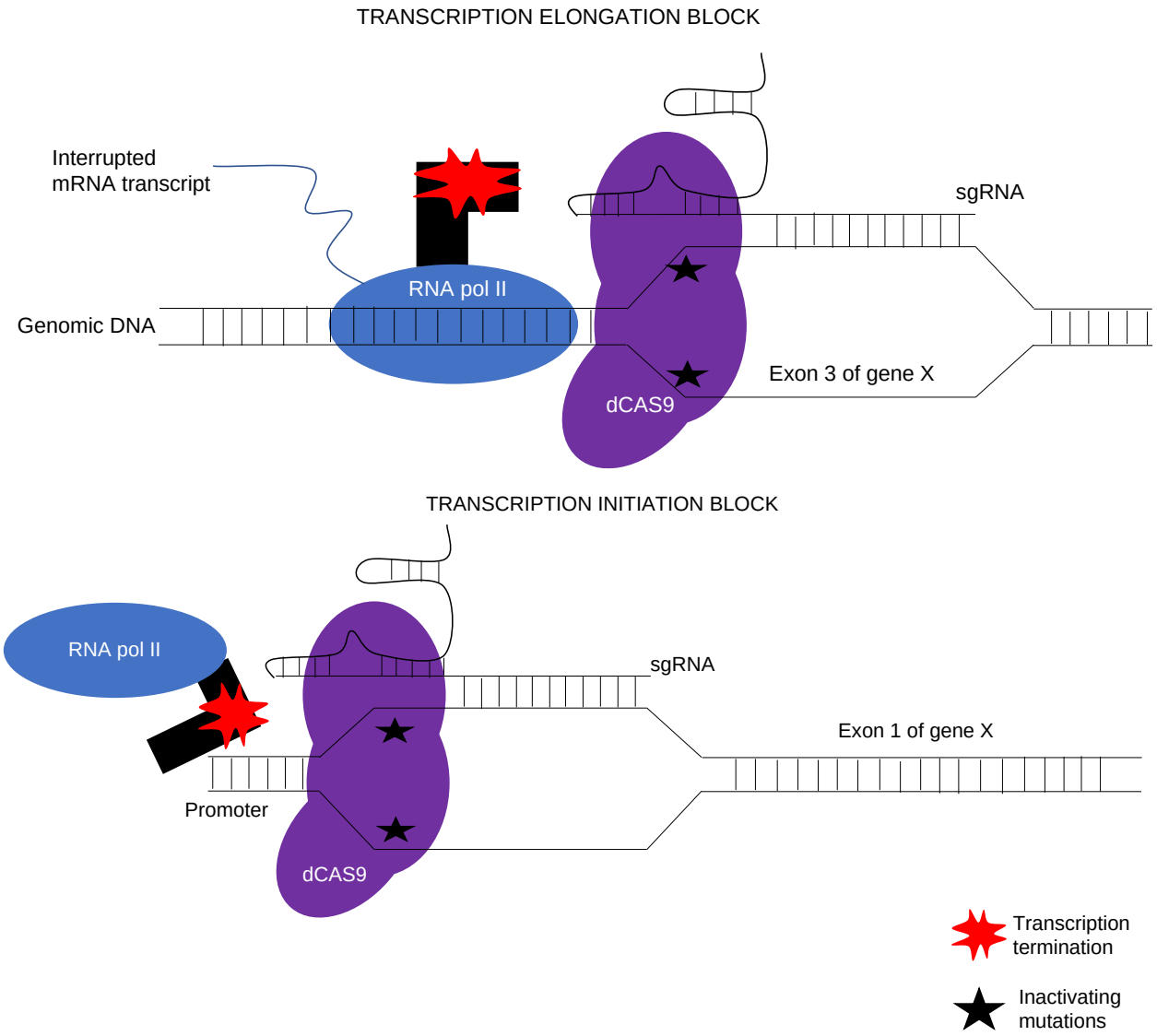
A key characteristic of CRISPR-Cas9 technology is its ability to bind DNA theoretically anywhere in the genome where PAM sites are available. In addition, CRISPR coupled with effector domains can be used to regulate genes at the transcriptional level. CRISPR interference (CRISPRi) is a genetic perturbation technique that represses human cell gene expression in a sequence-specific manner. CRISPRi uses a modified Cas9 protein, in which the two cleavage domains, HNH and RuvC, acquired D10A and H840A mutations. These mutations result in a catalytically inactive Cas9 protein (dCas9), which remains able to bind DNA. The dCas9 can interfere with gene transcription either through blocking the passage of RNA polymerase if recruited to a transcription start site or through fusion with repressive protein domains to induce epigenetic repression (Fig3.2 a and b) (Gilbert et al., 2013). The pioneering study from Lei et al first demonstrated dCas9:sgRNA complex could efficiently repress gene expression in E.coli in a controlled and reversible manner (Qi et al., 2013). Native elongating transcript sequencing (NET-seq) data further illustrated how CRISPRi complex blocked RNA polymerase from elongation to block gene transcription. Multiple sgRNAs could be delivered at the same time to interrogate multiple sites of the E.coli genome with

very high precision. Moreover, evidence from the LacZ β-galactosidase assays from the dCas9 knockdown strain revealed that this technology can interrogate the regulatory functions (activating or repressing) of genes and cis-elements in the complex regulatory network. Most importantly, CRISPRi can knock down targeted gene expression in mammalian cells (Qi et al., 2013).

Transcriptional repression mediated by Krüppel-associated box-zinc finger proteins (KRAB-ZFPs) require interaction with chromatin-remodelling factors. Genome-wide studies of KRAB–ZFP DNA-binding patterns have been carried out to illustrate its transcription repression model in vivo. The study suggested KRAB and its co-repressor KRAB-associated protein-1 (KAP1) could silence promoters by changing chromatin status. Silenced promoters exhibited a loss of histone H3-acetylation, an increase in H3 lysine 9 trimethylation (H3K9me3) and reduced RNA Pol II recruitment. Therefore, CRISPRi's gene-silencing ability can be further enhanced by fusing dCas9 to KRAB. The new complex recruited histone-modifying proteins, ultimately creating a tightly wound, inactive heterochromatin histone state (Peddle et al., 2020). Using a similar approach, dCas9 can also fuse with transcription activators, such as VP64 and p65, to increase gene transcription. Zhang et al introduced the dCAS9 Synergistic Activation Mediator (dCas9-SAM) system, which utilises MS2, P65, HSF1 proteins to recruit various transcriptional factors working synergistically to activate the gene of interest (Zhang et al., 2015).

Overall, although this new technology is far from fully under control, problems such as off-targets effects will continue to show up on the road to translating CRISPRa/i to therapies that are able to make an impact on patients. Nevertheless, it provides an alternative method to study gene transcription. in addition to conventional DNA-based CRISPR gene editing. More importantly, CRISPRa/i provides an effective and simple way for scientists to decode the noncoding regions (such as CREs) of the human genome.

(A)


TRANSCRIPTION ELONGATION BLOCK
Interrupted
mRNA transcript
sgRNA
RNA pol II
Genomic DNA
Exon 3 of gene X
dCAS9
TRANSCRIPTION INITIATION BLOCK
RNA pol II
sgRNA
Exon 1 of gene X
Promoter
dCAS9
Transcription
termination
Inactivating
mutations

(B)

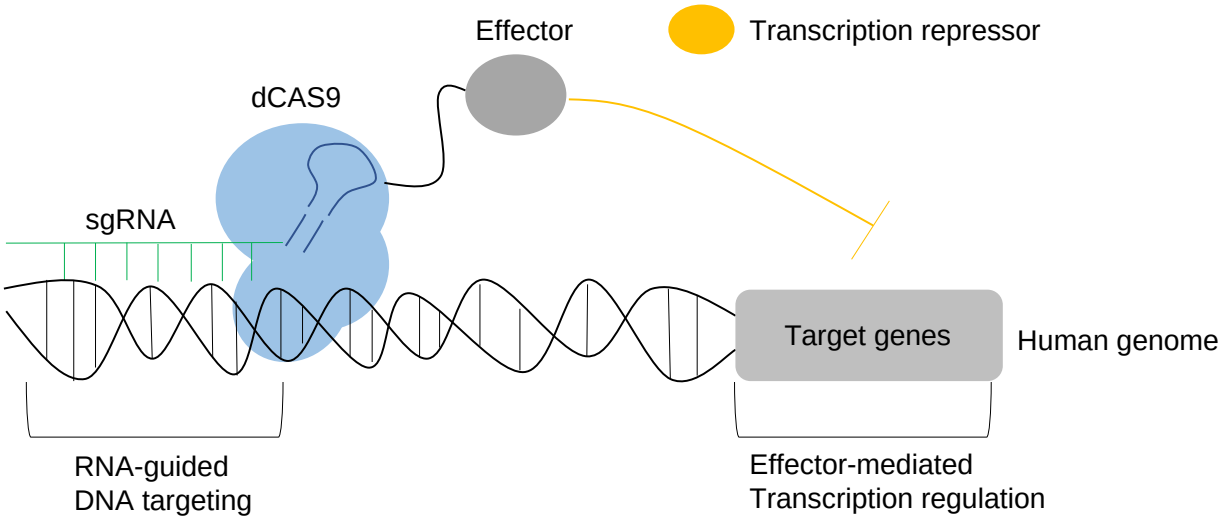



**Figure 3.2 Summary of CRISPR interference mechanisms** (A) Diagram illustrating dCas9 of CRISPRi can either block the RNA polymerase to elongate and transcribe mRNA (elongation blocking) or prevent the RNA polymerase from transcription at the beginning (initiation blocking) (B) Schematic representation of sgRNA-guided dCas9 linking to an effector protein which has transcription repression activity to turn an off-target gene. Figure adapted from (Gilbert et al., 2013)

### 3.1.4 Aims and objectives for this chapter

GWAS uses high-throughput genotyping data, such as SNP arrays, exome and whole-genome sequencing, to identify genetic variants that alter disease susceptibility. Many GWAS variants are in the CREs, but how individual variants affect gene expression is largely unknown. In this chapter, my work first focused on using computational approaches to identify DILI-associated SNPs overlapping with hepatic CREs, with the goal of carrying out a large-scale CRISPR screen to identify CREs that are important for bile acid homeostasis in human hepatocytes. I planned to use CRISPRa/i and the iHEP-based bile transportation assay I described in Chapter 2 to test if epigenetic modulation of CREs harbouring DILI-risk SNPs leads to cholestasis phenotypes. Unfortunately, due to the unexpected global COVID-19 outbreak, it was not possible to carry out the experimental part of this project.

This chapter illustrated the following:

1. How I identified the DILI-associated genetic variants.
2. How I identified hepatic CREs containing DILI-associated genetic variants.

## 3.2 Methods

### 3.2.1 GWAS DILI variants' LD block generation

I first queried the GWAS catalogue (https://www.ebi.ac.uk/gwas/) to search for all DILI-relevant studies. I retrieved all lead SNPs with an association P-value lower than 0.05 that were not representing *HLA* alleles. I decided to exclude all *HLA* variants from the analysis due to this locus' low recombination frequency nature (Cullen et al., 1997).This means DILI SNPs identified from this study in *HLA* regions might be very far away from the actual DILI causal variants and therefore results in unnecessary downstream disease functional characterization studies. All variants in high linkage disequilibrium (LD) with lead DILI SNPs are likely to be inherited together. Therefore, LD SNPs might also contribute to DILI phenotype. The LD variants were identified using the LD proxy tool (*https://ldlink.nci.nih.gov/?tab=ldproxy*) (Mitchell et al, 2015). Populations were selected according to the patient cohort, and I considered SNPs with r2 > 0.8 in high LD with the lead SNP for each locus. Thus, a curated list of lead DILI SNPs, and SNPs with high LD were generated.

### 3.2.2 Identification of regulatory DILI variants

The goal for this part of the project was to identify the DILI variants from the previous section overlapping with CREs of relevant liver cell/tissue, such as the HepG2 cell line and adult liver tissue; all CREs were available on the ENCODE portal. On the University of California, Santa Cruz (UCSC) genome browser, I connected the track hub 'Roadmap Epigenomics Data Complete Collection at Wash U VizHub' to assess human liver epigenomic datasets derived from the Roadmap Epigenomics Project. Then, I set up a new track from the hub track, 'chromHMM'. Here, the 'chromHMM' and three sample types, 'HepG2', 'Adult Liver' and 'HepG2 TF binding sites', were selected for establishing subtracks. Finally, curated DILI SNPs were selected by comparing and integrating the information from each track by using bedtools (Aaron et al, 2010). SNPs that overlap with regions marked by chromHMM as 'enhancer', 'flank_active_TSS region' or 'Active_TSS region' were candidate functional SNPs. Those SNPs were used in the subsequent analysis.

1. The following command line codes were used to identify DILI variants located in enhancers and active TSS regions of adult liver tissue.

1) The command line for selecting the enhancers from adult liver tissue chromHMM: grep ‘Enh’/Adult Liver tissue chromHMM.bed
2) The command line for intersecting DILI variants with the enhancer sites of adult liver tissue: bedtools intersect-a/Adult Liver tissue chromHMMe-b/DILI variants collection.txt
3) The command line for selecting the TSS from adult liver tissue chromHMM: grep ‘TssA’/Adult Liver tissue chromHMM.bed
4) The command line for intersecting DILI variants with the TSS sites of adult liver tissue:
5) Bedtools intersect-a/Adult Liver tissue chromHMMTSS-b/DILI variants collection.txt

2. The following command line codes were used to identify DILI variants located in enhancers and active TSS regions of the HepG2 cell line.

1) The command line for selecting the enhancers from HepG2 cell line chromHMM: $ grep ‘Enh’/HepG2 cell line chromHMM.bed
2) The command line for intersecting DILI variants with the enhancer sites of HepG2 cell line: bedtools intersect-a/HepG2 cell line chromHMMe-b/DILI variants collection.txt
3) The command line for selecting the TSS from HepG2 cell line: $ grep ‘TssA’/HepG2 cell line chromHMM.bed
4) The command line for intersecting DILI variants with the TSS sites of HepG2 cell line dataset: $ bedtools intersect-a/HepG2 cell line chromHMMTSS-b/DILI variants collection.txt

3. The following command line codes were used to identify DILI variants located in TFBS of the HepG2 cell line.

1) The command line for intersecting DILI variants with the TFBS of HepG2 cell line dataset: $ bedtools intersect-a/HepG2TFBS.bed-b/DILI variants collection.txt

### 3.2.3 Nonsynonymous coding variant identification

Table_annovar.pl program was used to find nonsynonymous variants from the curated DILI variants list. The DILI variants bed file was converted to ANNOVAR's input variant file VCF format by using the following command line:
$ convert2annovar.pl-format rsid/DILI variants collection.txt/hg19_DILI variants collection.txt>/DILI variants collection.avinput

Then the following command line was used to generate the output file, which includes nonsynonymous variant information (table_annovar.pl is a program from ANNOVAR):
$ perl table_annovar.pl DILI variants collection.avinput humandb/-buildver hg19-out myanno-remove-protocol refGene,cytoBand,exac03,avsnp147,dbnsfp30a-operation gx,r,f,f,f-nastring.-csvout-polish-xref hg19_refGene.txt

## 3.3 Results

### 3.3.1 Functional annotation of DILI-associated SNPs

The first part of this project involved filtering previous published GWAS results. There were 10 lead variants that have been identified in GWAS studies across 10 different loci; they are shown along with their nearest gene in Table 5. To enlarge the CRISRP screening sites, all SNPs in LD with the 10 lead DILI SNPs were all included and rearranged into a bed file format (Table S2). Among all the 446 DILI variants in Table S2, there is one identified nonsynonymous coding variant near Protein tyrosine phosphatase, non-receptor type 22 (*PTPN22*), which altered the protein coding sequence (Table 6, Table S3) (Cirulli et al., 2019; Kowalec et al., 2018; Nicoletti et al., 2017a; Parham et al., 2016; Suvichapanich et al., 2019a; Urban et al., 2012a). Adult liver/HepG2 chromHMMs were downloaded as described in the methods section, then command line bedtools were used to select another two files, 'enhancer' and 'TSS', from the chromHMMs and used in further analysis. By using command line bedtools, the coordinates of DILI-associated variants were intersected with those of liver-active CREs (enhancers and promoters).

To summarise, there are 15 DILI variants localised in four adult liver CRE loci which account for 40% of the total lead DILI variant loci used in this study (Fig3.3a). Furthermore, 34% of the DILI variants are in the lipopolysaccharide-responsive and beige-like anchor protein (*LRBA*) locus. In contrast, only 13% of the DILI variants are in the peroxisome proliferator-activated receptors (*PPARG*) locus (Fig3.3a). Similarly, there are 26 DILI HepG2 TSS variants located in four lead DILI loci; this accounts for 40% of the total lead DILI variant loci used in this study (Fig3.3b). Among the DILI variants in HepG2 cell CRE regions, 46% of the DILI variants are in the ST6 beta-galactoside alpha-2,6-sialyltransferase 1 (*ST6GAL1*) locus, only 7% are in the *PPARG* locus (Fig3.3b). All the above 41 DILI variants located in CREs are not protein-coding sequence variants but could contribute to the human DILI phenotype via gene transcription regulation. Moreover, these 41 DILI variants come from CREs of four independent genetic loci: *PTPN22*, *LRBA*, *PPARG* and Tumour Protein D52 (*TPD52*) (Fig 3.4 a b c d). Last, these non-*HLA* nonsynonymous coding variants identified near lead variant *PTPN22* account for 10% of total lead DILI variant loci used in this study (Fig 3.3c). More importantly, the *PTPN22* locus has nonsynonymous coding DILI variants and CRE DILI variants. This makes this locus more susceptible to transcription regulation alteration via CREs or change in amino acid sequence via missense mutation, which can potentially lead to DILI phenotype.

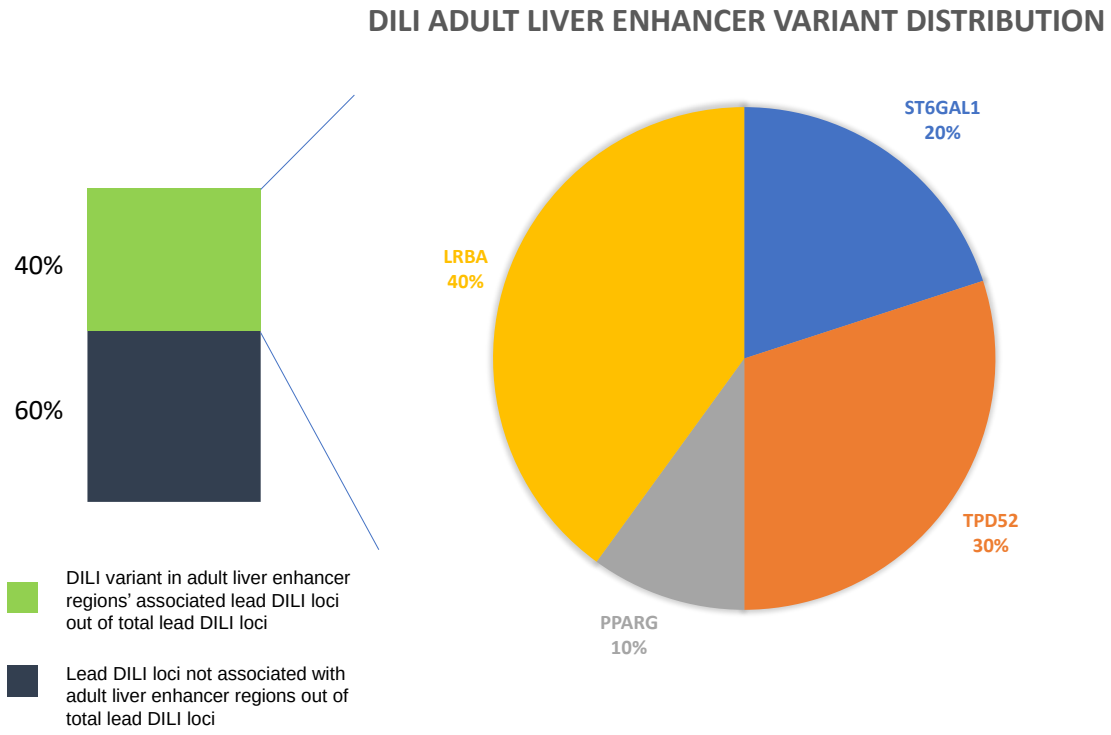
**Table 3.1 Lead DILI variants and their loci from GWAS**

| DILI variants | Locus | ref | P value/effect size |
|---|---|---|---|
| rs10937275 | *ST6GAL1* | (Kaliyaperumal et al., 2009) | $1.4 \times 10^{-8}$ |
| rs7828135 | *TPD52* | (Parham et al., 2016) | $4.5 \times 10^{-8}$ |
| rs72631567 | *LINC01249;LINC01248* | (Urban et al., 2012b) | $9.7 \times 10^{-9}$ |
| rs2205986 | *IRF6* | (Kowalec et al., 2018) | $2.3 \times 10^{-8}$ |
| rs17036170 | *PPARG* | (Urban et al., 2012b) | $2.0 \times 10^{-8}$ |
| rs116561224 | *CDH19;MIR5011* | (Nicoletti et al., 2017b) | $7.1 \times 10^{-9}$ |
| rs1495741 | *NAT2;PSD3* | (Suvichapanich et al., 2019b) | $7 \times 10^{-11}$ |
| rs114811931 | *LINC02159;GABRB2* | (Martens et al., 1985) | $3 \times 10^{-9}$ |
| rs28521457 | *LRBA* | (Urban et al., 2012b) | $4.8 \times 10^{-9}$ |
| rs2476601 | *PTPN22* | (Cirulli et al., 2019) | $1.2 \times 10^{-9}$ |

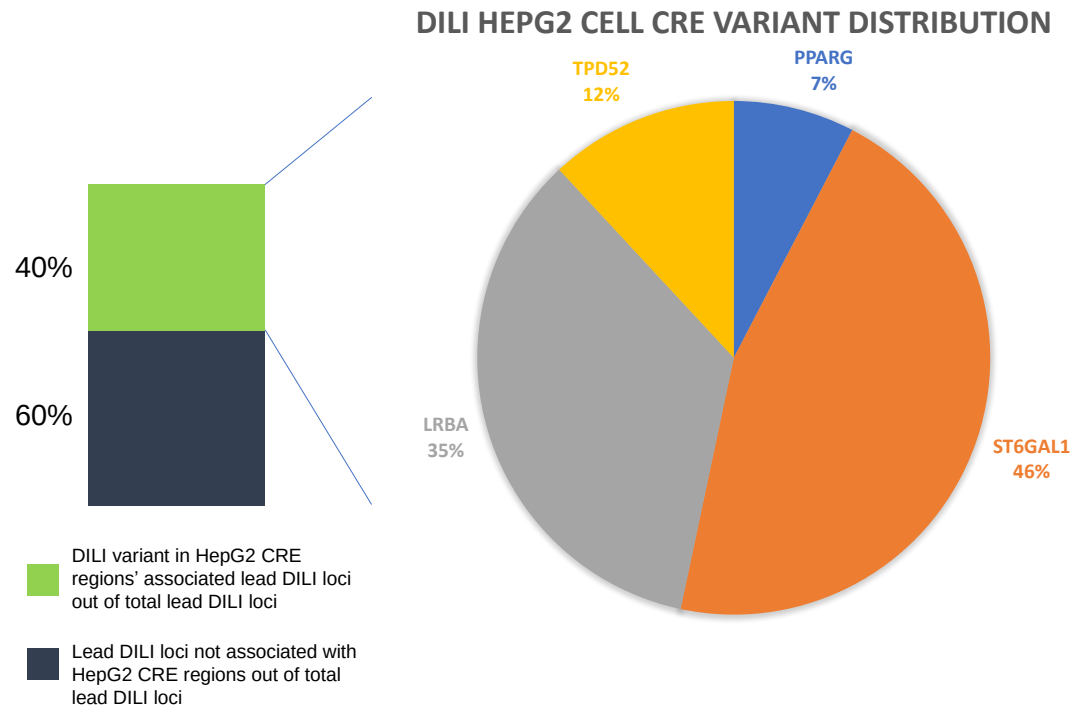
**Table 3.2 DILI nonsynonymous coding variants and their loci**

| DILI coding variants | Locus |
|---|---|
| rs2476601 | *PTPN22* |

(A)


DILI ADULT LIVER ENHANCER VARIANT DISTRIBUTION
40%
60%
DILI variant in adult liver enhancer regions' associated lead DILI loci out of total lead DILI loci
Lead DILI loci not associated with adult liver enhancer regions out of total lead DILI loci
ST6GAL1 20%
TPD52 30%
PPARG 10%
LRBA 40%

(B)


DILI HEPG2 CELL CRE VARIANT DISTRIBUTION
40%
60%
DILI variant in HepG2 CRE regions' associated lead DILI loci out of total lead DILI loci
Lead DILI loci not associated with HepG2 CRE regions out of total lead DILI loci
TPD52 12%
PPARG 7%
LRBA 35%
ST6GAL1 46%

(C)

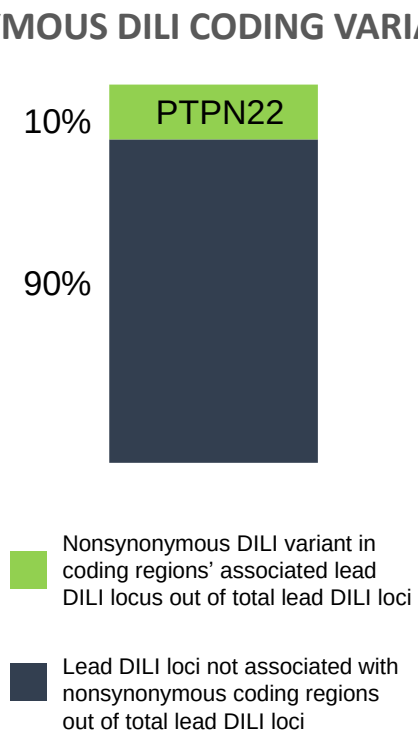
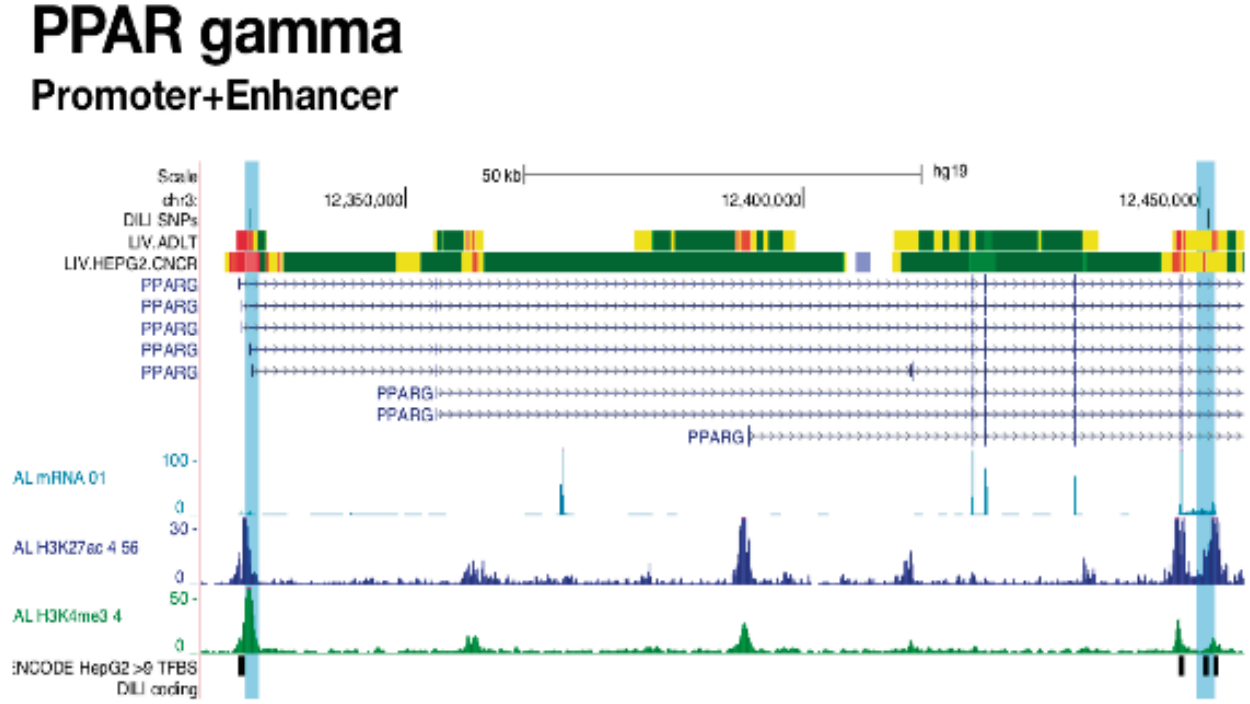
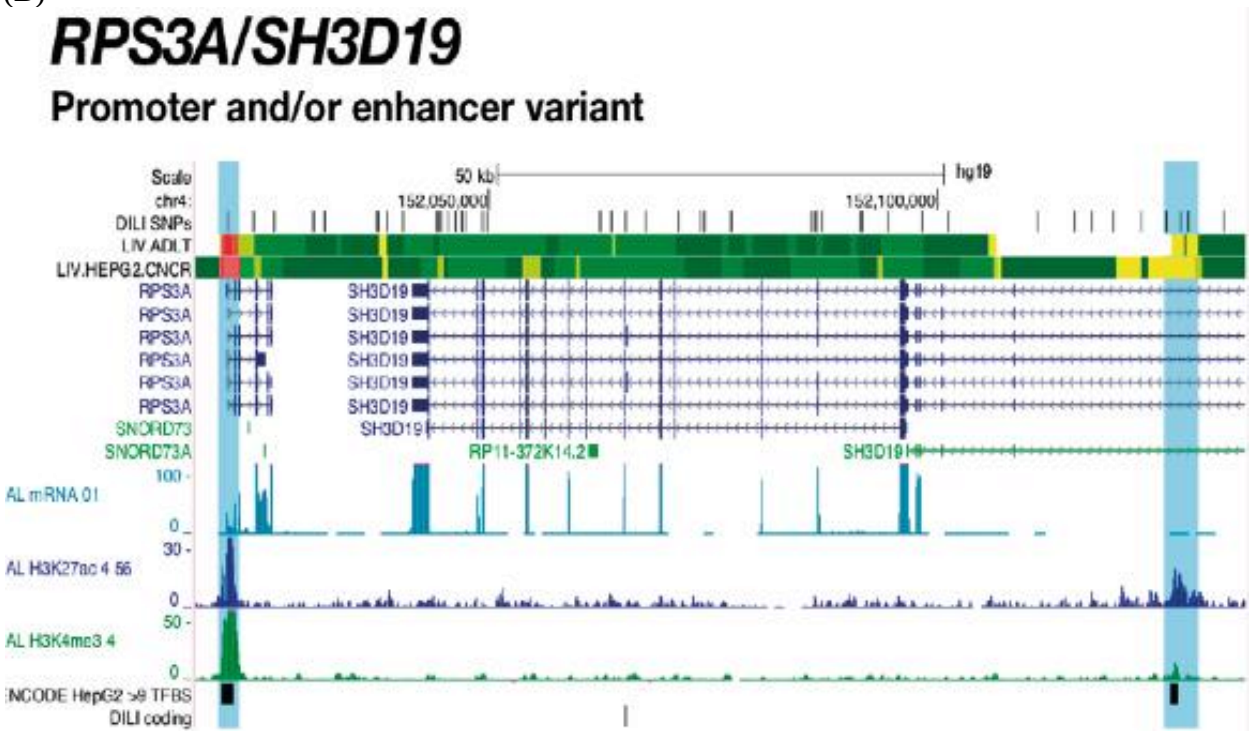



**Figure 3.3 Summary of DILI variant distribution** (A) Pie chart illustrating DILI variants in adult liver CRE sites associated with lead DILI loci, representing 40% of the total lead DILI variant loci used in this study. 34% of the DILI variants are located mostly in the *ST6GAL1* locus, followed by 33% of variants in the *LRBA* locus, then 20% of variants in the *TPD52* locus and, last, 13% of variants in the *PPARG* locus. (B) Pie chart illustrating DILI variants in HepG2 cells' CRE sites associated with the lead DILI loci, representing 40% of the total lead DILI variant loci used in this study. 46% of the DILI variants are located mostly in the *ST6GAL1* locus, followed by 35% of variants in the *LRBA* locus, then 12% of variants in *TPD52* locus and, last, 7% of variants in the *PPARG* locus. (C) Bar chart illustrating DILI nonsynonymous coding variant in lead DILI locus *PTPN22*, representing 10% of the total lead DILI variant loci used in this study.

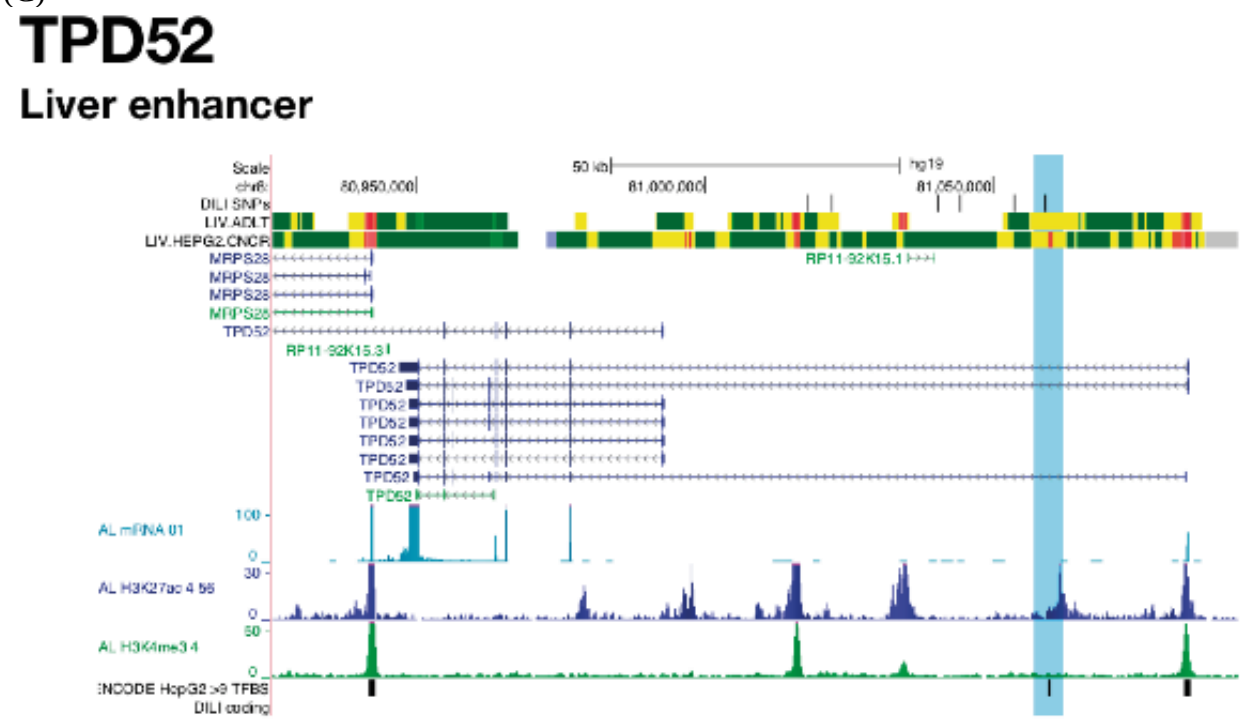
(A)


PPAR gamma
Promoter+Enhancer
Scale
chr3:
50 kb
hg19
12,350,000
12,400,000
12,450,000
DILI SNPs
LIV.ADLT
LIV.HEPG2.CNCR
PPARG
AL mRNA 01
AL H3K27ac 4 56
AL H3K4me3 4
ENCODE HepG2 >9 TFBS
DILI coding

(B)


RPS3A/SH3D19
Promoter and/or enhancer variant
Scale
chr4:
50 kb
hg19
152,050,000
152,100,000
DILI SNPs
LIV.ADLT
LIV.HEPG2.CNCR
RPS3A
SH3D19
SNORD73
SNORD73A
RP11-372K14.2
AL mRNA 01
AL H3K27ac 4 56
AL H3K4me3 4
ENCODE HepG2 >9 TFBS
DILI coding

(C)


TPD52
Liver enhancer
Scale
chr8:
50 kb
hg19
80,950,000
81,000,000
81,050,000
DILI SNPs
LIV.ADLT
LIV.HEPG2.CNCR
MRPS28
RP11-32K15.1
TPD52
AL mRNA 01
AL H3K27ac 4 56
AL H3K4me3 4
ENCODE HepG2 >9 TFBS
DILI coding

(D)

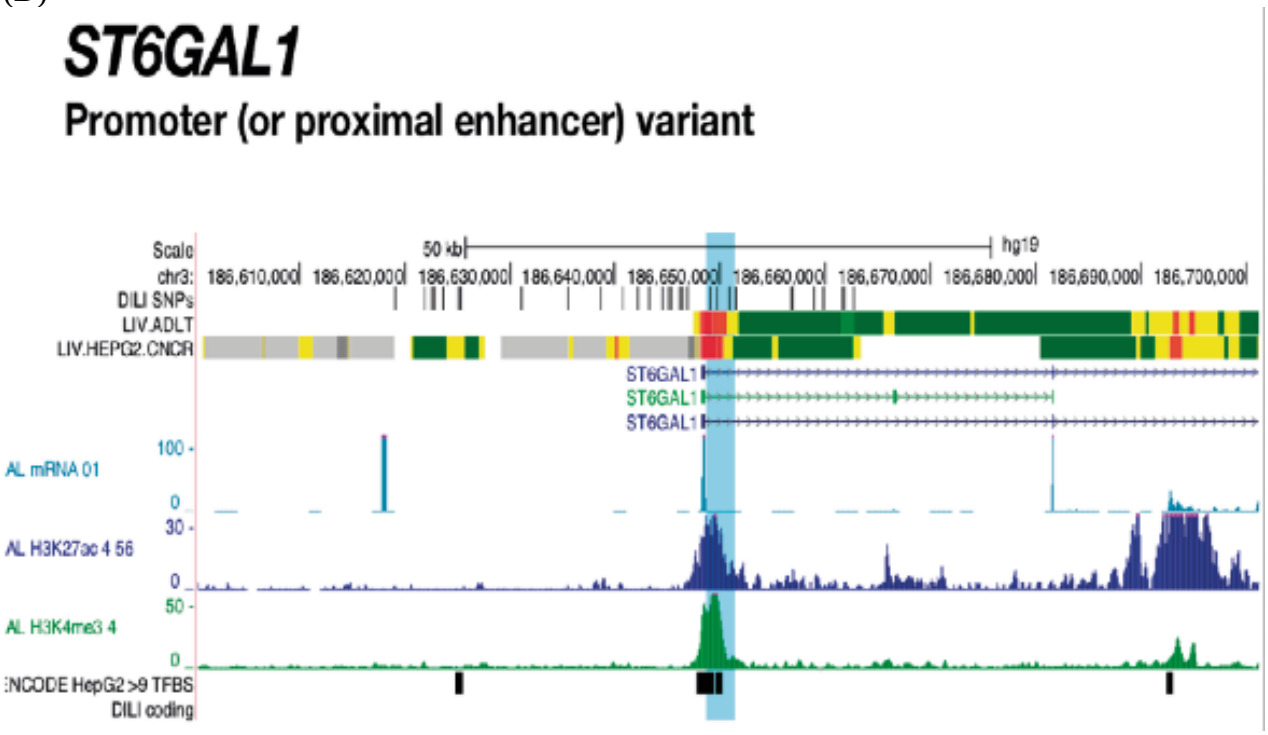

ST6GAL1
Promoter (or proximal enhancer) variant
Scale
chr3:
50 kb
hg19
186,610,000
186,620,000
186,630,000
186,640,000
186,650,000
186,660,000
186,670,000
186,680,000
186,690,000
186,700,000
DILI SNPs
LIV.ADLT
LIV.HEPG2.CNCR
ST6GAL1
AL mRNA 01
AL H3K27ac 4 56
AL H3K4me3 4
ENCODE HepG2 >9 TFBS
DILI coding

**Figure 3.4 Novel DILI SNPs in liver gene CREs** (A) Diagram illustrating liver HEPG2 cancer cell chromHMM tracks and DILI variant tracks at the top. The ROIs here are near the PPAR gamma gene. There are two DILI variants (highlighted in blue) that have been identified to overlap with H3K27ac and H3K4me3 markers of adult liver enhancers and promoter sites; moreover, these two variants also overlap with adult liver mRNA expression and HepG2 cell TFBS. (B) Diagram illustrating four DILI variants (highlighted in blue) identified near 40S ribosomal protein S3a (RPS3A)/SH3 Domain Containing 19 (*SH3D19*), overlapping with H3K27ac and H3K4me3 markers of adult liver enhancers and/or promoter sites. Moreover, these four variants also overlap with adult liver mRNA expression and HepG2 cell TFBS (C) Diagram illustrating two DILI variants (highlighted in blue) identified near *TPD52*, overlapping with H3K27ac and H3K4me3 markers of adult liver enhancer sites. Moreover, these two variants also overlap with adult liver mRNA expression and HepG2 cell TFBS (D) Diagram illustrating four DILI variants (highlighted in blue) identified near *ST6GAL1*, overlapping with H3K27ac and H3K4me3 markers of adult liver promoter/proximal enhancer sites. Moreover, these four variants also overlap with adult liver mRNA expression and HepG2 cell TFBS.

## 3.4 Discussion and conclusion

This chapter's wet lab work was heavily disrupted by the COVID-19 outbreak. Nevertheless, most of the *in silico* analyses had been completed at the time this thesis was drafted. For almost all human diseases, individual susceptibility is influenced by genetic variation. Therefore, solving the relationship between genetic variants and disease phenotypes is highly crucial for understanding diseases' molecular mechanisms and for novel treatment development.

From 1980–1990, linkage analysis and fine mapping within large multiplex pedigrees were used to map rare, monogenic diseases. Sanger sequencing was performed after the potential linked locus was identified to lock down the disease-causing allele. Cellular and animal models were then used to carry out the functional studies for the disease allele. From early 2000, high-throughput sequencing (HTS) technology allowed the rare disease studies to switch from genetic characterisation of small numbers of individuals with similar clinical phenotypes to genome-wide sequencing of large cohorts of phenotypically diverse patients. How to accurately determine the disease-causing genetic variants and their penetration remains problematic for the rare disease research community. Future highly parallelised *in vitro* cellular assays will help to test all disease-

causing variants' functional effects. Moreover, when HTS data combines with RNA-sequencing and DNA Methylation readouts, scientists will be able to identify previously cryptic causal genetic variants, such as noncoding variants for rare disease patients (Koido et al., 2020).

However, for common diseases, such as DILI, it is more difficult to determine the causing variants because the diseases are often polygenic, and the variants are pleiotropic. To make variant analysis more complicated, complex disease-risk alleles are distributed across a continuous spectrum of frequencies and effect sizes. Nevertheless, population-wide genetics research projects, such as GWAS, enable scientists to establish associations between multiple variants of complex disease traits. GWAS results have revealed that the most common disease risk variants were mapped to noncoding regions, and these variants' functions are largely unknown (Koido et al., 2020). DILI variants in *HLA* locus were heavily studied, and previous literature suggested they were playing critical roles in DILI formation in patients via known mechanisms such as presenting processed drug components to activate T cells to degrade hepatocytes (Kindmark et al., 2008; Pichler et al., 2015). However, GWAS reported DILI variants in *HLA* locus and variants in LD with them are likely to be inherited from generation to generation for most of the human population due to this *HLA* locus' extremely low recombination frequency (Cullen et al., 1997). In addition, SNPs in *HLA* regions also associated with multiple other complicated disorders such as type 2 diabetes and Parkinson's disease (Raya-Sandino et al., 2017). Thus, the actual DILI casual variants might be located very far away from *HLA* locus, hence most variants in *HLA* locus are predicted not to have downstream disease functional phenotypes, therefore the *HLA* DILI variants were removed from this study.

Therefore, this study tried to determine if GWAS reported DILI variants located in CRE sites. If so, how do they influence transcriptional regulation? Most importantly, can DILI CRE variants contribute to DILI cholestasis phenotypes *in vitro*? Due to time restraints, only the DILI CRE variant results have been generated. If more time was allowed, I would use the genotype-tissue expression (GTEx) portal (*https://gtexportal.org/home/*) to calculate each DILI CRE variants' expression quantitative trait loci (eQTL) scores. This would allow me to predict associations between the variants identified in this study and DILI-related genes in the human liver. However, DILI CRE variants' eQTL scores in isolation are insufficient to define the specific causal variants

responsible for DILI phenotypes. Additional CRISPRa/i screens and luciferase reporter assays should be used to investigate the functional characterisations of DILI CRE variants.

Every DILI CRE variant leading to functional phenotype changes, such as altered gene expression, should be further validated by introducing relevant point mutations to the healthy cellular models. The other important question is what cellular model should I use for DILI CRE variant studies? A human physiologically relevant cellular model which allows effective genetic and epigenetic manipulations is desirable. This *in vitro* model should ideally include comparable genetic/epigenetic landscapes and liver-cell function to human primary liver cells. More importantly, a highly reproducible assay with clear disease-relevant biological endpoints (bile acid accumulation) is needed in this DILI disease model. Once the DILI causal variant phenotype link is established, technologies, such as ChIA-PET or Hi-C chromosome conformation capture, should be used to generate DILI causal variants' 3D chromatin interaction map to further delineate the epigenetic mechanism of DILI (Miguel-Escalada et al., 2019). Among the four lead DILI loci identified in liver CRE regions from this study, variants located in promoters or enhancers of *PPARG* were highly likely to alter gene expression. *PPARG* was a ligand-dependent TF and a nuclear receptor responsible for sensing hormones, vitamins, endogenous metabolites, and xenobiotic compounds.

The nuclear receptors control the expression of a very large number of genes (Martin et al., 2010). Thomas et al. first identified variant rs17036170, which was upstream of the *PPARG* gene and had a strong association with diclofenac-induced liver injury from a GWAS study of DILI in 783 individuals of European ancestry (Urban et al., 2012a). Early COS-1 cell reporter assay results demonstrated non-steroidal anti-inflammatory drugs (NSAID) diclofenac was an antagonist of *PPARG*, and more importantly, *PPARG* activation had been shown to influence inflammation, which could be linked to the immune-mediated DILI mechanism (Adamson et al., 2002; Zieleniak et al., 2008). Moreover, peroxisome proliferator-activated receptor (PPAR) nuclear receptor pathways were highly relevant to cholestasis, PPAR-regulated genes were reported to impact bile acid metabolism and synthesis for example repression of synthesis via cytochrome P450 7A1 (*CYP7A1*) and cytochrome P450 27A1 (*CYP27A1*) uptake via *NTCP* expression (Onofrio and Hirschfield, 2020). Therefore, the novel DILI variants that resided in *PPARG* CRE regions

identified in this study will help us understand the molecular mechanism behind diclofenac-caused DILI.

The novel DILI variants found in the promoter region of *RPS3A* could also help explain the mechanism of DILI. The lead DILI variant was reported within an intronic region near *LRBA*. This GWAS study recruited patients from Europe, the USA and South America, and there were 11 representative DILIs (hepatocellular death in particular) significantly associated with this SNP (Nicoletti et al., 2017b). The variants in LD, with the lead DILI locus *LRBA*, yielded a positive result: RPS3A, which is responsible for catalysing protein synthesis. An organic pollutant polybrominated diphenyl ether-99 (PBDE-99) had been shown to downregulate *RPS3A* in HepaRG. Meanwhile, a mouse study had shown PBDE-99 caused over-accumulation of unconjugated bile acid, which could lead to cholestasis (Zhang et al., 2020a). Recent studies highlight RPS3A's importance in nuclear factor kappa light chain enhancer of activated B cells (NK-KB) signalling, as its function is an integral part of nuclear factor kappa-light-chain-enhancer of activated B cells (NFKB) to interact with P65 to stabilise NFKB's association with certain DNA binding sites. NFKB signalling is well-known for the development of inflammatory disease, which could lead to DILI (Wan and Lenardo, 2010). Last, co-immunoprecipitation results suggested that C/EBP Homologous Protein (CHOP) is bound to RPS3A in murine erythroleukaemia cells, and RPS3A appeared to regulate CHOP and cell apoptosis (Cui et al., 2000). The link between RPS3A and cholestasis or DILI is still unknown. Nevertheless, my study has provided some novel, nonHLA-related DILI variants to study the cause of the disease.

In addition, the lead DILI variant near *LRBA* had another LD SNP in the intronic region of *SH3D19*. This SNP co-localised with the enhancer region of adult liver and HepG2 cells. To the best of my knowledge, there were no available DILI phenotypes associated with *SH3D19*. This gene encoded multiple SH3 domain-containing proteins, which interacted with EBP and members of the ADAM (metalloendopeptidase) family. First, SH3 domain-containing protein EEN shares some structural homology with the SH3D19 protein. Yam et al. demonstrated EBP interacting simultaneously with EEN and Sos, a guanine-nucleotide exchange factor for Ras in HeLa cells. Co-expression of EBP and EEN suppresses Ras-induced cellular transformation and Ras-mediated activation of ETS Like-1 protein (ELK-1) (Yam et al., 2004). BSEP translocation involves

PI3K/PIPs/Ras/Raf/Erk signalling pathways; therefore, SH3D19 could indirectly cause drug-induced cholestasis by altering key bile salt transporter's translocation (Anwer, 2004). Second, mouse *Eve-1* gene translated to Met-371, which was the human SH3D19 counterpart, had been suggested to positively regulate ADAM; hence, the shedding of Epidermal Growth Factor Receptor (EGFR) ligands and transactivation (Tanaka et al., 2004). Svinka et al. generated conditional EGFR KO in hepatocytes and cholangiocytes of *Mdr2*−/−mice. This led to strongly aggravated liver damage phenotypes, suggesting EGFR signalling was pivotal to protecting mice from getting cholestatic liver disease (Svinka et al., 2017). How the DILI enhancer variants in *SH3D19* locus cause DILI is still largely unknown in humans; hence, it requires further investigation.

Lapatinib is associated with a low incidence of serious liver injury. Parham et al. carried out a GWAS study on 1,194 and 1,259 women with EGFR 2-positive early breast cancer randomised to lapatinib and placebo, respectively, who consented to provide a blood sample for pharmacogenetic analysis. GWAS results suggested one intronic variant in *TPD52* was significantly associated with Laptinib-induced alanine aminotransferase (ALT) elevation (Parham et al., 2016). My *in-silico* analysis suggests that this leads DILI SNP to reside in the liver enhancer region of the tumour protein D52 (*TPD52*) locus. *TPD52* had been shown to cause metastatic phenotype transformation when ectopically expressed in fibroblasts. It was also involved in homologous recombination-dependent DNA damage repair in osteosarcoma cells. More recent studies indicated *TPD52* regulated endosomal trafficking in secretory cell types and was critically regulated by ser136 phosphorylation (Byrne et al., 2014). Most *TPD52* literature was cancer-related; therefore, its role in DILI is largely unknown and requires further investigation.

All the *in silico* analysis results I generated in this study would not have any practical meaning until they were tested in the wet lab functional assay from Chapter 2. There are three main experiments that should be carried out if time allows. First, the gRNA sequence, designed to target all DILI variants in the four lead loci identified from this study, then gRNAs cloned into CRISPRa/i plasmids. The backbone of CRISPRa plasmid can be found at (https://www.addgene.org/75112/). CRISPRi plasmids can be found at (https://www.addgene.org/118155/). Finally, all plasmids should be packaged into lentivirus

separately. Lentivirus would be used to transfect healthy iHEPs. Each well would accommodate one gRNA only. In general, three different gRNAs should be designed to target the same variant loci to increase the result reliability. By manipulating DILI loci's epigenetic landscapes in healthy iHEPs, these experiments are expected to alter CDF flow, causing intrahepatic CDF accumulation. Second, reported DILI drugs from GWAS should be added to healthy iHEPs to observe if any DILI-related cellular phenotype appears in the *in vitro* assay. In theory, all identified DILI variants are predicted to alter bile transportation when drugs are added to the healthy iHEPs, but this does not exclude contribution to hepatocellular death as a DILI disease phenotype. Third, once scientists recognise that certain drugs can cause DILI phenotype in the iHEP-based *in vitro* assay, using CRISPRa/i to target the DILI variants also becomes a way to rescue the DILI phenotype.

This is a proof-of-concept experiment designed to test if we can use CRISPR-based epigenetic manipulation to safely allow patients to take drugs that normally cause severe liver injury. It is not clear if the DILI CREs identified only regulate the genes nearby or if they regulate other remote genes within the TAD. Therefore, it will be helpful to use Hi-C to create a regulatory map for each CRE DILI variant from our list once a clear phenotype is identified from the functional assay. In addition, the comparison between ES-HEP and adult liver deoxyribonuclease-I (DNAse-1) and ChIP-seq results is only the beginning of the second project in this chapter. There are a few wet-lab experiments that remain to be addressed.

This chapter aimed to gain more understanding of how mutations in CREs could lead to DILI. Currently, our understanding of the area is still lacking. My study in this chapter used *in silico* methods to generate a list of DILI variants in liver CRE regions, which could contribute to the disease phenotype. I believe that by further establishing the links between these variants and DILI, *in vitro* phenotype will help the DILI community to better understand molecular mechanisms of DILI and generate novel treatments for patients.

# Overall summary

This thesis has demonstrated that generation of PFIC secondary to *TJP2* mutation patient specific iPSC and its relevant isogenic control iPSC are feasible. However, there are two unanswered questions remaining for this study. Firstly, can iPSC be reprogrammed from patients with multiple backgrounds such as different sexes, ages, and genetic make-up? Secondly, it is not known if the iPSC isogenic control with patient specific point mutation p.Y261Sfs*50 can be generated using CRISPR-Cas9? Future experiments should focus on addressing these two questions by designing iPSC disease models with relevant genetic make-ups as close as possible to the patients and to ensure the reproducibility of such methods in modeling larger cohort of patients of this disease. The somatic cell to iPSC reprogramming method is highly inefficient, this can be improved by delivering synthetic mRNAs encoding Yamanaka factors to somatic cells cultured in microfluidic device (Bell et al., 2016). Recent advancement in CRISPR-Cas9 based genetic engineering tool-prime editor could be used to introduce the patient specific point mutation p.Y261Sfs*50 to healthy iPSC and generate the patient specific iPSC isogenic control in a more efficient manner (Anzalone et al., 2019). Introduction of the disease iPSCs into a chemically defined in vitro hepatocyte differentiation platform reproduced key facet of the PFIC with *TJP2* mutation patients' disease phenotype such as change of canaliculi structure (Sambrotta et al., 2014d). The frameshift mutation p.Y261Sfs*50 caused premature mRNA production which were then degraded by NMD, hence no ZO2 protein was detected from the patient with this mutation (Sambrotta et al., 2014d). The iHEP$^{TJP2\text{-}KO}$ had very low *TJP2* gene expression and no ZO2 protein was detected by immune florescence staining. On the other hand, iHEP$^{PFIC\text{-}patient1}$ had some *TJP2* gene expression, and no proteins were detected by immune florescence staining. The mechanism behind higher *TJP2* gene expression in iHEP$^{PFIC\text{-}patient1}$ remains unknown, it could be explained by potential escape mechanism from nonsense-mediated mRNA degradation such as reinitiate translation at downstream start codon (Neu-Yilik et al., 2011). No mRNA was detected from iHEP$^{TJP2\text{-}KO}$ cells due to the absence of a critical exon of within the cells. Future experiments should include multiple qPCR primers targeting various parts of the *TJP2* locus, and antibodies targeting different epitopes of the ZO2 protein. In addition, iHEP$^{PFIC\text{-}patient1}$ and iHEP$^{TJP2\text{-}KO}$ demonstrated key disease phenotype - bile salt transport abnormality. Matrigel sandwich and spheroids iHEP models in here demonstrated two separate bile salt transport abnormality functional phenotypes. This suggested the final disease phenotypes are highly dependent on the cell culture system used for the

investigation. Future experiments should consider integrating iHEPs in a more complexed culture model such as the microfluidic device due to its similar nature to *in vivo* human liver set-up. Polarity is another important part of PFIC secondary to TJP2 deficiency disease investigation. Previously claudin-1's failure of localisation in the canalicular membrane was observed in patients with severe TJP2 deficiency (Sambrotta et al., 2014d). It is possible the missing of ZO2 protein de-stabilizes the junctional structure and consequently leads to leakage of bile into the paracellular space of the liver parenchyma in patients. More importantly there are two questions remaining unanswered, firstly how well does the healthy iHEP in a culture system reflect *in vivo* human canaliculi polarity? Secondly, how does the polarity changes if TJP2 is mutated in such a model? Future experiments should focus on staining other key apical and basolateral membrane proteins in the existing *in vitro* iHEP canaliculi model. Once a fully polarized canaliculi membrane is established in an iHEP *in vitro* model, other tight junction proteins such as claudins and occludins should also be stained and look for polarity changes assemble the patients' canaliculi in the iHEP models.

The next goal is to use genetic engineering tools to correct the point mutation p.Y261Sfs*50 in patient iPSC. The corrected iPSC differentiated iHEPs should no longer have a cholestatic disease phenotype. In theory, this iPSC based cell and gene therapy can be used to treat PFIC patients with TJP2 mutation. This concept had been proofed by Dr Tamir Rashid's work on using ZNF to correct the *SERPINA1* mutation in iPSC derived from A1AT deficient patient, and polymers accumulated in the diseased iHEPs were successfully removed (Yusa et al., 2011).

In addition, the iHEPs generated from iPSCs in this study are relatively immature and heterogenous. In addition, this iHEP differentiation protocol is a long process with high inter-batch variability. Current problems in iHEP differentiation can lead to inadequate cholestatic disease phenotype on-set and can have a negative impact on the following disease mechanism studies and drug screening. Therefore, a shortened iPSC to hepatocytes differentiation protocol with higher cellular purity and functionality is desirable. Primary human cells and foetal organs' RNA seq and ChIP-seq results from ENCODE can be used as a guidance to help iPSC differentiate to somatic cells with higher function and purity (Rackham et al., 2016). More recently, forward reprogramming technology allowed iPSC to differentiate to high quality neurons, skeletal

myocytes, and oligodendrocytes in an efficient and reproducible manner (Pawlowski et al., 2017). Both approaches can be extrapolated to enhance iHEP differentiation.

The third chapter of this thesis presented a collection of DILI associated variants located in CRE regions, which can be used to further establish the connection between functional DILI disease phenotype and the patients' genetic composition. All lead DILI variants identified from this study are in genes which previously reported to either directly related to DILI or having important impacts on liver function which might indirectly related to DILI. The bioinformatic analysis highlighted the variants with the most potential to contribute to DILI, thus gives researchers confidence to move on to the transcriptional and the functional investigation of these DILI variants identified. Complex disorders such as DILI tend to be genetically heterogenous, which points toward a likelihood of multiple genes and pathways being implicated in DILI development (Stephens et al., 2012). Future research should not only focus on the single variant's contribution to DILI, but also variants' collective contribution to DILI. As soon as the research community has more understanding on how GWAS DILI genetic variants regulate genes in liver cells, clinician can start to advise patients on what drugs to take based on their genetic information, so DILI can be avoided as much as possible.

Overall, my PhD work has introduced a novel patient iPSC-based platform to investigate the biology of PFIC with TJP2 deficiency. In addition, the technology innovation of combining *ex-vivo* tissue engineering, low/medium-throughput imaging assay, and genetic manipulation together established an efficient and effective disease model which has huge potential for industry and academic research. This project initially set out to seed the iHEPs into a microwell with 3D protein coating, aimed to study canaliculi and hepatocytes transportation at a single cell level (Zhang et al., 2020b). However, the microwell building requires extensive expertise in tissue engineering and takes a long time to establish. Then the research attention shifted to build a simpler fit-for-purpose low/medium-throughput cholestatic disease model which accurately reflected few key patients' disease phenotypes. When we tried to use this novel cholestatic disease model to study more complexed disease like DILI, the wet lab research has been interrupted by COVID-19 outbreak, but the lead DILI variants identified from the *in-silico* analysis should be pursued further to unlock their functional applications and their contribution to DILI.

# References

1. F. Duclos, F. Rodius, K. Wrogemann, J. L. Mandel, M. Koenig, The friedreich ataxia region: Characterization of two novel genes and reduction of the critical region to 300 kb. *Human Molecular Genetics* (1994), doi:10.1093/hmg/3.6.909.
2. A. Chlenski, K. V. Ketels, G. I. Korovaitseva, M. S. Talamonti, R. Oyasu, D. G. Scarpelli, Organization and expression of the human zo-2 gene (tjp-2) in normal and neoplastic tissues. Biochimica et Biophysica Acta - Gene Structure and Expression (2000), doi:10.1016/S0167-4781(00)00185-8.
3. M. Itoh, M. Furuse, K. Morita, K. Kubota, M. Saitou, S. Tsukita, Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and ZO-3, with the COOH termini of claudins. Journal of Cell Biology (1999), doi:10.1083/jcb.147.6.1351.
4. A. Traweger, R. Fuchs, I. A. Krizbai, T. M. Weiger, H. C. Bauer, H. Bauer, The tight junction protein ZO-2 localizes to the nucleus and interacts with the heterogeneous nuclear ribonucleoprotein scaffold attachment factor-B. Journal of Biological Chemistry (2003), doi:10.1074/jbc.M206821200.
5. T. Walsh, S. B. Pierce, D. R. Lenz, Z. Brownstein, O. Dagan-Rosenfeld, H. Shahin, W. Roeb, S. McCarthy, A. S. Nord, C. R. Gordon, Z. Ben-Neriah, J. Sebat, M. Kanaan, M. K. Lee, M. Frydman, M. C. King, K. B. Avraham, Genomic duplication and overexpression of TJP2/ZO-2 leads to altered expression of apoptosis genes in progressive nonsyndromic hearing loss DFNA51. American Journal of Human Genetics (2010), doi:10.1016/j.ajhg.2010.05.011.
6. S. Lechuga, L. Alarcón, J. Solano, M. Huerta, E. Lopez-Bayghen, L. González-Mariscal, Identification of ZASP, a novel protein associated to Zona occludens-2. Experimental Cell Research (2010), doi:10.1016/j.yexcr.2010.09.008.
7. M. Sambrotta, S. Strautnieks, E. Papouli, P. Rushton, B. E. Clark, D. A. Parry, C. V. Logan, L. J. Newbury, B. M. Kamath, S. Ling, T. Grammatikopoulos, B. E. Wagner, J. C. Magee, R. J. Sokol, G. Mieli-Vergani, J. D. Smith, C. A. Johnson, P. McClean, M. A. Simpson, A. S. Knisely, L. N. Bull, R. J. Thompson, Mutations in TJP2 cause progressive cholestatic liver disease. Nature Genetics (2014), doi:10.1038/ng.2918.
8. L. N. Bull, R. J. Thompson, Progressive Familial Intrahepatic Cholestasis. Clinics in Liver Disease (2018), , doi:10.1016/j.cld.2018.06.003.
9. A. M. Van der Bliek, F. Baas, T. Ten Houte de Lange, P. M. Kooiman, T. Van der Velde-Koerts, P. Borst, The human mdr3 gene encodes a novel P-glycoprotein homologue and gives rise to alternatively spliced mRNAs in liver. EMBO J (1987), doi:10.1002/j.1460-2075.1987.tb02653.x.
10. C. R. Lincke, J. J. M. Smit, T. Van der Velde-Koerts, P. Borst, Structure of the human MDR3 gene and physical mapping of the human MDR locus. Journal of Biological Chemistry (1991), doi:10.1016/s0021-9258(19)67788-4.

11. N. D. Weber, L. Odriozola, J. Martínez-García, V. Ferrer, A. Douar, B. Bénichou, G. González-Aseguinolaza, C. Smerdou, Gene therapy for progressive familial intrahepatic cholestasis type 3 in a clinically relevant mouse model. Nature Communications (2019), doi:10.1038/s41467-019-13614-3.
12. S. Ruetz, P. Gros, Phosphatidylcholine translocase: A physiological role for the mdr2 gene. Cell. 77, 1071–1081 (1994). doi: 10.1016/0092-8674(94)90446-4.
13. H. Bauer, J. Zweimueller-Mayer, P. Steinbacher, A. Lametschwandtner, H. C. Bauer, The dual role of zonula occludens (ZO) proteins. Journal of Biomedicine and Biotechnology. 2010 (2010), , doi:10.1155/2010/402593.
14. Z. Mirjam, D. Punita, B. Thomas, Tight junction proteins in gastrointestinal and liver disease. Gut. 68, 547–561 (2019), ,doi:10.1136/GUTJNL-2018-316906
15. O. Rosmorduc, R. Poupon, Low phospholipid associated cholelithiasis: Association with mutation in the MDR3/ABCB4 gene. Orphanet Journal of Rare Diseases. 2 (2007), , doi:10.1186/1750-1172-2-29.
16. M. Sambrotta, R. J. Thompson, Mutations in TJP2, encoding zona occludens 2, and liver disease. Tissue Barriers (2015), doi:10.1080/21688370.2015.1026537.
17. M. Allen, Compelled by the Diagram: Thinking through C. H. Waddington's Epigenetic Landscape. Contemporaneity: Historical Presence in Visual Culture (2015), doi:10.5195/contemp.2015.143.
18. J. B. GURDON, The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J Embryol Exp Morphol (1962). Issn: 00220752
19. K. H. S. Campbell, J. McWhir, W. A. Ritchie, I. Wilmut, Sheep cloned by nuclear transfer from a cultured cell line. Nature (1996), doi:10.1038/380064a0.
20. C. A. Cowan, J. Atienza, D. A. Melton, K. Eggan, Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science (2005), doi:10.1126/science.1116447.
21. K. Takahashi, S. Yamanaka, Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. 126, 663–676 (2016). doi: 10.1016/j.cell.2006.07.024
22. K. Okita, T. Ichisaka, S. Yamanaka, Generation of germline-competent induced pluripotent stem cells. Nature (2007), doi:10.1038/nature05934.
23. K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell (2007), doi:10.1016/j.cell.2007.11.019.
24. J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin, J. A. Thomson, Induced pluripotent stem cell lines derived from human somatic cells. Science (1979) (2007), doi:10.1126/science.1151526.

25. H. Hong, K. Takahashi, T. Ichisaka, T. Aoi, O. Kanagawa, M. Nakagawa, K. Okita, S. Yamanaka, Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature (2009), doi:10.1038/nature08235.
26. X. shan Yue, M. Fujishiro, M. Toyoda, T. Akaike, Y. Ito, Reprogramming of somatic cells induced by fusion of embryonic stem cells using hemagglutinating virus of Japan envelope (HVJ-E). Biochemical and Biophysical Research Communications. 394, 1053–1057 (2010). doi: 10.1016/J.BBRC.2010.03.122
27. A. E. Omole, A. O. J. Fakoya, Ten years of progress and promise of induced pluripotent stem cells: Historical origins, characteristics, mechanisms, limitations, and potential applications. PeerJ. 2018, 4370 (2018). doi: 10.7717/PEERJ.4370
28. J. A. Thomson, Embryonic Stem Cell Lines Derived from Human Blastocysts. Science (1979). 282, 1145–1147 (1998). doi:10.1126/science.282.5391.1145
29. M. C. N. Marchetto, G. W. Yeo, O. Kainohana, M. Marsala, F. H. Gage, A. R. Muotri, Transcriptional signature and memory retention of human-induced pluripotent stem cells. PLoS One. 4 (2009), doi:10.1371/JOURNAL.PONE.0007076.

30. M. H. Chin, M. J. Mason, W. Xie, S. Volinia, M. Singer, C. Peterson, G. Ambartsumyan, O. Aimiuwu, L. Richter, J. Zhang, I. Khvorostov, V. Ott, M. Grunstein, N. Lavon, N. Benvenisty, C. M. Croce, A. T. Clark, T. Baxter, A. D. Pyle, M. A. Teitell, M. Pelegrini, K. Plath, W. E. Lowry, Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. Cell Stem Cell (2009), doi:10.1016/j.stem.2009.06.008.
31. L. Vallier, T. Touboul, S. Brown, C. Cho, B. Bilican, M. Alexander, J. Cedervall, S. Chandran, L. Ährlund-Richter, A. Weber, R. A. Pedersen, Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. Stem Cells (2009), doi:10.1002/stem.199.
32. S. J. I. Blackford, S. S. Ng, J. M. Segal, A. J. F. King, A. L. Austin, D. Kent, J. Moore, M. Sheldon, D. Ilic, A. Dhawan, R. R. Mitry, S. T. Rashid, Validation of Current Good Manufacturing Practice Compliant Human Pluripotent Stem Cell-Derived Hepatocytes for Cell-Based Therapy. Stem Cells Translational Medicine (2019), doi:10.1002/sctm.18-0084.
33. O. Bar-Nur, H. A. Russ, S. Efrat, N. Benvenisty, Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell (2011), doi:10.1016/j.stem.2011.06.007.
34. C. Bock, E. Kiskinis, G. Verstappen, H. Gu, G. Boulting, Z. D. Smith, M. Ziller, G. F. Croft, M. W. Amoroso, D. H. Oakley, A. Gnirke, K. Eggan, A. Meissner, Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. Cell (2011), doi:10.1016/j.cell.2010.12.032.

35. M. C. Puri, A. Nagy, Concise review: Embryonic stem cells versus induced pluripotent stem cells: The game is on. Stem Cells (2012), , doi:10.1002/stem.788.
36. R. Teshigawara, K. Hirano, S. Nagata, J. Ainscough, T. Tada, OCT4 activity during conversion of human intermediately reprogrammed stem cells to iPSCs through mesenchymalepithelial transition. Development (Cambridge) (2016), doi:10.1242/dev.130344.
37. Y. Buganim, D. A. Faddah, A. W. Cheng, E. Itskovich, S. Markoulaki, K. Ganz, S. L. Klemm, A. Van Oudenaarden, R. Jaenisch, Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 150, 1209–1222 (2012). doi: 10.1016/j.cell.2012.08.023
38. L. A. Boyer, I. L. Tong, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, R. A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 122, 947–956 (2005). doi: 10.1016/j.cell.2005.08.020
39. Y. S. Ang, S. Y. Tsai, D. F. Lee, J. Monk, J. Su, K. Ratnakumar, J. Ding, Y. Ge, H. Darr, B. Chang, J. Wang, M. Rendl, E. Bernstein, C. Schaniel, I. R. Lemischka, Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell (2011), doi:10.1016/j.cell.2011.03.003.
40. A. Banito, S. T. Rashid, J. C. Acosta, S. de Li, C. F. Pereira, I. Geti, S. Pinho, J. C. Silva, V. Azuara, M. Walsh, L. Vallier, J. Gil, Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev. 23, 2134–2139 (2009), , doi:10.1101/gad.1811609
41. F. Soldner, R. Jaenisch, Stem Cells, Genome Editing, and the Path to Translational Medicine. Cell (2018), , doi:10.1016/j.cell.2018.09.010.
42. S. T. Rashid, S. Corbineau, N. Hannan, S. J. Marciniak, E. Miranda, G. Alexander, I. Huang-Doran, J. Griffin, L. Ahrlund-Richter, J. Skepper, R. Semple, A. Weber, D. A. Lomas, L. Vallier, Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. Journal of Clinical Investigation (2010), doi:10.1172/JCI43122.
43. S. M. Choi, Y. Kim, J. S. Shim, J. T. Park, R. H. Wang, S. D. Leach, J. O. Liu, C. Deng, Z. Ye, Y. Y. Jang, Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. Hepatology (2013), doi:10.1002/hep.26237.
44. K. Imagawa, K. Takayama, S. Isoyama, K. Tanikawa, M. Shinkai, K. Harada, M. Tachibana, F. Sakurai, E. Noguchi, K. Hirata, M. Kage, K. Kawabata, R. Sumazaki, H. Mizuguchi, Generation of a bile salt export pump deficiency model using patient-specific induced pluripotent stem cell-derived hepatocyte-like cells. Scientific Reports (2017), doi:10.1038/srep41806.
45. T. Shinozawa, M. Kimura, Y. Cai, N. Saiki, Y. Yoneyama, R. Ouchi, H. Koike, M. Maezawa, R.-R. Zhang, A. Dunn, A. Ferguson, S. Togo, K. Lewis, W. Thompson, A. Asai, T. Takebe, High-Fidelity Drug-Induced Liver Injury Screen Using Human Pluripotent Stem Cell–Derived Organoids. Gastroenterology (2020), doi:10.1053/j.gastro.2020.10.002.

46. S. Zhang, S. Chen, W. Li, X. Guo, P. Zhao, J. Xu, Y. Chen, Q. Pan, X. Liu, D. Zychlinski, H. Lu, M. D. Tortorella, A. Schambach, Y. Wang, D. Pei, M. A. Esteban, Rescue of ATP7B function in hepatocyte-like cells from Wilson's disease induced pluripotent stem cells using gene therapy or the chaperone drug curcumin. Human Molecular Genetics (2011), doi:10.1093/hmg/ddr223.
47. X. Hu, C. Mao, L. Fan, H. Luo, Z. Hu, S. Zhang, Z. Yang, H. Zheng, H. Sun, Y. Fan, J. Yang, C. Shi, Y. Xu, Modeling Parkinson's Disease Using Induced Pluripotent Stem Cells. Stem Cells International (2020), , doi:10.1155/2020/1061470.
48. J. L. Corbett, S. A. Duncan, iPSC-Derived Hepatocytes as a Platform for Disease Modeling and Drug Discovery. Frontiers in Medicine (2019), , doi:10.3389/fmed.2019.00265.
49. J. A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. Science (1979) (2014), , doi:10.1126/science.1258096.
50. Y. G. Kim, J. Cha, S. Chandrasegaran, Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A (1996), doi:10.1073/pnas.93.3.1156.
51. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics (2010), , doi:10.1038/nrg2842.
52. K. Yusa, S. T. Rashid, H. Strick-Marchand, I. Varela, P. Q. Liu, D. E. Paschon, E. Miranda, A. Ordóñez, N. R. F. Hannan, F. J. Rouhani, S. Darche, G. Alexander, S. J. Marciniak, N. Fusaki, M. Hasegawa, M. C. Holmes, J. P. Di Santo, D. A. Lomas, A. Bradley, L. Vallier, Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. Nature (2011), doi:10.1038/nature10424.
53. J. K. Joung, J. D. Sander, TALENs: A widely applicable technology for targeted genome editing. Nature Reviews Molecular Cell Biology (2013), , doi:10.1038/nrm3486.
54. Q. Ding, Y. K. Lee, E. A. K. Schaefer, D. T. Peters, A. Veres, K. Kim, N. Kuperwasser, D. L. Motola, T. B. Meissner, W. T. Hendriks, M. Trevisan, R. M. Gupta, A. Moisan, E. Banks, M. Friesen, R. T. Schinzel, F. Xia, A. Tang, Y. Xia, E. Figueroa, A. Wann, T. Ahfeldt, L. Daheron, F. Zhang, L. L. Rubin, L. F. Peng, R. T. Chung, K. Musunuru, C. A. Cowan, A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell (2013), doi:10.1016/j.stem.2012.11.011.
55. W. T. Hendriks, C. R. Warren, C. A. Cowan, Genome Editing in Human Pluripotent Stem Cells: Approaches, Pitfalls, and Solutions. Cell Stem Cell (2016), , doi:10.1016/j.stem.2015.12.002.
56. A. V. Anzalone, P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, J. M. Levy, P. J. Chen, C. Wilson, G. A. Newby, A. Raguram, D. R. Liu, Search-and-replace genome editing without double-strand breaks or donor DNA. Nature (2019), doi:10.1038/s41586-019-1711-4.

57. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. Science (1979) (2013), doi:10.1126/science.1231143.
58. A. M. Drozd, M. P. Walczak, S. Piaskowski, E. Stoczynska-Fidelus, P. Rieske, D. P. Grzela, Generation of human iPSCs from cells of fibroblastic and epithelial origin by means of the oriP/EBNA-1 episomal reprogramming system. Stem Cell Research and Therapy (2015), doi:10.1186/s13287-015-0112-3.
59. A. Radzisheuskaya, G. Le Bin Chia, R. L. Dos Santos, T. W. Theunissen, L. F. C. Castro, J. Nichols, J. C. R. Silva, A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nature Cell Biology (2013), doi:10.1038/ncb2742.
60. M. Ohgushi, M. Matsumura, M. Eiraku, K. Murakami, T. Aramaki, A. Nishiyama, K. Muguruma, T. Nakano, H. Suga, M. Ueno, T. Ishizaki, H. Suemori, S. Narumiya, H. Niwa, Y. Sasai, Molecular pathway and cell state responsible for dissociation-induced apoptosis in human pluripotent stem cells. Cell Stem Cell. 7, 225–239 (2010). doi: 10.1016/j.stem.2010.06.018
61. S. P. Medvedev, A. I. Shevchenko, S. M. Zakian, Induced Pluripotent Stem Cells: Problems and Advantages when Applying them in Regenerative Medicine. Acta Naturae (2010), doi:10.32607/20758251-2010-2-2-18-27.
62. K. Washburn, G. Halff, Hepatocellular carcinoma and liver transplantation. Current Opinion in Organ Transplantation (2011), doi:10.1097/MOT.0b013e3283465756.
63. V. Iansante, R. R. Mitry, C. Filippi, E. Fitzpatrick, A. Dhawan, Human hepatocyte transplantation for liver disease: Current status and future perspectives. Pediatric Research (2018), , doi:10.1038/pr.2017.284.
64. E. A. Jones, M. H. Abel, H. R. Woodland, The possible role of mesodermal growth factors in the formation of endoderm in Xenopus laevis. Roux's Archives of Developmental Biology (1993), doi:10.1007/BF02427884.
65. J. Cai, Y. Zhao, Y. Liu, F. Ye, Z. Song, H. Qin, S. Meng, Y. Chen, R. Zhou, X. Song, Y. Guo, M. Ding, H. Deng, Directed differentiation of human embryonic stem cells into functional hepatic cells. Hepatology (2007), doi:10.1002/hep.21582.
66. A. B. McLean, K. A. D'Amour, K. L. Jones, M. Krishnamoorthy, M. J. Kulik, D. M. Reynolds, A. M. Sheppard, H. Liu, Y. Xu, E. E. Baetge, S. Dalton, Activin A Efficiently Specifies Definitive Endoderm from Human Embryonic Stem Cells Only When Phosphatidylinositol 3-Kinase Signaling Is Suppressed. STEM CELLS (2007), doi:10.1634/stemcells.2006-0219.
67. J. K. Mfopou, B. Chen, I. Mateizel, K. Sermon, L. Bouwens, Noggin, Retinoids, and Fibroblast Growth Factor Regulate Hepatic or Pancreatic Fate of Human Embryonic Stem Cells. Gastroenterology (2010), doi:10.1053/j.gastro.2010.02.056.

68. A. Kamiya, T. Kinoshita, A. Miyajima, Oncostatin M and hepatocyte growth factor induce hepatic maturation via distinct signaling pathways. FEBS Letters (2001), doi:10.1016/S0014-5793(01)02140-8.
69. C. Schmidt, F. Bladt, S. Goedecke, V. Brinkmann, W. Zschiesche, M. Sharpe, E. Gherardi, S. Goedecke, Scatter factor/hepatocyte growth factor is essential for liver development. Nature (1995), doi:10.1038/373699a0.
70. K. Takayama, M. Inamura, K. Kawabata, K. Katayama, M. Higuchi, K. Tashiro, A. Nonaka, F. Sakurai, T. Hayakawa, M. Kusuda Furue, H. Mizuguchi, Efficient generation of functional hepatocytes from human embryonic stem cells and induced pluripotent stem cells by HNF4α transduction. Molecular Therapy (2012), doi:10.1038/mt.2011.234.
71. Q. J. Zhou, L. X. Xiang, J. Z. Shao, R. Z. Hu, Y. L. Lu, H. Yao, L. C. Dai, In vitro differentiation of hepatic progenitor cells from mouse embryonic stem cells induced by sodium butyrate. Journal of Cellular Biochemistry (2007), doi:10.1002/jcb.20970.
72. D. Fu, Y. Wakabayashi, Y. Ido, J. Lippincott-Schwartz, I. M. Arias, Regulation of bile canalicular network formation and maintenance by AMP-activated protein kinase and LKB1. Journal of Cell Science (2010), doi:10.1242/jcs.068098.
73. D. Fu, Y. Wakabayashi, J. Lippincott-Schwartz, I. M. Arias, Bile acid stimulates hepatocyte polarization through a cAMP-Epac-MEK-LKB1- AMPK pathway. Proc Natl Acad Sci U S A (2011), doi:10.1073/pnas.1018376108.
74. A. Woods, A. J. Heslegrave, P. J. Muckett, A. P. Levene, M. Clements, M. Mobberley, T. A. Ryder, S. Abu-Hayyeh, C. Williamson, R. D. Goldin, A. Ashworth, D. J. Withers, D. Carling, LKB1 is required for hepatic bile acid transport and canalicular membrane integrity in mice. Biochemical Journal (2011), doi:10.1042/BJ20101721.
75. L. Homolya, D. Fu, P. Sengupta, M. Jarnik, J. P. Gillet, L. Vitale-Cross, J. S. Gutkind, J. Lippincott-Schwartz, I. M. Arias, LKB1/AMPK and PKA Control ABCB11 trafficking and polarization in hepatocytes. PLoS ONE (2014), doi:10.1371/journal.pone.0091921.
76. D. Schuppan, Structure of the extracellular matrix in normal and fibrotic liver: Collagens and glycoproteins. Seminars in Liver Disease. 10 (1990), pp. 1–10. doi: 10.1055/s-2008-1040452
77. H. Kipp, N. Pichetshote, I. M. Arias, Transporters on demand. Intrahepatic pools of canalicular ATP binding cassette transporters in rat liver. Journal of Biological Chemistry. 276, 7218–7224 (2001). doi: 10.1074/jbc.M007794200
78. E. V. Polishchuk, M. Concilli, S. Iacobacci, G. Chesi, N. Pastore, P. Piccolo, S. Paladino, D. Baldantoni, S. C. D. vanIJzendoorn, J. Chan, C. J. Chang, A. Amoresano, F. Pane, P. Pucci, A. Tarallo, G. Parenti, N. Brunetti-Pierri, C. Settembre, A. Ballabio, R. S. Polishchuk, Wilson Disease Protein ATP7B Utilizes Lysosomal Exocytosis to Maintain Copper Homeostasis. Developmental Cell. 29, 686–700 (2014). doi: 10.1016/j.devcel.2014.04.033
79. Y. Wakabayashi, J. Lippincott-Schwartz, I. M. Arias, Intracellular trafficking of bile salt export pump (ABCB11) in polarized hepatic cells: Constitutive cycling between the

canalicular membrane and rab11-positive endosomes. Molecular Biology of the Cell. 15, 3485–3496 (2004). doi: 10.1091/mbc.E03-10-0737.
80. P. Gissen, I. M. Arias, Structural and functional hepatocyte polarity and liver disease. Journal of Hepatology (2015), , doi:10.1016/j.jhep.2015.06.015.
81. L. González-Mariscal, H. Gallego-Gutiérrez, L. González-González, C. Hernández-Guzmán, ZO-2 is a master regulator of gene expression, cell proliferation, cytoarchitecture, and cell size. International Journal of Molecular Sciences. 20 (2019), , doi:10.3390/ijms20174128.
82. M. Huerta, R. Muñoz, R. Tapia, E. Soto-Reyes, L. Ramírez, F. Recillas-Targa, L. González-Mariscal, E. López-Bayghen, Cyclin D1 is transcriptionally down-regulated by ZO-2 via an E box and the transcription factor c-Myc. Molecular Biology of the Cell. 18, 4826–4836 (2007). doi: 10.1091/mbc.E07-02-0109
83. M. Kazmierczak, S. L. Harris, P. Kazmierczak, P. Shah, V. Starovoytov, K. K. Ohlemiller, M. Schwander, Progressive hearing loss in mice carrying a mutation in Usp53. Journal of Neuroscience. 35, 15582–15598 (2015).
doi: 10.1523/JNEUROSCI.1965-15.2015
84. A. Raya-Sandino, A. Castillo-Kauil, A. Domínguez-Calderón, L. Alarcón, D. Flores-Benitez, F. Cuellar-Perez, B. López-Bayghen, B. Chávez-Munguía, J. Vázquez-Prado, L. González-Mariscal, Zonula occludens-2 regulates Rho proteins activity and the development of epithelial cytoarchitecture and barrier function. Biochimica et Biophysica Acta - Molecular Cell Research. 1864, 1714–1733 (2017).
doi: 10.1016/j.bbamcr.2017.05.016
85. T. Oka, E. Remue, K. Meerschaert, B. Vanloo, C. Boucherie, D. Gfeller, G. D. Bader, S. S. Sidhu, J. Vandekerckhove, J. Gettemans, M. Sudol, Functional complexes between YAP2 and ZO-2 are PDZ domain-dependent, and regulate YAP2 nuclear localization and signalling. Biochemical Journal (2010), doi:10.1042/BJ20100870.
86. E. Remue, K. Meerschaert, T. Oka, C. Boucherie, J. Vandekerckhove, M. Sudol, J. Gettemans, TAZ interacts with zonula occludens-1 and -2 proteins in a PDZ-1 dependent manner. FEBS Letters (2010), doi:10.1016/j.febslet.2010.09.020.
87. K. Gross-Steinmeyer, P. L. Stapleton, J. H. Tracy, T. K. Bammler, T. Lehman, S. C. Strom, D. L. Eaton, Influence of Matrigel-overlay on constitutive and inducible expression of nine genes encoding drug-metabolizing enzymes in primary human hepatocytes. Xenobiotica (2005), doi:10.1080/00498250500137427.
88. J. C. Y. Dunn, M. L. Yarmush, H. G. Koebe, R. G. Tompkins, Hepatocyte function and extracellular matrix geometry: long-term culture in a sandwich configuration. The FASEB Journal (1989), doi:10.1096/fasebj.3.2.2914628.
89. B. Swift, N. D. Pfeifer, K. L. R. Brouwer, Sandwich-cultured hepatocytes: An in vitro model to evaluate hepatobiliary transporter-based drug interactions and hepatotoxicity. Drug Metabolism Reviews (2010), , doi:10.3109/03602530903491881.

90. K. Vekemans, F. Braet, Structural and functional aspects of the liver and liver sinusoidal cells in relation to colon carcinoma metastasis. World Journal of Gastroenterology (2005), , doi:10.3748/wjg.v11.i33.5095.
91. J. A. Bonzo, K. Rose, K. Freeman, E. Deibert, K. B. Amaral, S. S. Ferguson, M. E. Andersen, R. P. Witek, E. L. LeCluyse, Differential Effects of Trovafloxacin on TNF-α and IL-6 Profiles in a Rat Hepatocyte–Kupffer Cell Coculture System. Applied In Vitro Toxicology (2015), doi:10.1089/aivt.2014.0004.
92. A. Asai, E. Aihara, C. Watson, R. Mourya, T. Mizuochi, P. Shivakumar, K. Phelan, C. Mayhew, M. Helmrath, T. Takebe, J. Wells, J. A. Bezerra, Paracrine signals regulate human liver organoid maturation from induced pluripotent stem cells. Development (Cambridge) (2017), doi:10.1242/dev.142794.
93. S. R. Khetani, S. N. Bhatia, Microscale culture of human liver cells for drug development. Nature Biotechnology (2008), doi:10.1038/nbt1361.
94. D. F. G. Hendriks, L. F. Puigvert, S. Messner, W. Mortiz, M. Ingelman-Sundberg, Hepatic 3D spheroid models for the detection and study of compounds with cholestatic liability. Scientific Reports (2016), doi:10.1038/srep35434.
95. C. C. Bell, D. F. G. Hendriks, S. M. L. Moro, E. Ellis, J. Walsh, A. Renblom, L. Fredriksson Puigvert, A. C. A. Dankers, F. Jacobs, J. Snoeys, R. L. Sison-Young, R. E. Jenkins, Å. Nordling, S. Mkrtchian, B. K. Park, N. R. Kitteringham, C. E. P. Goldring, V. M. Lauschke, M. Ingelman-Sundberg, Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease. Scientific Reports (2016), doi:10.1038/srep25187.
96. K. Takayama, K. Kawabata, Y. Nagamoto, K. Kishimoto, K. Tashiro, F. Sakurai, M. Tachibana, K. Kanda, T. Hayakawa, M. K. Furue, H. Mizuguchi, 3D spheroid culture of hESC/hiPSC-derived hepatocyte-like cells for drug toxicity testing. Biomaterials (2013), doi:10.1016/j.biomaterials.2012.11.029.
97. S. S. Ng, K. Saeb-Parsy, S. J. I. Blackford, J. M. Segal, M. P. Serra, M. Horcas-Lopez, D. Y. No, S. Mastoridis, W. Jassem, C. W. Frank, N. J. Cho, H. Nakauchi, J. S. Glenn, S. T. Rashid, Human iPS derived progenitors bioengineered into liver organoids using an inverted colloidal crystal poly (ethylene glycol) scaffold. Biomaterials (2018), doi:10.1016/j.biomaterials.2018.07.043.
98. L. A. Vernetti, N. Senutovitch, R. Boltz, R. DeBiasio, T. Ying Shun, A. Gough, D. L. Taylor, A human liver microphysiology platform for investigating physiology, drug safety, and disease models. Experimental Biology and Medicine. 241, 101 (2016). doi:10.1177/1535370215592121
99. M. N. Bin Ramli, Y. S. Lim, C. T. Koe, D. Demircioglu, W. Tng, K. A. U. Gonzales, C. P. Tan, I. Szczerbinska, H. Liang, E. L. Soe, Z. Lu, C. Ariyachet, K. M. Yu, S. H. Koh, L. P. Yaw, N. H. B. Jumat, J. S. Y. Lim, G. Wright, A. Shabbir, Y. Y. Dan, H. H. Ng, Y. S. Chan, Human Pluripotent Stem Cell-Derived Organoids as Models of Liver Disease. Gastroenterology (2020), doi:10.1053/j.gastro.2020.06.010.

100. A. Gough, A. Soto-Gutierrez, L. Vernetti, M. R. Ebrahimkhani, A. M. Stern, D. L. Taylor, Human biomimetic liver microphysiology systems in drug development and precision medicine. Nature Reviews Gastroenterology and Hepatology (2020), , doi:10.1038/s41575-020-00386-1.
101. A. P. Ramme, L. Koenig, T. Hasenberg, C. Schwenk, C. Magauer, D. Faust, A. K. Lorenz, A. C. Krebs, C. Drewell, K. Schirrmann, A. Vladetic, G. C. Lin, S. Pabinger, W. Neuhaus, F. Bois, R. Lauster, U. Marx, E. M. Dehne, Autologous induced pluripotent stem cell-derived four-organ-chip. Future Science OA (2019), doi:10.2144/fsoa-2019-0065.
102. K. Viravaidya, A. Sin, M. L. Shuler, Development of a Microscale Cell Culture Analog to Probe Naphthalene Toxicity. Biotechnology Progress (2004), doi:10.1021/bp0341996.
103. V. Starokozhko, R. Greupink, P. van de Broek, N. Soliman, S. Ghimire, I. A. M. de Graaf, G. M. M. Groothuis, Rat precision-cut liver slices predict drug-induced cholestatic injury. Archives of Toxicology (2017), doi:10.1007/s00204-017-1960-7.
104. R. Rao, G. Samak, Bile duct epithelial tight junctions and barrier function. Tissue Barriers (2013), doi:10.4161/tisb.25718.
105. Z. Hannoun, C. Steichen, N. Dianat, A. Weber, A. Dubart-Kupperschmitt, The potential of induced pluripotent stem cell derived hepatocytes. Journal of Hepatology (2016), , doi:10.1016/j.jhep.2016.02.025.
106. M. B. Esch, G. J. Mahler, T. Stokol, M. L. Shuler, Body-on-a-chip simulation with gastrointestinal tract and liver tissues suggests that ingested nanoparticles have the potential to cause liver injury. Lab on a Chip (2014), doi:10.1039/c4lc00371c.
107. S. Messner, I. Agarkova, W. Moritz, J. M. Kelm, Multi-cell type human liver microtissues for hepatotoxicity testing. Archives of Toxicology (2013), doi:10.1007/s00204-012-0968-2.
108. T. S. Jou, E. E. Schneeberger, W. J. Nelson, Structural and functional regulation of tight junctions by RhoA and Rac1 small GTPases. Journal of Cell Biology (1998), doi:10.1083/jcb.142.1.101.
109. A. W. Overeem, K. Klappe, S. Parisi, P. Klöters-Planchy, L. Mataković, M. du Teil Espina, C. A. Drouin, K. H. Weiss, S. C. D. van IJzendoorn, Pluripotent stem cell-derived bile canaliculi-forming hepatocytes to study genetic liver diseases involving hepatocyte polarity. Journal of Hepatology. 71, 344–356 (2019). doi: 10.1016/j.jhep.2019.03.031
110. N. S. Arden, A. C. Fisher, K. Tyner, L. X. Yu, S. L. Lee, M. Kopcha, Industry 4.0 for pharmaceutical manufacturing: Preparing for the smart factories of the future. Int J Pharm. 602 (2021), doi:10.1016/J.IJPHARM.2021.120554.
111. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I.

Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA. Science (1979). 337, 1190–1195 (2012). doi: 10.1126/science.1222794
112. Z. Julian, B. Dominique, K. Bat, M. Danuta, D. Peter, R. Johanna, L. Tsui, Identification of mutations in exons 1 through 8 of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Genomics. 10, 229–235 (1991). doi: 10.1016/0888-7543(91)90504-8
113. M. Koido, E. Kawakami, J. Fukumura, Y. Noguchi, M. Ohori, Y. Nio, P. Nicoletti, G. P. Aithal, A. K. Daly, P. B. Watkins, H. Anayama, Y. Dragan, T. Shinozawa, T. Takebe, Polygenic architecture informs potential vulnerability to drug-induced liver injury. Nature Medicine (2020), doi:10.1038/s41591-020-1023-0.
114. N. Chalasani, E. Björnsson, Risk Factors for Idiosyncratic Drug-Induced Liver Injury. Gastroenterology. 138 (2010), pp. 2246–2259. doi: 10.1053/j.gastro.2010.04.001
115. Y. S. Huang, H. Der Chern, W. J. Su, J. C. Wu, S. C. Chang, C. H. Chiang, F. Y. Chang, S. D. Lee, Cytochrome P450 2E1 genotype and the susceptibility to antituberculosis drug-induced hepatitis. Hepatology. 37, 924–930 (2003). doi: 10.1053/jhep.2003.50144
116. A. K. Daly, C. P. Day, Candidate gene case-control association studies: Advantages and potential pitfalls. Br J Clin Pharmacol. 52 (2001), pp. 489–499. doi: 10.1046/j.0306-5251.2001.01510.x
117. A. Raúl, L. Isabel, A. Anabel, G.-C. Miren, G.-R. Elena, B. Rafael, F. Carmen, P. Gloria, R. Manuel, C. Raquel, D. José, J. Manuel, R. Luis, N. Flor, M.-V. Rafael, N. José, S. Javier, de la C. Felipe, H. Ramón, HLA class II genotype influences the type of liver injury in drug-induced idiosyncratic liver disease. Hepatology. 39, 1603–1612 (2004). doi: 10.1002/HEP.20215
118. C. Stephens, M. Á. López-Nevot, F. Ruiz-Cabello, E. Ulzurrun, G. Soriano, M. Romero-Gómez, A. Moreno-Casares, M. I. Lucena, R. J. Andrade, HLA Alleles Influence the Clinical Signature of Amoxicillin-Clavulanate Hepatotoxicity. PLoS ONE. 8 (2013), doi:10.1371/journal.pone.0068111.
119. F. Abascal, R. Acosta, N. J. Addleman, J. Adrian, V. Afzal, B. Aken, J. A. Akiyama, O. Al Jammal, H. Amrhein, S. M. Anderson, G. R. Andrews, I. Antoshechkin, K. G. Ardlie, J. Armstrong, M. Astley, B. Banerjee, A. A. Barkal, I. H. A. Barnes, I. Barozzi, D. Barrell, G. Barson, D. Bates, U. K. Baymuradov, C. Bazile, M. A. Beer, S. Beik, M. A. Bender, R. Bennett, L. P. B. Bouvrette, B. E. Bernstein, A. Berry, A. Bhaskar, A. Bignell, S. M. Blue, D. M. Bodine, C. Boix, N. Boley, T. Borrman, B. Borsari, A. P. Boyle, L. A. Brandsmeier, A. Breschi, E. H. Bresnick, J. A. Brooks, M. Buckley, C. B. Burge, R. Byron, E. Cahill, L. Cai, L. Cao, M. Carty, R. G. Castanon, A. Castillo, H. Chaib, E. T. Chan, D. R. Chee, S. Chee, H. Chen, H. Chen, J. Y. Chen, S. Chen, J. M. Cherry, S. B. Chhetri, J. S. Choudhary, J. Chrast, D. Chung, D. Clarke, N. A. L. Cody, C. J. Coppola, J. Coursen, A. M. D'Ippolito, S. Dalton, C. Danyko, C. Davidson, J. Davila-Velderrain, C. A. Davis, J. Dekker, A. Deran, G. DeSalvo, G. Despacio-Reyes, C. N. Dewey, D. E. Dickel, M. Diegel, M. Diekhans, V. Dileep, B. Ding, S.

Djebali, A. Dobin, D. Dominguez, S. Donaldson, J. Drenkow, T. R. Dreszer, Y. Drier, M. O. Duff, D. Dunn, C. Eastman, J. R. Ecker, M. D. Edwards, N. El-Ali, S. I. Elhajjajy, K. Elkins, A. Emili, C. B. Epstein, R. C. Evans, I. Ezkurdia, K. Fan, P. J. Farnham, N. Farrell, E. A. Feingold, A. M. Ferreira, K. Fisher-Aylor, S. Fitzgerald, P. Flicek, C. S. Foo, K. Fortier, A. Frankish, P. Freese, S. Fu, X. D. Fu, Y. Fu, Y. Fukuda-Yuzawa, M. Fulciniti, A. P. W. Funnell, I. Gabdank, T. Galeev, M. Gao, C. G. Giron, T. H. Garvin, C. A. Gelboin-Burkhart, G. Georgolopoulos, M. B. Gerstein, B. M. Giardine, D. K. Gifford, D. M. Gilbert, D. A. Gilchrist, S. Gillespie, T. R. Gingeras, P. Gong, A. Gonzalez, J. M. Gonzalez, P. Good, A. Goren, D. U. Gorkin, B. R. Graveley, M. Gray, J. F. Greenblatt, E. Griffiths, M. T. Groudine, F. Grubert, M. Gu, R. Guigó, H. Guo, Y. Guo, Y. Guo, G. Gursoy, M. Gutierrez-Arcelus, J. Halow, R. C. Hardison, M. Hardy, M. Hariharan, A. Harmanci, A. Harrington, J. L. Harrow, T. B. Hashimoto, R. D. Hasz, M. Hatan, E. Haugen, J. E. Hayes, P. He, Y. He, N. Heidari, D. Hendrickson, E. F. Heuston, J. A. Hilton, B. C. Hitz, A. Hochman, C. Holgren, L. Hou, S. Hou, Y. H. E. Hsiao, S. Hsu, H. Huang, T. J. Hubbard, J. Huey, T. R. Hughes, T. Hunt, S. Ibarrientos, R. Issner, M. Iwata, O. Izuogu, T. Jaakkola, N. Jameel, C. Jansen, L. Jiang, P. Jiang, A. Johnson, R. Johnson, I. Jungreis, M. Kadaba, M. Kasowski, M. Kasparian, M. Kato, R. Kaul, T. Kawli, M. Kay, J. C. Keen, S. Keles, C. A. Keller, D. Kelley, M. Kellis, P. Kheradpour, D. S. Kim, A. Kirilusha, R. J. Klein, B. Knoechel, S. Kuan, M. J. Kulik, S. Kumar, A. Kundaje, T. Kutyavin, J. Lagarde, B. R. Lajoie, N. J. Lambert, J. Lazar, A. Y. Lee, D. Lee, E. Lee, J. W. Lee, K. Lee, C. S. Leslie, S. Levy, B. Li, H. Li, N. Li, X. Li, Y. I. Li, Y. Li, Y. Li, Y. Li, J. Lian, M. W. Libbrecht, S. Lin, Y. Lin, D. Liu, J. Liu, P. Liu, T. Liu, X. S. Liu, Y. Liu, Y. Liu, M. Long, S. Lou, J. Loveland, A. Lu, Y. Lu, E. Lécuyer, L. Ma, M. Mackiewicz, B. J. Mannion, M. Mannstadt, D. Manthravadi, G. K. Marinov, F. J. Martin, E. Mattei, K. McCue, M. McEown, G. McVicker, S. K. Meadows, A. Meissner, E. M. Mendenhall, C. L. Messer, W. Meuleman, C. Meyer, S. Miller, M. G. Milton, T. Mishra, D. E. Moore, H. M. Moore, J. E. Moore, S. H. Moore, J. Moran, A. Mortazavi, J. M. Mudge, N. Munshi, R. Murad, R. M. Myers, V. Nandakumar, P. Nandi, A. M. Narasimha, A. K. Narayanan, H. Naughton, F. C. P. Navarro, P. Navas, J. Nazarovs, J. Nelson, S. Neph, F. J. Neri, J. R. Nery, A. R. Nesmith, J. S. Newberry, K. M. Newberry, V. Ngo, R. Nguyen, T. B. Nguyen, T. Nguyen, A. Nishida, W. S. Noble, C. S. Novak, E. M. Novoa, B. Nuñez, C. W. O'Donnell, S. Olson, K. C. Onate, E. Otterman, H. Ozadam, M. Pagan, T. Palden, X. Pan, Y. Park, E. C. Partridge, B. Paten, F. Pauli-Behn, M. J. Pazin, B. Pei, L. A. Pennacchio, A. R. Perez, E. H. Perry, D. D. Pervouchine, N. N. Phalke, Q. Pham, D. H. Phanstiel, I. Plajzer-Frick, G. A. Pratt, H. E. Pratt, S. Preissl, J. K. Pritchard, Y. Pritykin, M. J. Purcaro, Q. Qin, G. Quinones-Valdez, I. Rabano, E. Radovani, A. Raj, N. Rajagopal, O. Ram, L. Ramirez, R. N. Ramirez, D. Rausch, S. Raychaudhuri, J. Raymond, R. Razavi, T. E. Reddy, T. M. Reimonn, B. Ren, A. Reymond, A. Reynolds, S. K. Rhie, J. Rinn, M. Rivera, J. C. Rivera-Mulia, B. Roberts, J. M. Rodriguez, J. Rozowsky, R. Ryan, E. Rynes, D. N. Salins, R. Sandstrom, T. Sasaki, S. Sathe, D. Savic, A. Scavelli, J. Scheiman, C. Schlaffner, J. A. Schloss, F. W. Schmitges, L. H. See, A. Sethi, M. Setty, A. Shafer, S. Shan, E. Sharon, Q. Shen, Y. Shen, R. I. Sherwood, M. Shi, S. Shin, N.

Shoresh, K. Siebenthall, C. Sisu, T. Slifer, C. A. Sloan, A. Smith, V. Snetkova, M. P. Snyder, D. V. Spacek, S. Srinivasan, R. Srivas, G. Stamatoyannopoulos, J. A. Stamatoyannopoulos, R. Stanton, D. Steffan, S. Stehling-Sun, J. S. Strattan, A. Su, B. Sundararaman, M. M. Suner, T. Syed, M. Szynkarek, F. Y. Tanaka, D. Tenen, M. Teng, J. A. Thomas, D. Toffey, M. L. Tress, D. E. Trout, G. Trynka, J. Tsuji, S. A. Upchurch, O. Ursu, B. Uszczynska-Ratajczak, M. C. Uziel, A. Valencia, B. Van Biber, A. G. van der Velde, E. L. Van Nostrand, Y. Vaydylevich, J. Vazquez, A. Victorsen, J. Vielmetter, J. Vierstra, A. Visel, A. Vlasova, C. M. Vockley, S. Volpi, S. Vong, H. Wang, M. Wang, Q. Wang, R. Wang, T. Wang, W. Wang, X. Wang, Y. Wang, N. K. Watson, X. Wei, Z. Wei, H. Weisser, S. M. Weissman, R. Welch, R. E. Welikson, Z. Weng, H. J. Westra, J. W. Whitaker, C. White, K. P. White, A. Wildberg, B. A. Williams, D. Wine, H. N. Witt, B. Wold, M. Wolf, J. Wright, R. Xiao, X. Xiao, J. Xu, J. Xu, K. K. Yan, Y. Yan, H. Yang, X. Yang, Y. W. Yang, G. G. Yardımcı, B. A. Yee, G. W. Yeo, T. Young, T. Yu, F. Yue, C. Zaleski, C. Zang, H. Zeng, W. Zeng, D. R. Zerbino, J. Zhai, L. Zhan, Y. Zhan, B. Zhang, J. Zhang, J. Zhang, K. Zhang, L. Zhang, P. Zhang, Q. Zhang, X. O. Zhang, Y. Zhang, Z. Zhang, Y. Zhao, Y. Zheng, G. Zhong, X. Q. Zhou, Y. Zhu, J. Zimmerman, M. P. Snyder, T. R. Gingeras, J. E. Moore, Z. Weng, M. B. Gerstein, B. Ren, R. C. Hardison, J. A. Stamatoyannopoulos, B. R. Graveley, E. A. Feingold, M. J. Pazin, M. Pagan, D. A. Gilchrist, B. C. Hitz, J. M. Cherry, B. E. Bernstein, E. M. Mendenhall, D. R. Zerbino, A. Frankish, P. Flicek, R. M. Myers, Perspectives on ENCODE. Nature. 583 (2020), pp. 693–698. doi: 10.1038/s41586-020-2449-8

120. O.Shanta, A. Noor, J. Sebat, The effects of common structural variants on 3D chromatin structure. BMC Genomics. 21 (2020), doi: 10.1186/S12864-020-6516-1

121. M. Creyghton, A. Cheng, G. Welstead, T. Kooistra, B. Carey, E. Steine, J. Hanna, M. Lodato, G. Frampton, PA. Sharp, LA. Boyer, RA. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 107, 21931–21936 (2010). doi: 10.1073/PNAS.1016071107

122. F. Shaliu, W. Qin, M. Jill, P. Michael, P. Henry, F. Kaili, G. Cuihua, J. Cizhong, Z. Ruixin, K. Anshul, L. Aiping, W. Zhiping, Differential analysis of chromatin accessibility and histone modifications for predicting mouse developmental enhancers. Nucleic Acids Res. 46, 11184–11201 (2018). doi: 10.1093/NAR/GKY753

123. M. Glenn A, E. Sara K, G. Michael R, Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 7, 29–59 (2006). doi: 10.1146/ANNUREV.GENOM.7.080505.115623

124. S. Serge, B. Paul, B. Douglas, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A. 103, 1412–1417 (2006). doi: 10.1073/PNAS.0510310103

125. J. Kamel, B. Giorgio, Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene. 333, 143–149 (2004). doi: 10.1016/J.GENE.2004.02.043

126. S. Ilakya, T. Anu, G. Kristina, L. Rodríguez. M, K. Arthur, M. Zong, K. Dorota, Õ. Kadri, Ö. Tiit, R. Aarthi, L. Oscar, M. Pierre, J. D. Ashik, M. Ville, P. Calvin, L. Anna, L.

Aldons, H. Sami, R. Casey, P. Jussi, P. Päivi, K. Minna, Integrative analysis of liver-specific non-coding regulatory SNPs associated with the risk of coronary artery disease. Am J Hum Genet. 108, 411–430 (2021). doi: 10.1016/J.AJHG.2021.02.006
127. B. Ewan, S. John A, D. Anindya, G. Roderic, G. Thomas R, M. Elliott H, W. Zhiping, Snyder. Michael, D. Emmanouil T, T. Robert E, K. Michael S, T. Christopher M, N. Shane, K. Christoph M, A. Saurabh, Malhotra. Ankit, A. Ivan, G. Jason A, A. Robert M, F. Paul, B. Patrick J, C. Hua, C. Nigel P, C. Gayle K, D. Sean, D. Nathan , D. Pawandeep, D. Shane C, D. Michael O, F. Heike, G. Paul G, G. Jeff, H. Michael, H. Andrew, H. Richard, J. Keith D, Johnson. Brett E, J. Ericka M, F. Tristan T, R. Elizabeth R, K. Neerja, L. Kirsten, L. Gregory C, N. Patrick A, N. Fidencio, P. Stephen C J, S. Peter J, S. Richard, S. Anthony, V. David, W. Molly, Wilcox. Sarah, Yu. Man, C. Francis S, D. Job, L. Jason D, T. Thomas D, C. Gregory E, S. Shamil, N. William S, D. Ian, D. France, R. Alexandre, K, Philipp R, Joel, Z. Deyou, C. Robert, F. Adam, H. Jennifer, G. Srinka, S. Albin, H. Ivo L, B. Robert, K. Damian, D. Sujit, C. Jill, H. Heather A, S. Edward A, L. Julien, A, Josep F, S. Atif, F. Christoph, F. Claudia, H, Jana, L. Manja, M. Kristin, T. Andrea, W. Stefan, K. Jan, E. Olof, P. Jakob S, H. Nancy, T. Ruth, S. David, M. Nicholas, D. Mark C, Thomas. Daryl J, W. Matthew T, G. James, D. Jorg, B. Ian, Zhao. Xiaodong, S. K G, Sung. Wing Kin, Ooi. Hong Sain, Chiu. Kuo Ping, F. Sylvain, A. Tyler, B. Michael, P. Lior, T. Michael L, V. Alfonso, C. Siew Who, C, Chiou Yu, U. Catherine, M. Caroline, W. Carine, C. Evelyn, C. Taane G, B. James B, G Madhavan, P. Sandeep, T. Hari, C. Jacqueline, H, Charlotte N, K. Chikatoshi, K, Jun, N. Ugrappa, W. Jiaqian, L. Zheng, L. Jin, N. Peter, Z. Xueqing, B. Peter, M. John S, C. Piero, H. Yoshihide, W. Sherman, H. Tim, M. Richard M, R. Jane, S. Peter F, L. Todd M, W. Chia Lin, R. Yijun, S. Kevin, G. Mark, A. Stylianos E, Fu. Yutao, G. Eric D, K. Ulaş, S. Adam, T. James, L. Laura A, W. Kris A, G. Peter J, F. Elise A, G. Mark S, C. Gregory M, A. George, D. Colin N, H. Minmei, N. Sergey, M. Juan I, W. Simon, P. Fabio, M. Tim, H. Haiyan, Z. Nancy R, H. Ian, M. James C, U. Abel, P. Benedict, S. Michael, C. Deanna, R. Kate, K. W James, S. Eric A, B, Serafim, G. Nick, H. David, M. Webb, S. Arend, T. Nathan D, Z. Zhengdong D, B. Leah, S. Rhona, K. David C, A. Adam, E. Stefan, B. Mark C, K. Jonghwan, B. Akshay A, J. Nan, L. Jun, Y. Fei, V. Vinsensius B, L. Ng, Patrick Y, Annie, M, Zarmik, Z. Zhou, Xu. Xiaoqin, S. Sharon, O. Matthew J, I. David, S. Michael A, R. Todd A, M. Kyle J, R, Alvaro, W. Ola, K. Jan, F. Joanna C, C. Phillippe, B. Alexander W, D. Oliver M, E. Peter D, L, Cordelia F. Nix, E. Ghia, H. Stephen, U. Alexander E, K. Peter, C. Sara Van, H. Nate, K. Tae Hoon, W. Kun, Q. Chunxu, H. Gary, L. Rosa, G. Christopher K, M. Geoff, Aldred, Shelley Force, Cooper, Sara J, Halees, Anason, L. Jane M, S. Hennady P, Z. Xiaoling, X. Mousheng, H. Jaafar N S, Y. Yong, I. Vishwanath R, G. Roland D, W. Claes, F. Peggy J, R. Bing, H. Rachel A, H. Angie S, T. Heather, C. Hiram, H. Jennifer.T, A. Barber, K. Robert M, K, Donna, A, Lluis, B. Christine P, B. Paul I W De, K. Andrew D, Lopez-Bigas, Nuria, M, Joel D and Stranger, Barbara E, W. Abigail, D. Eugene, D. Antigone, E. Eduardo, H. Julian, Z. Michael C, A. E. Xavier, B. Gerard G,G, Xiaobin, Nancy F and Idol, Valerie V B, Maskeri, Baishali, McDowell, Jennifer C, Park, Morgan,T. Pamela J, Y. Alice C, B. Robert W, Muzny, Donna

M S. Erica, W. David A, Worley, Kim C, Jiang, Huaiyang, Weinstock, George M, Gibbs, Richard A, Graves, Tina and Fulton, Robert, Mardis, Elaine R, Wilson, Richard K, Clamp, Michele, Cuff, James, Gnerre, Sante, Jaffe, David B,Chang, Jean L, Lindblad-Toh, Kerstin, Lander, Eric S, Koriabine, Maxim, Nefedov, Mikhail, Osoegawa, Kazutoyo and Yoshinaga, Yuko and Zhu, Baoli and Jong, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 447, 799–816 (2007). doi: 10.1038/nature05874

128. L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, L. S. Qi, XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 154, 442 (2013). doi: 10.1016/j.cell.2013.06.044

129. L.Qi, M.Larson, L.Gilbert, J.Doudna, J.Weissman, A.Arkin, W.Lim, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 152, 1173–1183 (2013). doi: 10.1016/J.CELL.2013.02.022

130. L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell. 152, 1173–1183 (2013). doi: 10.1016/J.CELL.2013.02.022

131. C. F. Peddle, L. E. Fry, M. E. McClements, R. E. Maclaren, Crispr interference–potential application in retinal disease. International Journal of Molecular Sciences. 21 (2020), , doi:10.3390/ijms21072329.

132. Y. Zhang, C. Yin, T. Zhang, F. Li, W. Yang, R. Kaminski, P. R. Fagan, R. Putatunda, W. bin Young, K. Khalili, W. Hu, CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. Scientific Reports 2015 5:1. 5, 1–14 (2015). doi: 10.1038/srep16277

133. M.Cullen, J. Noble, Erlich.H, Thorpe. K, Beck. S, Klitz. W, Trowsdale. J, Carrington. M, Characterization of recombination in the HLA class II region. Am J Hum Genet. 60, 397–407 (1997). Issn: 0002-9297

134. M.Machiela, Chanock. SJ, LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 31, 3555–3557 (2015). doi: 10.1093/BIOINFORMATICS/BTV402

135. A.Quinlan, Hall.IM, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841–842 (2010). doi: 10.1093/BIOINFORMATICS/BTQ033

136. E. T. Cirulli, P. Nicoletti, K. Abramson, R. J. Andrade, E. S. Bjornsson, N. Chalasani, R. J. Fontana, P. Hallberg, Y. J. Li, M. I. Lucena, N. Long, M. Molokhia, M. R. Nelson, J. A. Odin, M. Pirmohamed, T. Rafnar, J. Serrano, K. Stefánsson, A. Stolz, A. K. Daly, G. P. Aithal, P. B. Watkins, F. Bessone, E. Bjornsson, I. Cascorbi, J. F. Dillon, C. P. Day, P. Hallberg, N. Hernández, L. Ibanez, G. A. Kullak-Ublick, T. Laitinen, D. Larrey, A. Maitland-van der Zee, J. H. Martin, D. Menzies, S. Qin, M. Wadelius, A Missense Variant in PTPN22

is a Risk Factor for Drug-induced Liver Injury. Gastroenterology. 156, 1707-1716.e2 (2019). doi:10.1093/BIOINFORMATICS/BTQ033
137. P. Nicoletti, G. P. Aithal, E. S. Bjornsson, R. J. Andrade, A. Sawle, M. Arrese, H. X. Barnhart, E. Bondon-Guitton, P. H. Hayashi, F. Bessone, A. Carvajal, I. Cascorbi, E. T. Cirulli, N. Chalasani, A. Conforti, S. A. Coulthard, M. J. Daly, C. P. Day, J. F. Dillon, R. J. Fontana, J. I. Grove, P. Hallberg, N. Hernández, L. Ibáñez, G. A. Kullak-Ublick, T. Laitinen, D. Larrey, M. I. Lucena, A. H. Maitland-van der Zee, J. H. Martin, M. Molokhia, M. Pirmohamed, E. E. Powell, S. Qin, J. Serrano, C. Stephens, A. Stolz, M. Wadelius, P. B. Watkins, A. Floratos, Y. Shen, M. R. Nelson, T. J. Urban, A. K. Daly, Association of Liver Injury From Specific Drugs, or Groups of Drugs, With Polymorphisms in HLA and Other Genes in a Genome-Wide Association Study. Gastroenterology (2017), doi:10.1053/j.gastro.2016.12.016.
138. S. Suvichapanich, S. Wattanapokayakit, T. Mushiroda, H. Yanai, C. Chuchottawon, T. Kantima, S. Nedsuwan, W. Suwankesawong, C. Sonsupap, R. Pannarunothai, S. Tumpattanakul, W. Bamrungram, A. Chaiwong, S. Mahasirimongkol, S. Mameechai, W. Panthong, N. Klungtes, A. Munsoo, U. Chauychana, M. Maneerat, K. Fukunaga, Y. Omae, K. Tokunaga, Genomewide association study confirming the association of nat2 with susceptibility to antituberculosis drug-induced liver injury in Thai patients. Antimicrobial Agents and Chemotherapy (2019), doi:10.1128/AAC.02692-18.
139. T. J. Urban, Y. Shen, A. Stolz, N. Chalasani, R. J. Fontana, J. Rochon, D. Ge, K. V. Shianna, A. K. Daly, M. I. Lucena, M. R. Nelson, M. Molokhia, G. P. Aithal, A. Floratos, I. Pe'Er, J. Serrano, H. Bonkovsky, T. J. Davern, W. M. Lee, V. J. Navarro, J. A. Talwalkar, D. B. Goldstein, P. B. Watkins, Limited contribution of common genetic variants to risk for liver injury due to a variety of drugs. Pharmacogenetics and Genomics. 22, 784–795 (2012). doi: 10.1097/FPC.0B013E3283589A76
140. K. Kowalec, G. E. B. Wright, B. I. Drögemöller, F. Aminkeng, A. P. Bhavsar, E. Kingwell, E. M. Yoshida, A. Traboulsee, R. A. Marrie, M. Kremenchutzky, T. L. Campbell, P. Duquette, N. Chalasani, M. Wadelius, P. Hallberg, Z. Xia, P. L. De Jager, J. C. Denny, M. F. Davis, C. J. D. Ross, H. Tremlett, B. C. Carleton, Common variation near IRF6 is associated with IFN-β-induced liver injury in multiple sclerosis. Nature Genetics. 50 (2018), pp. 1081–1085. doi: 10.1038/S41588-018-0168-Y
141. L. R. Parham, L. P. Briley, L. Li, J. Shen, P. J. Newcombe, K. S. King, A. J. Slater, A. Dilthey, Z. Iqbal, G. McVean, C. J. Cox, M. R. Nelson, C. F. Spraggs, Comprehensive genome-wide evaluation of lapatinib-induced liver injury yields a single genetic signal centered on known risk allele HLA-DRB1∗07:01. Pharmacogenomics Journal. 16, 180–185 (2016). doi: 10.1038/TPJ.2015.40
142. Daly. AK, Donaldson. Bhatnagar.P, Y.Shen, Pe'er,I, Floratos.A, Daly. MJ, Goldstein.DB, John. S, Nelson. MR, Graham. J, Park. BK, Dillon. JF, Bernal. W, Cordell. HJ, Pirmohamed. M, Aithal. GP, Day. CP, HLA-B*5701 genotype is a major determinant of

drug-induced liver injury due to flucloxacillin. Nat Genet. 41, 816–819 (2009). doi: 10.1038/NG.379
143. T. J. Urban, Y. Shen, A. Stolz, N. Chalasani, R. J. Fontana, J. Rochon, D. Ge, K. V. Shianna, A. K. Daly, M. I. Lucena, M. R. Nelson, M. Molokhia, G. P. Aithal, A. Floratos, I. Pe'Er, J. Serrano, H. Bonkovsky, T. J. Davern, W. M. Lee, V. J. Navarro, J. A. Talwalkar, D. B. Goldstein, P. B. Watkins, Limited contribution of common genetic variants to risk for liver injury due to a variety of drugs. Pharmacogenetics and Genomics. 22, 784–795 (2012). doi: 10.1097/FPC.0B013E3283589A76
144. P. Nicoletti, G. P. Aithal, E. S. Bjornsson, R. J. Andrade, A. Sawle, M. Arrese, H. X. Barnhart, E. Bondon-Guitton, P. H. Hayashi, F. Bessone, A. Carvajal, I. Cascorbi, E. T. Cirulli, N. Chalasani, A. Conforti, S. A. Coulthard, M. J. Daly, C. P. Day, J. F. Dillon, R. J. Fontana, J. I. Grove, P. Hallberg, N. Hernández, L. Ibáñez, G. A. Kullak-Ublick, T. Laitinen, D. Larrey, M. I. Lucena, A. H. Maitland-van der Zee, J. H. Martin, M. Molokhia, M. Pirmohamed, E. E. Powell, S. Qin, J. Serrano, C. Stephens, A. Stolz, M. Wadelius, P. B. Watkins, A. Floratos, Y. Shen, M. R. Nelson, T. J. Urban, A. K. Daly, Association of Liver Injury From Specific Drugs, or Groups of Drugs, With Polymorphisms in HLA and Other Genes in a Genome-Wide Association Study. Gastroenterology. 152, 1078–1089 (2017). doi: 10.1053/j.gastro.2016.12.016
145. L.Martens, R.Verbeeck, F. Callens, Matthys. PF, Boesman. ER, Dermaut. LR, Apatite crystallite alignment in sound human tooth enamel studied by E.S.R. J Biol Buccale. 13, 347–353 (1985). ISSN: 0301-3952
146. M. Claussnitzer, Cho. JH, Collins. R, Cox. NJ, Dermitzakis. ET, Hurles. ME, K. S, Kathiresan. EE, Lindgren. CM, MacArthur. DG, North. KN, Plon. SE, Rehm. HL, Risch. N, Rotimi. CN, Shendure. J, Soranzo. N, McCarthy. MI, A brief history of human disease genetics. Nature. 577, 179–189 (2020). doi: 10.1038/S41586-019-1879-7
147. W. Pichler, Adam. J, Watkins. S, Wuillemin. N, Yun. J, Yerly. D, Drug Hypersensitivity: How Drugs Stimulate T Cells via Pharmacological Interaction with Immune Receptors. Int Arch Allergy Immunol. 168, 13–24 (2015). doi: 10.1159/000441280
148. A. Kindmark, Jawaid. A, Harbron. CG, Barratt. BJ, Bengtsson. OF, Andersson. TB, Carlsson. S, Cederbrant. KE, Gibson. NJ, Armstrong. M, Lagerström-Fermér. ME, Dellsén. A , Brown. EM, Thornton. M, Dukes. C, Jenkins. SC, Firth. MA, Harrod. GO, Pinel. TH, Billing-Clason. SM, Cardon. LR, March. RE, Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. Pharmacogenomics J. 8, 186–195 (2008). doi:10.1038/SJ.TPJ.6500458
149. R. Scott, L. Scott, L. Marullo, KJ, Gaulton. M. Kaakinen, N. Pervjakova, TH. Pers, AD. Johnson, JD. Eicher, AU.Jackson, T. Ferreira, Y. Lee, C. Ma, V. Steinthorsdottir, G. Thorleifsson, L. Qi, NR. Van Zuydam, A. Mahajan, H. Chen, P. Almgren, BF. Voight, H. Grallert, JS. Ried,NW. Rayner, N. Robertson, LC. Karssen, EM. van Leeuwen, SM.

Willems, C. Fuchsberger, P. Kwan, TM. Teslovich, P. Chanda, M. Li, Y. Lu, C. Dina, D. Thuillier, L. Yengo, L. Jiang, T. Sparso, HA. Kestler, H. Chheda, L. Eisele, S. Gustafsson, RJ, Strawbridge, R. Benediktsson, AB. Hreidarsson, A. Kong, G. Sigurðsson, ND. Kerrison, J. Luan, L. Liang, T. Meitinger, M. Roden, B. Thorand, T. Esko, E. Mihailov, C. Fox, CT. Liu, D. Rybin, B. Isomaa, V. Lyssenko, T. Tuomi, DJ. Couper, JS. Pankow, N. Grarup, CT. Have, A. Linneberg, MC. Cornelis, RM. van Dam, DJ. Hunter, P. Kraft, Q. Sun, S. Edkins, KR. Owen, JRB. Perry, AR. Wood, E. Zeggini, J. Tajes-Fernandes, GR. Abecasis, LL. Bonnycastle, PS. Chines, HM. Stringham, HA. Koistinen, L. Kinnunen, B. Sennblad, S. Pechlivanis, D. Baldassarre, K. Gertow, SE. Humphries, E. Tremoli, N. Klopp, J. Meyer, G. Steinbach, R. Wennauer, JG. Eriksson, S. Männist, L. Peltonen, E. Tikkanen, G. Charpentier, E. Eury, S. Lobbens, B. Gigante, K. Leander, O. McLeod, EP. Bottinger, O. Gottesman, D. Ruderfer, and M, P. Kovacs, A. Tonjes, NM. Maruthur, C. Scapoli, R. Erbel, S. Moebus, U. de Faire, A. Hamsten, M. Stumvoll, P. Deloukas, PJ. Donnelly, TM. Frayling, AT. Hattersley, S. Ripatti, V. Salomaa, NL. Pedersen, BO. Boehm, RN. Bergman, FS. Collins, KL. Mohlke, J. Tuomilehto, T. Hansen, O. Pedersen, I. Barroso, L. Lannfelt, E. Ingelsson, L. Lind, CM. Lindgren, S. Cauchi, P. Froguel, RJF. Loos, B. Balkau, H. Boeing, PW. Franks, A. Barricarte Gurrea, D. Palli, YT. van der Schouw, D. Altshuler, LC. Groop, C. Langenberg, NJ. Wareham, E. Sijbrands, CM. van Duijn, JC. Florez, JB. Meigs, E. Boerwinkle, C. Gieger, K. Strauch, A. Metspalu, AD. Morris, CAN. Palmer, FB. Hu, U. Thorsteinsdottir, K. Stefansson, J. Dupuis, AP. Morris, M. Boehnke, MI. McCarthy, I. Prokopenko, An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 66, 2888–2902 (2017). doi: 10.2337/DB16-1253

150. B. Kunkle, B. Grenier-Boley, R. Sims, J. Bis, V. Damotte, A.Naj, A. Boland, M. Vronskaya, SJ. van der Lee, A. Amlie-Wolf, C. Bellenguez, A. Frizatti, V. Chouraki, ER. Martin, K. Sleegers, N. Badarinarayan, J. Jakobsdottir, KL. Hamilton-Nelson, S. Moreno-Grau, R. Olaso, R. Raybould, Y. Chen, A. Kuzma, M. Hiltunen, T. Morgan, S. Ahmad, BN. Vardarajan, J. Epelbaum, P. Hoffmann, M. Boada, GW. Beecham, JG. Garnier, D. Harold, AL. Fitzpatrick, O. Valladares, ML. Moutet, A. Gerrish, AV. Smith, L. Qu, D. Bacq, N. Denning, X. Jian, Y. Zhao, M. Del Zompo, NC. Fox, SH. Choi, I. Mateo, JT. Hughes, HH. Adams, J. Malamon, F. Sanchez-Garcia, Y. Patel, JA. Brody, BA. Dombroski, MCD. Naranjo, M. Daniilidou, G. Eiriksdottir, S. Mukherjee, D. Wallon, J. Uphill, T. Aspelund, LB. Cantwell, F. Garzia, D.Galimberti, E. Hofer, M. Butkiewicz, B. Fin, E. Scarpini, C, Sarnowski, WS. Bush, S. Meslage, J. Kornhuber, CC. White, Y. Song, RC. Barber, S. Engelborghs, S. Sordon, D. Voijnovic, PM. Adams, R. Vandenberghe, M, Mayhaus, LA. Cupples, MS. Albert, PP. De Deyn, W. Gu, JJ. Himali, D. Beekly, A. Squassina, AM. Hartmann, A. Orellana, D. Blacker, E. Rodriguez-Rodriguez, S. Lovestone, ME. Garcia, RS. Doody, C. Munoz-Fernadez, R. Sussams, H. Lin, TJ. Fairchild, YA. Benito, C. Holmes, H. Karamuji, MP. Frosch, H. Thonberg, W. Maier, G. Roshchupkin, B. Ghetti, V. Giedraitis, A. Kawalia, S. Li, RM. Huebinger, L. Kilander, S. Moebus, I. Herndez, MI. Kamboh, R. Brundin, J. Turton, Q. Yang, MJ. Katz, L. Concari, J. Lord, AS. Beiser, CD. Keene, S.

Helisalmi, I. Kloszewska, WA. Kukull, AM. Koivisto, A. Lynch, L. Tarraga, EB. Larson, A. Haapasalo, B. Lawlor, TH. Mosley, RB. Lipton, V. Solfrizzi, M. Gill, WT. Longstreth, TJ. Montine, V. Frisardi, M. Diez-Fairen, F. Rivadeneira, RC. Petersen, V. Deramecourt, I. Alvarez, F. Salani, A. Ciaramella, E. Boerwinkle, EM. Reiman, N. Fievet, JI. Rotter, JS. Reisch, O. Hanon, C. Cupidi, AG. Andre, DR. Royall, C. Dufouil, RG. Maletta, I. de Rojas, M. Sano, A. Brice, R. Cecchetti, PS. George-Hyslop, K. Ritchie, M. Tsolaki, DW. Tsuang, B. Dubois, D. Craig, CK. Wu, H. Soininen, D. Avramidou, RL. Albin, L. Fratiglioni, A. Germanou, LG. Apostolova, L. Keller, M. Koutroumani, SE. Arnold, F. Panza, O. Gkatzima, S. Asthana, D. Hannequin, P. Whitehead, CS. Atwood, P. Caffarra, H. Hampel, I. Quintela, L. Lannfelt, DC. Rubinsztein, LL. Barnes, F. Pasquier, S. Barral, B. McGuinness, TG. Beach, JA. Johnston, JT. Becker, P. Passmore, EH. Bigio, JM. Schott, TD. Bird, JD. Warren, BF. Boeve, MK. Lupton, JD. Bowen, P. Proitsi, A. Boxer, JF. Powell, JR. Burke, JSK. Kauwe, JM. Burns, M. Mancuso, JD. Buxbaum, U. Bonuccelli, NJ. Cairns, A. McQuillin, C. Cao, G. Livingston, CS. Carlson, NJ. Bass, CM. Carlsson, J. Hardy, RM. Carney, J. Bras, M. Carrasquillo, R. Guerreiro, M. Allen, HC. Chui, E. Fisher, C. Masullo, EA. Crocco, C. DeCarli, G. Bisceglio, M. Dick, L. Ma, R. Duara, N. Graff-Radford, DA. Evans, A. Hodges, KM. Faber, M. Scherer, KB. Fallon, M. Riemenschneider, DW. Fardo, R. Heun, MR. Farlow, S. Ferris, M. Leber, TM. Foroud, I. Heuser, DR. Galasko, I. Giegling, M. Gearing, DH. Geschwind, JR. Gilbert, J. Morris, RC. Green, K. Mayo, JH. Growdon, T. Feulner, RL. Hamilton, LE. Harrell, D. Drichel, LS. Honig, TD. Cushion, MJ. Huentelman, P. Hollingworth, CM. Hulette, BT. Hyman, R. Marshall, GP. Jarvik, A. Meggy, E. Abner, GE. Menzies, LW. Jin, G. Leonenko, LM. Real, GR. Jun, CT. Baldwin, D. Grozeva, A. Karydas, G. Russo, JA. Kaye, R. Kim, F. Jessen, NW. Kowall, B. Vellas, JH. Kramer, E. Vardy, FM. LaFerla, JJ. Lah, M. Dichgans, JB. Leverenz, D. Mann, AI. Levey, S. Pickering-Brown, AP. Lieberman, N. Klopp, KL. Lunetta, HE. Wichmann, CG. Lyketsos, K. Morgan, DC. Marson, K. Brown, F. Martiniuk, C. Medway, DC. Mash, E. Masliah, NM. Hooper, WC. McCormick, A. Daniele, SM. McCurry, A. Bayer, AN. McDavid, J. Gallacher, AC. McKee, H. van den Bussche, M. Mesulam, C. Brayne, BL. Miller, S. Riedel-Heller, CA. Miller, JW. Miller, A. Al-Chalabi, JC. Morris, CE. Shaw, AJ. Myers, J. Wiltfang, S. O'Bryant, JM. Olichney, V. Alvarez, JE. Parisi, AB. Singleton, HL. Paulson, J. Collinge, WR. Perry, S. Mead, E. Peskind, DH. Cribbs, M. Rossor, A. Pierce, NS. Ryan, WW. Poon, B. Nacmias, H. Potter, S. Sorbi, JF. Quinn, E. Sacchinelli, A. Raj .G, Spalletta, M. Raskind, C. Caltagirone, P. Boss, MD. Orfei, B. Reisberg, R. Clarke, C. Reitz, AD. Smith, JM. Ringman, D. Warden, ED. Roberson, G. Wilcock, E. Rogaeva, AC. Bruni, HJ. Rosen, M. Gallo, RN. Rosenberg, Y. Ben-Shlomo, MA. Sager, P. Mecocci, AJ. Saykin, P. Pastor, ML. Cuccaro, JM. Vance, JA. Schneider, LS. Schneider, S. Slifer, WW. Seeley, AG. Smith, JA. Sonnen, S. Spina, RA. Stern, RH. Swerdlow, M. Tang, RE. Tanzi, JQ. Trojanowski, JC. Troncoso, VM. Van Deerlin, LJ. Van Eldik, HV. Vinters, JP. Vonsattel, S. Weintraub, KA. Welsh-Bohmer, KC. Wilhelmsen, J. Williamson, TS. Wingo, RL. Woltjer, CB. Wright, CE. Yu, L. Yu , Y. Saba, A. Pilotto, MJ. Bullido, O. Peters, PK. Crane, D. Bennett, P. Bosco, E. Coto, V. Boccardi, PL. De Jager, A.

Lleo, N. Warner, OL. Lopez, M. Ingelsson, P. Deloukas, C. Cruchaga, C. Graff, R. Gwilliam, M. Fornage AM. Goate, P. Sanchez-Juan, PG. Kehoe, N. Amin, N. Ertekin-Taner, C. Berr, S. Debette, S. Love, LJ. Launer, SG. Younkin, JF. Dartigues, C. Corcoran, MA. Ikram , DW. Dickson, G. Nicolas, D. Campion, J. Tschanz, H. Schmidt, H. Hakonarson, J. Clarimon, R. Munger, R. Schmidt, LA. Farrer, C. Van Broeckhoven, M. O'Donovan, AL. DeStefano, L. Jones, JL. Haines, JF. Deleuze, MJ. Owen, V. Gudnason, R. Mayeux, V. Escott-Price, BM. Psaty, A. Ramirez, LS. Wang, A. Ruiz, CM. van Duijn, PA. Holmans, S. Seshadri, J. Williams, P. Amouyel, GD. Schellenberg, JC. Lambert, MA. Pericak-Vance, Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet. 51, 414–430 (2019). doi: 10.1038/S41588-019-0358-2

151. I. Miguel-Escalada, S. Bonàs-Guarch, I. Cebola, J. Ponsa-Cobas, J. Mendieta-Esteban, G. Atla, B. M. Javierre, D. M. Y. Rolando, I. Farabella, C. C. Morgan, J. García-Hurtado, A. Beucher, I. Morán, L. Pasquali, M. Ramos-Rodríguez, E. V. R. Appel, A. Linneberg, A. P. Gjesing, D. R. Witte, O. Pedersen, N. Grarup, P. Ravassard, D. Torrents, J. M. Mercader, L. Piemonti, T. Berney, E. J. P. de Koning, J. Kerr-Conte, F. Pattou, I. O. Fedko, L. Groop, I. Prokopenko, T. Hansen, M. A. Marti-Renom, P. Fraser, J. Ferrer, Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nature Genetics (2019), doi:10.1038/s41588-019-0457-0.

152. H. Martin, Role of PPAR-gamma in inflammation. Prospects for therapeutic intervention by food components. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis. 690 (2010), pp. 57–63. doi: 10.1016/j.mrfmmm.2009.09.009

153. A. Zieleniak, M. Wójcik, L. A. Woźniak, Structure and physiological functions of the human peroxisome proliferator-activated receptor γ. Archivum Immunologiae et Therapiae Experimentalis. 56 (2008), pp. 331–345. doi: 10.1007/s00005-008-0037-y

154. D. J. A. Adamson, D. Frew, R. Tatoud, C. Roland Wolf, C. N. A. Palmer, Diclofenac antagonizes peroxisome proliferator-activated receptor-γ signaling. Molecular Pharmacology. 61, 7–12 (2002). doi: 10.1124/mol.61.1.7

155. F. Q. Onofrio, G. M. Hirschfield, The Pathophysiology of Cholestasis and Its Relevance to Clinical Practice. Clinical Liver Disease. 15 (2020), pp. 110–114. doi: 10.1002/cld.894

156. A. Zhang, C. Y. Li, E. J. Kelly, L. Sheppard, J. Y. Cui, Transcriptomic profiling of PBDE-exposed HepaRG cells unveils critical lncRNA- PCG pairs involved in intermediary metabolism. PLoS ONE. 15, e0224644 (2020). doi: 10.1371/journal.pone.0224644

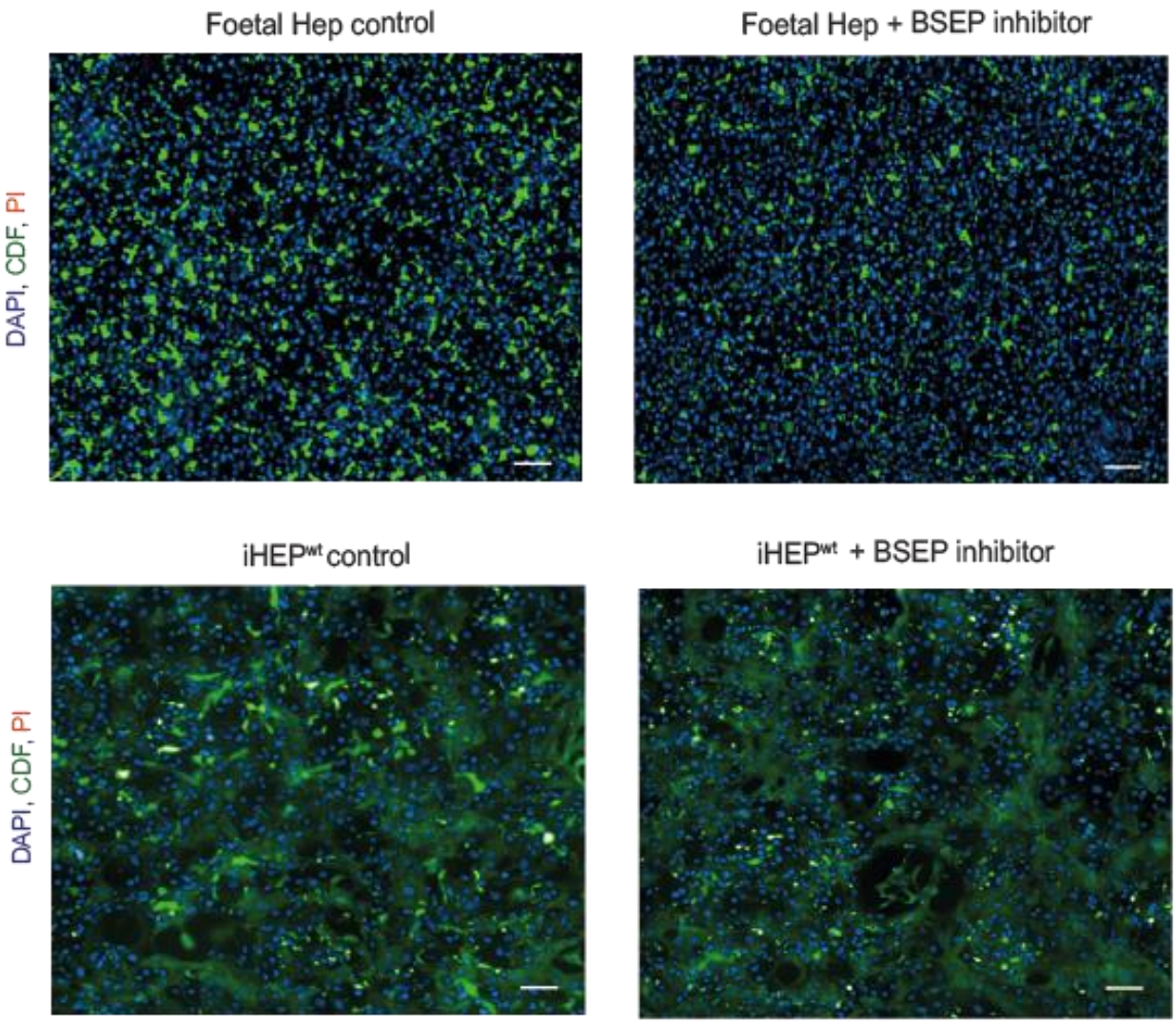
157. F. Wan, M. J. Lenardo, The nuclear signaling of NF-κB: Current knowledge, new insights, and future perspectives. Cell Research. 20 (2010), pp. 24–33. doi: 10.1038/cr.2009.137

158. K. Cui, M. Coutts, J. Stahl, A. J. Sytkowski, Novel interaction between the transcription factor CHOP (GADD153) and the ribosomal protein FTE/S3a modulates erythropoiesis. Journal of Biological Chemistry. 275, 7591–7596 (2000). doi: 10.1074/jbc.275.11.7591
159. J. W. P. Yam, D. Y. Jin, C. W. So, L. C. Chan, Identification and characterization of EBP, a novel EEN binding protein that inhibits Ras signaling and is recruited into the nucleus by the MLL-EEN fusion protein. Blood. 103, 1445–1453 (2004). doi: 10.1182/blood-2003-07-2452
160. M. S. Anwer, Cellular regulation of hepatic bile acid transport in health and cholestasis. Hepatology. 39, 581–590 (2004). doi: 10.1002/hep.20090
161. M. Tanaka, D. Nanba, S. Mori, F. Shiba, H. Ishiguro, K. Yoshino, N. Matsuura, S. Higashiyama, ADAM binding protein Eve-1 is required for ectodomain shedding of epidermal growth factor receptor ligands. Journal of Biological Chemistry. 279, 41950–41959 (2004). doi: 10.1074/jbc.M400086200
162. J. Svinka, S. Pflügler, M. Mair, H. U. Marschall, J. G. Hengstler, P. Stiedl, V. Poli, E. Casanova, G. Timelthaler, M. Sibilia, R. Eferl, Epidermal growth factor signaling protects from cholestatic liver injury and fibrosis. Journal of Molecular Medicine. 95, 109–117 (2017). doi: 10.1007/s00109-016-1462-8
163. J. A. Byrne, S. Frost, Y. Chen, R. K. Bright, Tumor protein D52 (TPD52) and cancer—oncogene understudy or understudied oncogene? Tumor Biology. 35, 7369–7382 (2014). doi: 10.1007/s13277-014-2006-x
164. C. Luni, S. Giulitti, E. Serena, L. Ferrari, A. Zambon, O. Gagliano, GG. Giobbe, F. Michielin, S. Knbel, A. Bosio, N. Elvassore, High-efficiency cellular reprogramming with microfluidics. Nat Methods. 13, 446–452 (2016). doi: 10.1038/NMETH.3832
165. G. Neu-Yilik, B. Amthor, N. H. Gehring, S. Bahri, H. Paidassi, M. W. Hentze, A. E. Kulozik, Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. RNA. 17, 843–854 (2011). doi: 10.1261/RNA.2401811
166. OJ. Rackham, J. Firas, H. Fang, ME. Oates, ML. Holmes, AS. Knaupp, H. Suzuki, CM. Nefzger, CO. Daub, JW. Shin, E. Petretto, AR. Forrest, Y. Hayashizaki, JM. Polo, J. Gough, A predictive computational framework for direct reprogramming between human cell types. Nat Genet. 48, 331–335 (2016). doi: 10.1038/NG.3487
167. M. Pawlowski, D. Ortmann, A. Bertero, JM. Tavares, RA. Pedersen, L. Vallier, MRN. Kotter, Inducible and Deterministic Forward Programming of Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes. Stem Cell Reports. 8, 803–812 (2017). doi: 10.1016/J.STEMCR.2017.02.016
168. C. Stephens, MI. Lucena, RJ. Andrade, Genetic variations in drug-induced liver injury(DILI): resolving the puzzle. Front Genet. 3 (2012) ,doi:10.3389/FGENE.2012.00253.
169. Y. Zhang, R. De Mets, C. Monzel, V. Acharya, P. Toh, J. F. L. Chin, N. Van Hul, I. C. Ng, H. Yu, S. S. Ng, S. Tamir Rashid, V. Viasnoff, Biomimetic niches reveal the minimal

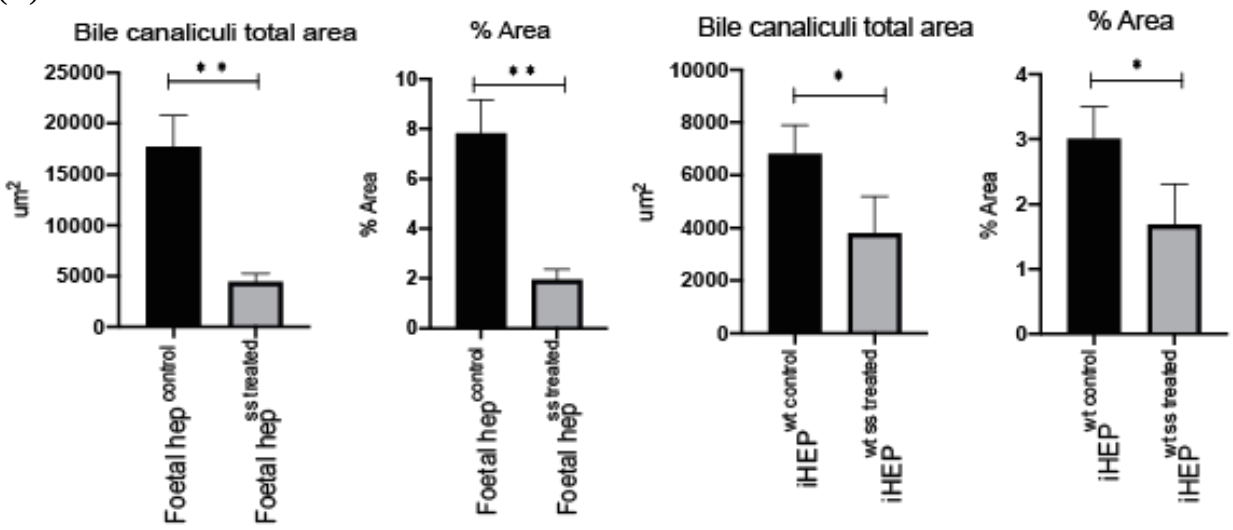
cues to trigger apical lumen formation in single hepatocytes. Nature Materials (2020), doi:10.1038/s41563-020-0662-3.

# Appendix 1

(A)


Foetal Hep control
Foetal Hep + BSEP inhibitor
DAPI, CDF, PI
iHEP^wt control
iHEP^wt + BSEP inhibitor
DAPI, CDF, PI

(B)




**Supplementary Figure 1 Foetal hepatocytes and iHEP$^{wt}$ treated with MRP2/BSEP inhibitor (sitaxsentan sodium)** (A) Representative confocal micrographs of iHEP$^{wt}$ and foetal HEP are taken by a high-throughput imaging system, CLS oppereta, and show the transport and accumulation of a fluorescent tracer (metabolites of CDF, green), representing the function and morphology of bile canaliculi of cells cultured in the Matrigel sandwich system. Nuclei are stained by DAPI (blue) to mark the position of the nuclei. iHEP$^{wt}$ and foetal HEP treated with SS display very few canaliculi-like structures and CLF accumulation, while the nontreated group displayed 'chicken-wire'-like fluorescent tracer accumulation patterns. Scale bar, 100 µm. (B) Morphological quantification of bile canalicular network in CDFDA assays using the ImageJ processing algorithm. The pattern of fluorescent tracers in SS-treated iHEP$^{wt}$ and foetal HEP show significantly lower total bile canalicular area and bile canalicular percentage area compared to that of the nontreated groups. Mean±SD, n = 3, unpaired student t-test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ns: nonsignificant.

**Table S1 Molecular biology primers used in this study**

| genes | primers 5' to 3' |
|---|---|
| TJP2 EXON5f | TTGTGGTCAAGAGGCCCC |
| TJP2 EXON5r | GTTCAGCCGGCTCCTCTC |
| TJP2EXON4/7f | TCATTCGTTTGCAGTTCAGC |
| TJP2EXON4/7r | CTGGCTGTCTCTCAACACCA |
| ALBf | CGTCGAGATGCACACAAGA |
| ALBr | GATACTGAGCAAAGGCAATCAAC |
| ASGR2f | CGTGGGTGACAAGATCACAT |
| ASGR2r | GGGAAGTGCTTCAGATGGAA |
| SERPINA1f | CCGAAGAGGCCAAGAAACAG |
| SERPINA1r | GGTCTCTCCCATTTGCCTTT |
| CYP2E1f | GCTGTGGTGCATGAGATT |
| CYP2E1r | ACTACGACTGTGCCCTT |
| RPL13Af | GTACTGCTGGGCCGGAA |
| RPL13Ar | GAGGAAAGCCAGGTACTTCAAC |

**Table S2 All DILI SNP used in this study**

| #chr | start | end | RS number | #chr | start | end | RS number |
|---|---|---|---|---|---|---|---|
| chr1 | 114377568 | 114377568 | rs2476601 | chr6 | 31389740 | 31389740 | rs116081995 |
| chr1 | 114303808 | 114303808 | rs6679677 | chr6 | 31387667 | 31387667 | rs75119533 |
| chr2 | 5232178 | 5232178 | rs72631567 | chr6 | 31387448 | 31387448 | rs79479695 |
| chr2 | 5232126 | 5232126 | rs77152121 | chr6 | 31383613 | 31383613 | rs115841246 |
| chr2 | 5232266 | 5232266 | rs72631568 | chr6 | 31382495 | 31382495 | rs140810304 |
| chr2 | 5230413 | 5230413 | rs72631566 | chr6 | 31380001 | 31380001 | rs372860257 |
| chr2 | 5227802 | 5227802 | rs74630953 | chr6 | 31378864 | 31378864 | rs41558312 |
| chr2 | 5227752 | 5227752 | rs72631565 | chr6 | 31378257 | 31378257 | rs144888775 |
| chr2 | 5227258 | 5227258 | rs146571690 | chr6 | 31373445 | 31373445 | rs2923003 |
| chr2 | 5227036 | 5227036 | rs72631564 | chr6 | 31373018 | 31373018 | rs202036493 |
| chr2 | 5224750 | 5224750 | rs72631562 | chr6 | 31371071 | 31371071 | rs59440261 |
| chr2 | 5224245 | 5224245 | rs77227894 | chr6 | 31365110 | 31365110 | rs138099588 |
| chr2 | 5222276 | 5222276 | rs72631561 | chr6 | 31363554 | 31363554 | rs201415628 |
| chr2 | 5221629 | 5221629 | rs1453781 | chr6 | 31351940 | 31351940 | rs114607072 |
| chr2 | 5221032 | 5221032 | rs72631560 | chr6 | 31351298 | 31351298 | rs113223977 |
| chr2 | 5220662 | 5220662 | rs72631559 | chr6 | 31346655 | 31346655 | rs58102217 |
| chr2 | 5234205 | 5234205 | rs72631569 | chr6 | 31345421 | 31345421 | rs114170382 |
| chr2 | 234641604 | 234641604 | rs75520741 | chr6 | 31335791 | 31335791 | rs57989216 |
| chr2 | 234580140 | 234580140 | rs2741045 | chr6 | 31335189 | 31335189 | rs112168410 |
| chr2 | 234580249 | 234580249 | rs2741046 | chr6 | 31331255 | 31331255 | rs115986568 |
| chr2 | 234579368 | 234579368 | rs2741044 | chr6 | 31329628 | 31329628 | rs111518019 |
| chr2 | 234582051 | 234582051 | rs2602376 | chr6 | 31329447 | 31329447 | rs113688927 |
| chr2 | 234568964 | 234568964 | rs2602374 | chr6 | 31329100 | 31329100 | rs112515516 |
| chr2 | 234565917 | 234565917 | rs2741042 | chr6 | 31324202 | 31324202 | rs41556417 |
| chr2 | 234563998 | 234563998 | rs2741038 | chr6 | 31323455 | 31323455 | rs41557415 |
| chr2 | 234561953 | 234561953 | rs2602373 | chr6 | 31322987 | 31322987 | rs1140487 |
| chr2 | 234558802 | 234558802 | rs2741036 | chr6 | 31322690 | 31322690 | rs41543314 |
| chr2 | 234548814 | 234548814 | rs2741034 | chr6 | 31322611 | 31322611 | rs41561016 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr2 | 234536152 | 234536152 | rs2741030 | chr6 | 31322224 | 31322224 | rs146647111 |
| chr2 | 234530160 | 234530160 | rs2741029 | chr6 | 31322210 | 31322210 | rs151341415 |
| chr2 | 234524229 | 234524229 | rs2602364 | chr6 | 31322108 | 31322108 | rs111301312 |
| chr2 | 234520525 | 234520525 | rs2602363 | chr6 | 31321776 | 31321776 | rs1057151 |
| chr2 | 234518914 | 234518914 | rs2741028 | chr6 | 31409677 | 31409677 | rs140991764 |
| chr2 | 234518011 | 234518011 | rs2741027 | chr6 | 31387810 | 31387810 | rs79411911 |
| chr2 | 234517745 | 234517745 | rs2741024 | chr6 | 31344182 | 31344182 | rs149663102 |
| chr2 | 234516714 | 234516714 | rs2741023 | chr6 | 31342005 | 31342005 | rs114118665 |
| chr2 | 234516548 | 234516548 | rs2602362 | chr6 | 31325884 | 31325884 | rs147887806 |
| chr2 | 234514273 | 234514273 | rs2741022 | chr6 | 31390361 | 31390361 | rs113265260 |
| chr2 | 234511883 | 234511883 | rs2741021 | chr6 | 31445771 | 31445771 | rs3021366 |
| chr2 | 234510688 | 234510688 | rs2741019 | chr6 | 31452292 | 31452292 | rs2905741 |
| chr2 | 234509911 | 234509911 | rs2741013 | chr6 | 31453290 | 31453290 | rs2905736 |
| chr2 | 234508963 | 234508963 | rs2741012 | chr6 | 31453711 | 31453711 | rs2905734 |
| chr2 | 234557086 | 234557086 | rs71058568 | chr6 | 31458936 | 31458936 | rs2905725 |
| chr3 | 186650790 | 186650790 | rs10937275 | chr6 | 31462648 | 31462648 | rs41293856 |
| chr3 | 186649683 | 186649683 | rs35102841 | chr6 | 31464091 | 31464091 | rs41293860 |
| chr3 | 186649678 | 186649678 | rs35461422 | chr6 | 31465917 | 31465917 | rs3828917 |
| chr3 | 186648929 | 186648929 | rs35376159 | chr6 | 31468404 | 31468404 | rs41293879 |
| chr3 | 186651388 | 186651388 | rs35923019 | chr6 | 31472338 | 31472338 | rs201725498 |
| chr3 | 186646817 | 186646817 | rs13059833 | chr6 | 31472345 | 31472345 | rs199923645 |
| chr3 | 186656777 | 186656777 | rs62292579 | chr6 | 31474820 | 31474820 | rs41293883 |
| chr3 | 186656802 | 186656802 | rs62292580 | chr6 | 31475628 | 31475628 | rs4959077 |
| chr3 | 186646335 | 186646335 | rs36118972 | chr6 | 31478960 | 31478960 | rs7757162 |
| chr3 | 186646140 | 186646140 | rs35166820 | chr6 | 31479223 | 31479223 | rs7757383 |
| chr3 | 186644463 | 186644463 | rs34695488 | chr6 | 31479535 | 31479535 | rs41293887 |
| chr3 | 186642111 | 186642111 | rs62294593 | chr6 | 31481133 | 31481133 | rs41293895 |
| chr3 | 186645080 | 186645080 | rs34421968 | chr6 | 31481753 | 31481753 | rs41293899 |
| chr3 | 186643240 | 186643240 | rs35508867 | chr6 | 31482626 | 31482626 | rs4959078 |
| chr3 | 186640688 | 186640688 | rs71322421 | chr6 | 31483892 | 31483892 | rs41293907 |
| chr3 | 186638620 | 186638620 | rs199886150 | chr6 | 31480005 | 31480005 | rs41293891 |
| chr3 | 186635584 | 186635584 | rs62294592 | chr6 | 31492279 | 31492279 | rs41293911 |
| chr3 | 186658866 | 186658866 | rs13061952 | chr6 | 31492870 | 31492870 | rs41293915 |
| chr3 | 186645278 | 186645278 | rs36086787 | chr6 | 31496381 | 31496381 | rs2516484 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr3 | 186631046 | 186631046 | rs371705710 | chr6 | 31503426 | 31503426 | rs2516474 |
| chr3 | 186625327 | 186625327 | rs115077394 | chr6 | 31509980 | 31509980 | rs41293919 |
| chr3 | 186625068 | 186625068 | rs62294586 | chr6 | 31511579 | 31511579 | rs41293923 |
| chr3 | 186621834 | 186621834 | rs62292548 | chr6 | 31519365 | 31519365 | rs28732141 |
| chr3 | 186619015 | 186619015 | rs62292542 | chr6 | 31521095 | 31521095 | rs28732142 |
| chr3 | 186623622 | 186623622 | rs115741745 | chr6 | 31530810 | 31530810 | rs28732143 |
| chr3 | 186622850 | 186622850 | rs115131245 | chr6 | 31532140 | 31532140 | rs370208598 |
| chr3 | 186622573 | 186622573 | rs149422997 | chr6 | 31546495 | 31546495 | rs3093668 |
| chr3 | 186621828 | 186621828 | rs57888171 | chr6 | 31546789 | 31546789 | rs3093726 |
| chr3 | 186659784 | 186659784 | rs34406259 | chr6 | 31547115 | 31547115 | rs3093727 |
| chr3 | 186659794 | 186659794 | rs62292583 | chr6 | 31556205 | 31556205 | rs28732144 |
| chr3 | 186638628 | 186638628 | rs77512271 | chr6 | 31558264 | 31558264 | rs28732145 |
| chr3 | 186638625 | 186638625 | rs74916631 | chr6 | 31562107 | 31562107 | rs28895015 |
| chr3 | 186661526 | 186661526 | rs35550772 | chr6 | 31566109 | 31566109 | rs199682092 |
| chr3 | 186661767 | 186661767 | rs34794275 | chr6 | 31569520 | 31569520 | rs17207190 |
| chr3 | 186662593 | 186662593 | rs13080971 | chr6 | 29828660 | 29828660 | rs2523822 |
| chr3 | 12330411 | 12330411 | rs17036170 | chr6 | 29822413 | 29822413 | rs2905756 |
| chr3 | 12450838 | 12450838 | rs147667955 | chr6 | 29822261 | 29822261 | rs2975033 |
| chr3 | 12173015 | 12173015 | rs111833688 | chr6 | 29822139 | 29822139 | rs2975034 |
| chr4 | 151680327 | 151680327 | rs28521457 | chr6 | 29821567 | 29821567 | rs2734982 |
| chr4 | 151682768 | 151682768 | rs13340305 | chr6 | 29803880 | 29803880 | rs9295821 |
| chr4 | 151686176 | 151686176 | rs57790879 | chr6 | 29826092 | 29826092 | rs2517840 |
| chr4 | 151671230 | 151671230 | rs17027065 | chr6 | 29834199 | 29834199 | rs2734972 |
| chr4 | 151699039 | 151699039 | rs28558492 | chr6 | 29818159 | 29818159 | rs4711200 |
| chr4 | 151596942 | 151596942 | rs28544421 | chr6 | 29799942 | 29799942 | rs28724925 |
| chr4 | 151573490 | 151573490 | rs10017937 | chr6 | 29835026 | 29835026 | rs2734968 |
| chr4 | 151678178 | 151678178 | rs77391623 | chr6 | 29835518 | 29835518 | rs2734967 |
| chr4 | 151692477 | 151692477 | rs9884732 | chr6 | 29841990 | 29841990 | rs2734966 |
| chr4 | 151664792 | 151664792 | rs10028040 | chr6 | 29842409 | 29842409 | rs2523812 |
| chr4 | 151706532 | 151706532 | rs10004115 | chr6 | 29833840 | 29833840 | rs2523818 |
| chr4 | 151750977 | 151750977 | rs75524406 | chr6 | 29841280 | 29841280 | rs2844820 |
| chr4 | 151752049 | 151752049 | rs1289963 | chr6 | 29798794 | 29798794 | rs9380142 |
| chr4 | 151752425 | 151752425 | rs294528 | chr6 | 29849674 | 29849674 | rs2734960 |
| chr4 | 151771134 | 151771134 | rs1148656 | chr6 | 29819202 | 29819202 | rs2975035 |
| chr4 | 151588483 | 151588483 | rs10026518 | chr6 | 29803526 | 29803526 | rs4713250 |
| chr4 | 151783863 | 151783863 | rs76142287 | chr6 | 29802690 | 29802690 | rs9380145 |
| chr4 | 151566660 | 151566660 | rs10034965 | chr6 | 29802045 | 29802045 | rs9380143 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr4 | 151550574 | 151550574 | rs75799239 | chr6 | 29785101 | 29785101 | rs7776026 |
| chr4 | 151810591 | 151810591 | rs10015967 | chr6 | 29773158 | 29773158 | rs2394173 |
| chr4 | 151542950 | 151542950 | rs12505123 | chr6 | 29772098 | 29772098 | rs2394688 |
| chr4 | 151826409 | 151826409 | rs78868697 | chr6 | 29762389 | 29762389 | rs28724922 |
| chr4 | 151854194 | 151854194 | rs114216415 | chr6 | 29791033 | 29791033 | rs6934957 |
| chr4 | 151860397 | 151860397 | rs7657668 | chr6 | 29787202 | 29787202 | rs6911139 |
| chr4 | 151861133 | 151861133 | rs79108077 | chr6 | 29784503 | 29784503 | rs6457057 |
| chr4 | 151868502 | 151868502 | rs28861118 | chr6 | 29757715 | 29757715 | rs2735035 |
| chr4 | 151870837 | 151870837 | rs60850361 | chr6 | 29755760 | 29755760 | rs2743940 |
| chr4 | 151879620 | 151879620 | rs10033792 | chr6 | 29749118 | 29749118 | rs2735038 |
| chr4 | 151880765 | 151880765 | rs9992474 | chr6 | 29745175 | 29745175 | rs2517929 |
| chr4 | 151902531 | 151902531 | rs6849172 | chr6 | 29744623 | 29744623 | rs2975038 |
| chr4 | 151905456 | 151905456 | rs59746424 | chr6 | 29740166 | 29740166 | rs2844841 |
| chr4 | 151708133 | 151708133 | rs28572569 | chr6 | 29762303 | 29762303 | rs9357085 |
| chr4 | 151713863 | 151713863 | rs10021640 | chr6 | 29745075 | 29745075 | rs2517930 |
| chr4 | 151716904 | 151716904 | rs28627398 | chr6 | 29745350 | 29745350 | rs2517928 |
| chr4 | 151717214 | 151717214 | rs28591543 | chr6 | 29742257 | 29742257 | rs2743945 |
| chr4 | 151719931 | 151719931 | rs28630451 | chr6 | 29795751 | 29795751 | rs6932596 |
| chr4 | 151727220 | 151727220 | rs17027116 | chr6 | 29795747 | 29795747 | rs6932888 |
| chr4 | 151730948 | 151730948 | rs1966844 | chr6 | 29784693 | 29784693 | rs6457058 |
| chr4 | 151605322 | 151605322 | rs28379788 | chr6 | 29932897 | 29932897 | rs3903160 |
| chr4 | 151560598 | 151560598 | rs28802622 | chr6 | 29838642 | 29838642 | rs2844821 |
| chr4 | 151551823 | 151551823 | rs28539520 | chr6 | 29850335 | 29850335 | rs2523806 |
| chr4 | 151551658 | 151551658 | rs28501240 | chr6 | 29914089 | 29914089 | rs12153924 |
| chr4 | 151550668 | 151550668 | rs28540258 | chr6 | 29849333 | 29849333 | rs77274701 |
| chr4 | 151544008 | 151544008 | rs6849364 | chr6 | 29828051 | 29828051 | rs1611702 |
| chr4 | 151541832 | 151541832 | rs111851502 | chr6 | 29821268 | 29821268 | rs3094643 |
| chr4 | 151528104 | 151528104 | rs7655913 | chr6 | 29828467 | 29828467 | rs1611704 |
| chr4 | 151526846 | 151526846 | rs4260528 | chr6 | 29818229 | 29818229 | rs2734987 |
| chr4 | 151522607 | 151522607 | rs78439800 | chr6 | 29908525 | 29908525 | rs28749139 |
| chr4 | 151515915 | 151515915 | rs28641948 | chr6 | 29910752 | 29910752 | rs1136683 |
| chr4 | 151510151 | 151510151 | rs6858706 | chr6 | 29829482 | 29829482 | rs1611715 |
| chr4 | 151874128 | 151874128 | rs146774955 | chr6 | 29909279 | 29909279 | rs28749142 |
| chr4 | 151903350 | 151903350 | rs28537980 | chr6 | 29910378 | 29910378 | rs41546314 |
| chr4 | 151903634 | 151903634 | rs28415994 | chr6 | 29910450 | 29910450 | rs17885299 |
| chr4 | 151904376 | 151904376 | rs75442857 | chr6 | 29908239 | 29908239 | rs1143147 |
| chr4 | 151904498 | 151904498 | rs10007479 | chr6 | 29908680 | 29908680 | rs200453522 |
| chr4 | 151906743 | 151906743 | rs149367054 | chr6 | 29871167 | 29871167 | rs28994643 |
| chr4 | 151923264 | 151923264 | rs76211772 | chr6 | 29818123 | 29818123 | rs2523763 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr4 | 151930476 | 151930476 | rs143563179 | chr6 | 29825566 | 29825566 | rs1611680 |
| chr4 | 151950247 | 151950247 | rs10029299 | chr6 | 29820278 | 29820278 | rs3115627 |
| chr4 | 151972546 | 151972546 | rs72963695 | chr6 | 29911092 | 29911092 | rs1136702 |
| chr4 | 151974908 | 151974908 | rs11940627 | chr6 | 29908500 | 29908500 | rs28994656 |
| chr4 | 151978470 | 151978470 | rs72963696 | chr6 | 29937541 | 29937541 | rs11752303 |
| chr4 | 151981527 | 151981527 | rs12507141 | chr6 | 29907835 | 29907835 | rs28749130 |
| chr4 | 151982349 | 151982349 | rs78946251 | chr6 | 29910986 | 29910986 | rs17882350 |
| chr4 | 151984034 | 151984034 | rs6839157 | chr6 | 29928487 | 29928487 | rs2508035 |
| chr4 | 151994154 | 151994154 | rs77113754 | chr6 | 29932865 | 29932865 | rs9357088 |
| chr4 | 151997382 | 151997382 | rs142778287 | chr6 | 29937493 | 29937493 | rs4713274 |
| chr4 | 152002915 | 152002915 | rs77063077 | chr6 | 29937580 | 29937580 | rs4713275 |
| chr4 | 152003472 | 152003472 | rs80001494 | chr6 | 29937562 | 29937562 | rs71830922 |
| chr4 | 152003481 | 152003481 | rs78077687 | chr6 | 29926300 | 29926300 | rs11267511 |
| chr4 | 152007201 | 152007201 | rs74322476 | chr6 | 29935576 | 29935576 | rs4538750 |
| chr4 | 152010007 | 152010007 | rs6847518 | chr6 | 29935590 | 29935590 | rs4391295 |
| chr4 | 152010630 | 152010630 | rs146153670 | chr6 | 29936307 | 29936307 | rs4959037 |
| chr4 | 152012053 | 152012053 | rs78084868 | chr6 | 29936404 | 29936404 | rs6927487 |
| chr4 | 152015988 | 152015988 | rs75997360 | chr6 | 29937104 | 29937104 | rs12193110 |
| chr4 | 152020777 | 152020777 | rs2280283 | chr6 | 32305979 | 32305979 | rs3129900 |
| chr4 | 152023697 | 152023697 | rs146146423 | chr6 | 32318610 | 32318610 | rs3117119 |
| chr4 | 152025969 | 152025969 | rs7656771 | chr6 | 32336495 | 32336495 | rs3129938 |
| chr4 | 152030340 | 152030340 | rs147762399 | chr6 | 32336187 | 32336187 | rs3129934 |
| chr4 | 152031561 | 152031561 | rs7682480 | chr6 | 32320153 | 32320153 | rs3132963 |
| chr4 | 152037378 | 152037378 | rs7675432 | chr6 | 32316016 | 32316016 | rs3117125 |
| chr4 | 152037668 | 152037668 | rs112353547 | chr6 | 32338620 | 32338620 | rs9281766 |
| chr4 | 152038588 | 152038588 | rs6535764 | chr6 | 32338632 | 32338632 | rs138019081 |
| chr4 | 152040304 | 152040304 | rs76919918 | chr6 | 32289390 | 32289390 | rs9268234 |
| chr4 | 152043997 | 152043997 | rs113768215 | chr6 | 32279938 | 32279938 | rs9268205 |
| chr4 | 152044300 | 152044300 | rs10009307 | chr6 | 32266310 | 32266310 | rs9268155 |
| chr4 | 152044661 | 152044661 | rs78759735 | chr6 | 32266021 | 32266021 | rs9268154 |
| chr4 | 152045368 | 152045368 | rs28694456 | chr6 | 32259527 | 32259527 | rs9268148 |
| chr4 | 152046288 | 152046288 | rs76834528 | chr6 | 32237926 | 32237926 | rs6913182 |
| chr4 | 152046979 | 152046979 | rs3762846 | chr6 | 32232652 | 32232652 | rs9268072 |
| chr4 | 152047365 | 152047365 | rs28593016 | chr6 | 32245370 | 32245370 | rs9268103 |
| chr4 | 152049296 | 152049296 | rs77084366 | chr6 | 32255236 | 32255236 | rs201851759 |
| chr4 | 152049743 | 152049743 | rs76967763 | chr6 | 32263099 | 32263099 | rs9268149 |
| chr4 | 152062317 | 152062317 | rs75298151 | chr6 | 32253100 | 32253100 | rs66593511 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr4 | 152063662 | 152063662 | rs113391399 | chr6 | 32300809 | 32300809 | rs3129960 |
| chr4 | 152065197 | 152065197 | rs3736502 | chr6 | 32257444 | 32257444 | rs9268147 |
| chr4 | 152067460 | 152067460 | rs7699495 | chr8 | 18272881 | 18272881 | rs1495741 |
| chr4 | 152071066 | 152071066 | rs12501643 | chr8 | 18272535 | 18272535 | rs35246381 |
| chr4 | 152073536 | 152073536 | rs76659656 | chr8 | 18272635 | 18272635 | rs35570672 |
| chr4 | 152073557 | 152073557 | rs78718721 | chr8 | 18272503 | 18272503 | rs146812806 |
| chr4 | 152074049 | 152074049 | rs112161281 | chr8 | 18272466 | 18272466 | rs4921915 |
| chr4 | 152076989 | 152076989 | rs112419089 | chr8 | 18272438 | 18272438 | rs4921914 |
| chr4 | 152085914 | 152085914 | rs76755719 | chr8 | 18272377 | 18272377 | rs4921913 |
| chr4 | 152086244 | 152086244 | rs74976701 | chr8 | 18274398 | 18274398 | rs35942058 |
| chr4 | 152086611 | 152086611 | rs74439947 | chr8 | 18274443 | 18274443 | rs34987019 |
| chr4 | 152087065 | 152087065 | rs75271898 | chr8 | 18274684 | 18274684 | rs34537991 |
| chr4 | 152091342 | 152091342 | rs79599251 | chr8 | 18274614 | 18274614 | rs7812546 |
| chr4 | 152091609 | 152091609 | rs75832380 | chr8 | 18275189 | 18275189 | rs7816847 |
| chr4 | 152094544 | 152094544 | rs75715047 | chr8 | 81058819 | 81058819 | rs7828135 |
| chr4 | 152098353 | 152098353 | rs74768630 | chr8 | 81059454 | 81059454 | rs7814474 |
| chr4 | 152101230 | 152101230 | rs6858302 | chr8 | 81053736 | 81053736 | rs7835273 |
| chr4 | 152111196 | 152111196 | rs60953992 | chr8 | 81040453 | 81040453 | rs77290663 |
| chr4 | 152115280 | 152115280 | rs78296340 | chr8 | 81044270 | 81044270 | rs201811135 |
| chr4 | 152117145 | 152117145 | rs77852985 | chr8 | 81021965 | 81021965 | rs12545071 |
| chr4 | 152119596 | 152119596 | rs11337469 | chr8 | 81017668 | 81017668 | rs4740105 |
| chr4 | 152122773 | 152122773 | rs77657350 | chr18 | 64629498 | 64629498 | rs116561224 |
| chr4 | 152125521 | 152125521 | rs60367453 | chr18 | 64628547 | 64628547 | rs28618088 |
| chr4 | 152127159 | 152127159 | rs10635205 | chr18 | 64632044 | 64632044 | rs116157500 |
| chr4 | 152127924 | 152127924 | rs74883557 | chr18 | 64637471 | 64637471 | rs9944860 |
| chr4 | 152131850 | 152131850 | rs141015944 | chr18 | 64620604 | 64620604 | rs142287014 |
| chr4 | 152136896 | 152136896 | rs78264024 | chr18 | 64640928 | 64640928 | rs8089772 |
| chr4 | 152139882 | 152139882 | rs74575144 | chr18 | 64643684 | 64643684 | rs28458792 |
| chr4 | 152140506 | 152140506 | rs58496896 | chr18 | 64643685 | 64643685 | rs28662367 |
| chr4 | 152146844 | 152146844 | rs78465919 | chr18 | 64612995 | 64612995 | rs116125729 |
| chr4 | 152156928 | 152156928 | rs144857627 | chr18 | 64647271 | 64647271 | rs9948964 |
| chr4 | 152158519 | 152158519 | rs17633648 | chr18 | 64648224 | 64648224 | rs28897228 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr4 | 152159423 | 152159423 | rs147152551 | chr18 | 64610596 | 64610596 | rs9949630 |
| chr4 | 152159560 | 152159560 | rs17633654 | chr18 | 64607346 | 64607346 | rs76586541 |
| chr4 | 152159628 | 152159628 | rs74664063 | chr18 | 64604569 | 64604569 | rs368331878 |
| chr4 | 152159726 | 152159726 | rs72967545 | chr18 | 64603770 | 64603770 | rs183111972 |
| chr4 | 152163489 | 152163489 | rs7671332 | chr18 | 64601284 | 64601284 | rs28736561 |
| chr5 | 160684731 | 160684731 | rs114811931 | chr18 | 64599792 | 64599792 | rs76157070 |
| chr5 | 160582417 | 160582417 | rs141984490 | chr18 | 64599391 | 64599391 | rs28560289 |
| chr6 | 31431780 | 31431780 | rs2395029 | chr18 | 64598835 | 64598835 | rs8089722 |
| chr6 | 31430060 | 31430060 | rs115846244 | chr18 | 64659765 | 64659765 | rs9964921 |
| chr6 | 31421547 | 31421547 | rs144027808 | chr18 | 64659767 | 64659767 | rs9954056 |
| chr6 | 31412271 | 31412271 | rs77311173 | chr18 | 64674587 | 64674587 | rs9950332 |
| chr6 | 31411714 | 31411714 | rs74655380 | chr18 | 64675435 | 64675435 | rs28402898 |
| chr6 | 31405128 | 31405128 | rs112689184 | chr18 | 64675654 | 64675654 | rs60383679 |
| chr6 | 31402358 | 31402358 | rs148792134 | chr18 | 64677117 | 64677117 | rs150018650 |
| chr6 | 31400705 | 31400705 | rs138117378 | chr18 | 64668693 | 64668693 | rs8087787 |
| chr6 | 31400137 | 31400137 | rs138130755 | chr18 | 64578171 | 64578171 | rs373290353 |
| chr6 | 31390971 | 31390971 | rs116339333 | chr18 | 64578170 | 64578170 | rs370352946 |
| chr6 | 31390365 | 31390365 | rs201023435 | chr18 | 64577103 | 64577103 | rs114749128 |
| chr6 | 31390266 | 31390266 | rs116419909 | chr18 | 64576553 | 64576553 | rs117317106 |
| | | | | chr18 | 64575083 | 64575083 | rs78115464 |
| | | | | chr18 | 64573442 | 64573442 | rs8087835 |

## Table S3 ANNVOAR analysis for DILI SNPs

| Chr | Start | End | Ref | Alt | Func.refGene | Gene.refGene | GeneDetail.refGene | ExonicFunc.refGene | AAChange.refGene | Xref.refGene | cytoBand | Variant |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr1** | 114303808 | 114303808 | C | A | downstream | RSBN1 | dist=646 | . | . | . | 1p13.2 | rs6679677 |
| **chr1** | 114377568 | 114377568 | A | G | exonic | PTPN22 | . | nonsynonymous SNV | PTPN22:NM_012411:exon12:c.T1693C:p.W565R,PTPN22:NM_001308297:exon13:c.T1786C:p.W596R,PTPN22:NM_001193431:exon14:c.T1858C:p.W620R,PTPN22:NM_015967:exon14:c.T1858C:p.W620R | . | 1p13.2 | rs2476601 |
| **chr18** | 64573442 | 64573442 | G | C | intergenic | CDH19;MIR5011 | dist=302115;dist=175379 | . | . | . | 18q22.1 | rs8087835 |
| **chr18** | 64575083 | 64575083 | C | G | intergenic | CDH19;MIR5011 | dist=303756;dist=173738 | . | . | . | 18q22.1 | rs78115464 |
| **chr18** | 64576553 | 64576553 | T | C | intergenic | CDH19;MIR5011 | dist=305226;dist=172268 | . | . | . | 18q22.1 | rs117317106 |
| **chr18** | 64577103 | 64577103 | T | C | intergenic | CDH19;MIR5011 | dist=305776;dist=171718 | . | . | . | 18q22.1 | rs114749128 |
| **chr18** | 64578170 | 64578170 | A | G | intergenic | CDH19;MIR5011 | dist=306843;dist=170651 | . | . | . | 18q22.1 | rs370352946 |
| **chr18** | 64578171 | 64578171 | G | A | intergenic | CDH19;MIR5011 | dist=306844;dist=170650 | . | . | . | 18q22.1 | rs373290353 |
| **chr18** | 64598835 | 64598835 | G | A | intergenic | CDH19;MIR5011 | dist=327508;dist=149986 | . | . | . | 18q22.1 | rs8089722 |
| **chr18** | 64599391 | 64599391 | T | C | intergenic | CDH19;MIR5011 | dist=328064;dist=149430 | . | . | . | 18q22.1 | rs28560289 |
| **chr18** | 64599792 | 64599792 | T | C | intergenic | CDH19;MIR5011 | dist=328465;dist=149029 | . | . | . | 18q22.1 | rs76157070 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr18** | 64601284 | 64601284 | T | C | intergenic | CDH19;MIR5011 | dist=329957;dist=147537 | . | . | . | 18q22.1 | rs28736561 |
| **chr18** | 64603770 | 64603770 | C | G | intergenic | CDH19;MIR5011 | dist=332443;dist=145051 | . | . | . | 18q22.1 | rs183111972 |
| **chr18** | 64604569 | 64604569 | A | C | intergenic | CDH19;MIR5011 | dist=333242;dist=144252 | . | . | . | 18q22.1 | rs368331878 |
| **chr18** | 64607346 | 64607346 | G | A | intergenic | CDH19;MIR5011 | dist=336019;dist=141475 | . | . | . | 18q22.1 | rs76586541 |
| **chr18** | 64610596 | 64610596 | T | C | intergenic | CDH19;MIR5011 | dist=339269;dist=138225 | . | . | . | 18q22.1 | rs9949630 |
| **chr18** | 64612995 | 64612995 | T | C | intergenic | CDH19;MIR5011 | dist=341668;dist=135826 | . | . | . | 18q22.1 | rs116125729 |
| **chr18** | 64620604 | 64620604 | - | C | intergenic | CDH19;MIR5011 | dist=349277;dist=128217 | . | . | . | 18q22.1 | rs142287014 |
| **chr18** | 64628547 | 64628547 | T | A | intergenic | CDH19;MIR5011 | dist=357220;dist=120274 | . | . | . | 18q22.1 | rs28618088 |
| **chr18** | 64629498 | 64629498 | A | G | intergenic | CDH19;MIR5011 | dist=358171;dist=119323 | . | . | . | 18q22.1 | rs116561224 |
| **chr18** | 64632044 | 64632044 | G | A | intergenic | CDH19;MIR5011 | dist=360717;dist=116777 | . | . | . | 18q22.1 | rs116157500 |
| **chr18** | 64637471 | 64637471 | G | A | intergenic | CDH19;MIR5011 | dist=366144;dist=111350 | . | . | . | 18q22.1 | rs9944860 |
| **chr18** | 64640928 | 64640928 | G | T | intergenic | CDH19;MIR5011 | dist=369601;dist=107893 | . | . | . | 18q22.1 | rs8089772 |
| **chr18** | 64643684 | 64643684 | A | T | intergenic | CDH19;MIR5011 | dist=372357;dist=105137 | . | . | . | 18q22.1 | rs28458792 |
| **chr18** | 64643685 | 64643685 | A | T | intergenic | CDH19;MIR5011 | dist=372358;dist=105136 | . | . | . | 18q22.1 | rs28662367 |
| **chr18** | 64647271 | 64647271 | T | G | intergenic | CDH19;MIR5011 | dist=375944;dist=101550 | . | . | . | 18q22.1 | rs9948964 |
| **chr18** | 64648224 | 64648224 | G | C | intergenic | CDH19;MIR5011 | dist=376897;dist=100597 | . | . | . | 18q22.1 | rs28897228 |
| **chr18** | 64659765 | 64659765 | C | G | intergenic | CDH19;MIR5011 | dist=388438;dist=89056 | . | . | . | 18q22.1 | rs9964921 |
| **chr18** | 64659767 | 64659767 | T | C | intergenic | CDH19;MIR5011 | dist=388440;dist=89054 | . | . | . | 18q22.1 | rs9954056 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr18** | 64668693 | 64668693 | A | C | intergenic | CDH19;MIR5011 | dist=397366;dist=80128 | . | . | . | 18q22.1 | rs8087787 |
| **chr18** | 64674587 | 64674587 | C | G | intergenic | CDH19;MIR5011 | dist=403260;dist=74234 | . | . | . | 18q22.1 | rs9950332 |
| **chr18** | 64675435 | 64675435 | C | T | intergenic | CDH19;MIR5011 | dist=404108;dist=73386 | . | . | . | 18q22.1 | rs28402898 |
| **chr18** | 64675654 | 64675654 | G | A | intergenic | CDH19;MIR5011 | dist=404327;dist=73167 | . | . | . | 18q22.1 | rs60383679 |
| **chr18** | 64677117 | 64677117 | - | G | intergenic | CDH19;MIR5011 | dist=405790;dist=71704 | . | . | . | 18q22.1 | rs150018650 |
| **chr2** | 52206062 | 52206062 | A | T | intergenic | LINC01249;LINC01248 | dist=516850;dist=553611 | . | . | . | 2p25.2 | rs72631559 |
| **chr2** | 52210032 | 52210032 | G | A | intergenic | LINC01249;LINC01248 | dist=517220;dist=553241 | . | . | . | 2p25.2 | rs72631560 |
| **chr2** | 52216029 | 52216029 | T | C | intergenic | LINC01249;LINC01248 | dist=517817;dist=552644 | . | . | . | 2p25.2 | rs1453781 |
| **chr2** | 52222276 | 52222276 | T | C | intergenic | LINC01249;LINC01248 | dist=518464;dist=551997 | . | . | . | 2p25.2 | rs72631561 |
| **chr2** | 52242246 | 52242246 | G | - | intergenic | LINC01249;LINC01248 | dist=520434;dist=550027 | . | . | . | 2p25.2 | rs398090106 |
| **chr2** | 52247250 | 52247250 | T | C | intergenic | LINC01249;LINC01248 | dist=520938;dist=549523 | . | . | . | 2p25.2 | rs72631562 |
| **chr2** | 52270036 | 52270036 | A | G | intergenic | LINC01249;LINC01248 | dist=523224;dist=547237 | . | . | . | 2p25.2 | rs72631564 |
| **chr2** | 52272758 | 52272758 | - | T | intergenic | LINC01249;LINC01248 | dist=523446;dist=547015 | . | . | . | 2p25.2 | rs146571690 |
| **chr2** | 52277752 | 52277752 | A | G | intergenic | LINC01249;LINC01248 | dist=523940;dist=546521 | . | . | . | 2p25.2 | rs72631565 |
| **chr2** | 52278002 | 52278002 | G | A | intergenic | LINC01249;LINC01248 | dist=523990;dist=546471 | . | . | . | 2p25.2 | rs74630953 |
| **chr2** | 52304013 | 52304013 | G | A | intergenic | LINC01249;LINC01248 | dist=526601;dist=543860 | . | . | . | 2p25.2 | rs72631566 |
| **chr2** | 52321227 | 52321228 | GT | - | intergenic | LINC01249;LINC01248 | dist=528315;dist=542145 | . | . | . | 2p25.2 | rs398090107 |
| **chr2** | 52321778 | 52321778 | A | G | intergenic | LINC01249;LINC01248 | dist=528366;dist=542095 | . | . | . | 2p25.2 | rs72631567 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr2** | 52322 66 | 52322 66 | T | A | intergenic | LINC01249;LINC01248 | dist=528454;dist=542007 | . | . | . | 2p25.2 | rs72631568 |
| **chr2** | 52342 05 | 52342 05 | T | G | intergenic | LINC01249;LINC01248 | dist=530393;dist=540068 | . | . | . | 2p25.2 | rs72631569 |
| **chr2** | 23450 8963 | 23450 8963 | C | T | intergenic | USP40;UGT1A8 | dist=33535;dist=17328 | . | . | . | 2q37.1 | rs2741012 |
| **chr2** | 23450 9911 | 23450 9911 | C | T | intergenic | USP40;UGT1A8 | dist=34483;dist=16380 | . | . | . | 2q37.1 | rs2741013 |
| **chr2** | 23451 0688 | 23451 0688 | C | G | intergenic | USP40;UGT1A8 | dist=35260;dist=15603 | . | . | . | 2q37.1 | rs2741019 |
| **chr2** | 23451 1883 | 23451 1883 | C | T | intergenic | USP40;UGT1A8 | dist=36455;dist=14408 | . | . | . | 2q37.1 | rs2741021 |
| **chr2** | 23451 4273 | 23451 4273 | C | T | intergenic | USP40;UGT1A8 | dist=38845;dist=12018 | . | . | . | 2q37.1 | rs2741022 |
| **chr2** | 23451 6548 | 23451 6548 | A | C | intergenic | USP40;UGT1A8 | dist=41120;dist=9743 | . | . | . | 2q37.1 | rs2602362 |
| **chr2** | 23451 6714 | 23451 6714 | G | A | intergenic | USP40;UGT1A8 | dist=41286;dist=9577 | . | . | . | 2q37.1 | rs2741023 |
| **chr2** | 23451 7745 | 23451 7745 | G | A | intergenic | USP40;UGT1A8 | dist=42317;dist=8546 | . | . | . | 2q37.1 | rs2741024 |
| **chr2** | 23451 8011 | 23451 8011 | G | A | intergenic | USP40;UGT1A8 | dist=42583;dist=8280 | . | . | . | 2q37.1 | rs2741027 |
| **chr2** | 23451 8914 | 23451 8914 | G | A | intergenic | USP40;UGT1A8 | dist=43486;dist=7377 | . | . | . | 2q37.1 | rs2741028 |
| **chr2** | 23452 0525 | 23452 0525 | T | C | intergenic | USP40;UGT1A8 | dist=45097;dist=5766 | . | . | . | 2q37.1 | rs2602363 |
| **chr2** | 23452 4229 | 23452 4229 | G | A | intergenic | USP40;UGT1A8 | dist=48801;dist=2062 | . | . | . | 2q37.1 | rs2602364 |
| **chr2** | 23453 0160 | 23453 0160 | T | G | intronic | UGT1A8 | . | . | . | . | 2q37.1 | rs2741029 |
| **chr2** | 23453 6152 | 23453 6152 | C | T | intronic | UGT1A8 | . | . | . | . | 2q37.1 | rs2741030 |
| **chr2** | 23454 8814 | 23454 8814 | A | G | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | rs2741034 |
| **chr2** | 23455 7103 | 23455 7103 | - | ATT | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | . |
| **chr2** | 23455 8802 | 23455 8802 | G | A | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | rs2741036 |
| **chr2** | 23456 1953 | 23456 1953 | T | C | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | rs2602373 |
| **chr2** | 23456 3998 | 23456 3998 | G | A | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | rs2741038 |
| **chr2** | 23456 5917 | 23456 5917 | C | T | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | rs2741042 |
| **chr2** | 23456 8964 | 23456 8964 | C | T | intronic | UGT1A10;UGT1A8 | . | . | . | . | 2q37.1 | rs2602374 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr2** | 234579368 | 234579368 | G | A | intronic | UGT1A10; UGT1A8 | . | . | . | . | 2q37.1 | rs2741044 |
| **chr2** | 234580140 | 234580140 | C | T | intronic | UGT1A10; UGT1A8 | . | . | . | . | 2q37.1 | rs2741045 |
| **chr2** | 234580249 | 234580249 | T | C | intronic | UGT1A10; UGT1A8 | . | . | . | . | 2q37.1 | rs2741046 |
| **chr2** | 234582051 | 234582051 | C | T | intronic | UGT1A10; UGT1A8; UGT1A9 | . | . | . | . | 2q37.1 | rs2602376 |
| **chr2** | 234641604 | 234641604 | C | G | intronic | UGT1A10; UGT1A3; UGT1A4; UGT1A5; UGT1A6; UGT1A7; UGT1A8; UGT1A9 | . | . | . | . | 2q37.1 | rs75520741 |
| **chr3** | 12173015 | 12173015 | T | C | intronic | SYN2 | . | . | . | . | 3p25.2 | rs111833688 |
| **chr3** | 12330411 | 12330411 | G | A | intronic | PPARG | . | . | . | Carotid intimal medial thickness 1;Insulin resistance, severe, digenic, Autosomal dominant;Lipodystrophy, familial partial, type 3, Autosomal dominant;Obesity, severe, Autosomal recessive, Autosomal dominant, Multifactorial | 3p25.2 | rs17036170 |
| **chr3** | 12450838 | 12450838 | C | T | UTR3 | PPARG | NM_001354670:c.*3192C>T;NM_001354668:c.*3192C>T | . | . | Carotid intimal medial thickness 1;Insulin resistance, severe, digenic, Autosomal dominant;Lipodystrophy, familial partial, type 3, Autosomal dominant;Obesity, severe, Autosomal recessive, Autosomal dominant, Multifactorial | 3p25.2 | rs147667955 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr3** | 186619015 | 186619015 | G | A | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=39858;dist=29300 | . | . | . | 3q27.3 | rs62292542 |
| **chr3** | 186621828 | 186621828 | G | A | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=42671;dist=26487 | . | . | . | 3q27.3 | rs57888171 |
| **chr3** | 186621834 | 186621834 | T | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=42677;dist=26481 | . | . | . | 3q27.3 | rs62292548 |
| **chr3** | 186622573 | 186622573 | G | A | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=43416;dist=25742 | . | . | . | 3q27.3 | rs36038544 |
| **chr3** | 186622850 | 186622850 | T | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=43693;dist=25465 | . | . | . | 3q27.3 | rs74853841 |
| **chr3** | 186623622 | 186623622 | T | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=44465;dist=24693 | . | . | . | 3q27.3 | rs13081409 |
| **chr3** | 186625068 | 186625068 | C | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=45911;dist=23247 | . | . | . | 3q27.3 | rs62294586 |
| **chr3** | 186625327 | 186625327 | T | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=46170;dist=22988 | . | . | . | 3q27.3 | rs62294588 |
| **chr3** | 186631047 | 186631052 | TCCCGC | - | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=51890;dist=17263 | . | . | . | 3q27.3 | rs371705710 |
| **chr3** | 186635584 | 186635584 | C | T | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=56427;dist=12731 | . | . | . | 3q27.3 | rs62294592 |
| **chr3** | 186638621 | 186638622 | TC | - | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=59464;dist=9693 | . | . | . | 3q27.3 | rs199886150 |
| **chr3** | 186638625 | 186638625 | C | A | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=59468;dist=9690 | . | . | . | 3q27.3 | rs74916631 |
| **chr3** | 186638628 | 186638628 | A | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=59471;dist=9687 | . | . | . | 3q27.3 | rs77512271 |
| **chr3** | 186640688 | 186640688 | T | A | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=61531;dist=7627 | . | . | . | 3q27.3 | rs71322421 |
| **chr3** | 186642111 | 186642111 | T | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=62954;dist=6204 | . | . | . | 3q27.3 | rs62294593 |
| **chr3** | 186643240 | 186643240 | C | T | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=64083;dist=5075 | . | . | . | 3q27.3 | rs35508867 |
| **chr3** | 186644463 | 186644463 | A | G | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=65306;dist=3852 | . | . | . | 3q27.3 | rs34695488 |
| **chr3** | 186645080 | 186645080 | T | C | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=65923;dist=3235 | . | . | . | 3q27.3 | rs34421968 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr3** | 186645278 | 186645278 | G | A | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=66121;dist=3037 | . | . | . | 3q27.3 | rs36086787 |
| **chr3** | 186646140 | 186646140 | C | T | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=66983;dist=2175 | . | . | . | 3q27.3 | rs35166820 |
| **chr3** | 186646335 | 186646335 | T | C | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=67178;dist=1980 | . | . | . | 3q27.3 | rs36118972 |
| **chr3** | 186646817 | 186646817 | G | T | intergenic | ADIPOQ-AS1;ST6GAL1 | dist=67660;dist=1498 | . | . | . | 3q27.3 | rs13059833 |
| **chr3** | 186648929 | 186648929 | T | A | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs35376159 |
| **chr3** | 186649678 | 186649678 | A | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs35461422 |
| **chr3** | 186649683 | 186649683 | A | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs35102841 |
| **chr3** | 186650790 | 186650790 | A | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs10937275 |
| **chr3** | 186651388 | 186651388 | C | T | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs35923019 |
| **chr3** | 186656777 | 186656777 | C | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs62292579 |
| **chr3** | 186656802 | 186656802 | A | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs62292580 |
| **chr3** | 186658866 | 186658866 | G | C | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs13061952 |
| **chr3** | 186659784 | 186659784 | A | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs34406259 |
| **chr3** | 186659794 | 186659794 | C | T | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs62292583 |
| **chr3** | 186661526 | 186661526 | A | G | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs35550772 |
| **chr3** | 186661767 | 186661767 | T | C | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs34794275 |
| **chr3** | 186662593 | 186662593 | C | T | intronic | ST6GAL1 | . | . | . | . | 3q27.3 | rs13080971 |
| **chr4** | 151510151 | 151510151 | A | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs6858706 |
| **chr4** | 151515915 | 151515915 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28641948 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151522608 | 151522613 | AAAGAT | - | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs78439800 |
| **chr4** | 151526846 | 151526846 | A | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs4260528 |
| **chr4** | 151528104 | 151528104 | C | G | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs7655913 |
| **chr4** | 151541832 | 151541832 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs111851502 |
| **chr4** | 151542950 | 151542950 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs12505123 |
| **chr4** | 151544008 | 151544008 | C | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs6849364 |
| **chr4** | 151550574 | 151550574 | T | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs75799239 |
| **chr4** | 151550668 | 151550668 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28540258 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151551658 | 151551658 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28501240 |
| **chr4** | 151551823 | 151551823 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28539520 |
| **chr4** | 151560598 | 151560598 | T | G | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28802622 |
| **chr4** | 151566660 | 151566660 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10034965 |
| **chr4** | 151573490 | 151573490 | A | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10017937 |
| **chr4** | 151588483 | 151588483 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10026518 |
| **chr4** | 151596942 | 151596942 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28544421 |
| **chr4** | 151605322 | 151605322 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28379788 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151664792 | 151664792 | G | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10028040 |
| **chr4** | 151671230 | 151671230 | T | G | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs17027065 |
| **chr4** | 151678179 | 151678182 | TTAA | - | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs77391623 |
| **chr4** | 151680327 | 151680327 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28521457 |
| **chr4** | 151682768 | 151682768 | A | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs13340305 |
| **chr4** | 151686176 | 151686176 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs57790879 |
| **chr4** | 151692477 | 151692477 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs9884732 |
| **chr4** | 151699039 | 151699039 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28558492 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151706532 | 151706532 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10004115 |
| **chr4** | 151708133 | 151708133 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28572569 |
| **chr4** | 151713863 | 151713863 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10021640 |
| **chr4** | 151716904 | 151716904 | T | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28627398 |
| **chr4** | 151717214 | 151717214 | C | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28591543 |
| **chr4** | 151719931 | 151719931 | C | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28630451 |
| **chr4** | 151727220 | 151727220 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs17027116 |
| **chr4** | 151730948 | 151730948 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs1966844 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151750977 | 151750977 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs75524406 |
| **chr4** | 151752049 | 151752049 | G | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs1289963 |
| **chr4** | 151752425 | 151752425 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs294528 |
| **chr4** | 151771134 | 151771134 | A | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs1148656 |
| **chr4** | 151783863 | 151783863 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs76142287 |
| **chr4** | 151810591 | 151810591 | G | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10015967 |
| **chr4** | 151826409 | 151826409 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs78868697 |
| **chr4** | 151854194 | 151854194 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs114216415 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151860397 | 151860397 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs7657668 |
| **chr4** | 151861133 | 151861133 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs79108077 |
| **chr4** | 151868502 | 151868502 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28861118 |
| **chr4** | 151870837 | 151870837 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs60850361 |
| **chr4** | 151874128 | 151874128 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs146774955 |
| **chr4** | 151879620 | 151879620 | A | G | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10033792 |
| **chr4** | 151880765 | 151880765 | A | G | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs9992474 |
| **chr4** | 151902531 | 151902531 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs6849172 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151903350 | 151903350 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28537980 |
| **chr4** | 151903634 | 151903634 | A | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs28415994 |
| **chr4** | 151904376 | 151904376 | A | G | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs75442857 |
| **chr4** | 151904498 | 151904498 | T | C | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs10007479 |
| **chr4** | 151905456 | 151905456 | - | AA | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs59746424 |
| **chr4** | 151906743 | 151906743 | C | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs149367054 |
| **chr4** | 151923264 | 151923264 | G | A | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs76211772 |
| **chr4** | 151930476 | 151930476 | A | T | intronic | LRBA | . | . | . | Immunodeficiency, common variable, 8, with autoimmunity, Autosomal recessive | 4q31.3 | rs143563179 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 151950247 | 151950247 | G | A | intergenic | LRBA;RPS3A | dist=13368;dist=70478 | . | . | . | 4q31.3 | rs10029299 |
| **chr4** | 151972546 | 151972546 | G | A | intergenic | LRBA;RPS3A | dist=35667;dist=48179 | . | . | . | 4q31.3 | rs72963695 |
| **chr4** | 151974908 | 151974908 | A | G | intergenic | LRBA;RPS3A | dist=38029;dist=45817 | . | . | . | 4q31.3 | rs11940627 |
| **chr4** | 151978470 | 151978470 | G | A | intergenic | LRBA;RPS3A | dist=41591;dist=42255 | . | . | . | 4q31.3 | rs72963696 |
| **chr4** | 151981527 | 151981527 | C | A | intergenic | LRBA;RPS3A | dist=44648;dist=39198 | . | . | . | 4q31.3 | rs12507141 |
| **chr4** | 151982349 | 151982349 | A | G | intergenic | LRBA;RPS3A | dist=45470;dist=38376 | . | . | . | 4q31.3 | rs78946251 |
| **chr4** | 151984034 | 151984034 | G | A | intergenic | LRBA;RPS3A | dist=47155;dist=36691 | . | . | . | 4q31.3 | rs6839157 |
| **chr4** | 151994154 | 151994154 | A | G | intergenic | LRBA;RPS3A | dist=57275;dist=26571 | . | . | . | 4q31.3 | rs77113754 |
| **chr4** | 151997383 | 151997388 | TTATTT | - | intergenic | LRBA;RPS3A | dist=60504;dist=23337 | . | . | . | 4q31.3 | rs142778287 |
| **chr4** | 152002915 | 152002915 | A | G | intergenic | LRBA;RPS3A | dist=66036;dist=17810 | . | . | . | 4q31.3 | rs77063077 |
| **chr4** | 152003472 | 152003472 | G | A | intergenic | LRBA;RPS3A | dist=66593;dist=17253 | . | . | . | 4q31.3 | rs80001494 |
| **chr4** | 152003481 | 152003481 | G | T | intergenic | LRBA;RPS3A | dist=66602;dist=17244 | . | . | . | 4q31.3 | rs78077687 |
| **chr4** | 152007201 | 152007201 | A | G | intergenic | LRBA;RPS3A | dist=70322;dist=13524 | . | . | . | 4q31.3 | rs74322476 |
| **chr4** | 152010007 | 152010007 | C | T | intergenic | LRBA;RPS3A | dist=73128;dist=10718 | . | . | . | 4q31.3 | rs6847518 |
| **chr4** | 152010631 | 152010631 | T | - | intergenic | LRBA;RPS3A | dist=73752;dist=10094 | . | . | . | 4q31.3 | rs146153670 |
| **chr4** | 152012053 | 152012053 | C | T | intergenic | LRBA;RPS3A | dist=75174;dist=8672 | . | . | . | 4q31.3 | rs78084868 |
| **chr4** | 152015988 | 152015988 | C | T | intergenic | LRBA;RPS3A | dist=79109;dist=4737 | . | . | . | 4q31.3 | rs75997360 |
| **chr4** | 152020777 | 152020777 | C | T | UTR5 | RPS3A | NM_001267699:c.-28C>T | . | . | . | 4q31.3 | rs2280283 |
| **chr4** | 152023697 | 152023697 | - | G | intronic | RPS3A | . | . | . | . | 4q31.3 | rs146146423 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 152025969 | 152025969 | T | C | downstream | RPS3A;SNORD73A | dist=175 | . | . | . | 4q31.3 | rs7656771 |
| **chr4** | 152030340 | 152030340 | C | T | intergenic | RPS3A;SH3D19 | dist=4546;dist=11093 | . | . | . | 4q31.3 | rs147762399 |
| **chr4** | 152031561 | 152031561 | C | T | intergenic | RPS3A;SH3D19 | dist=5767;dist=9872 | . | . | . | 4q31.3 | rs7682480 |
| **chr4** | 152037378 | 152037378 | G | A | intergenic | RPS3A;SH3D19 | dist=11584;dist=4055 | . | . | . | 4q31.3 | rs7675432 |
| **chr4** | 152037668 | 152037668 | - | T | intergenic | RPS3A;SH3D19 | dist=11874;dist=3765 | . | . | . | 4q31.3 | rs112353547 |
| **chr4** | 152038588 | 152038588 | G | C | intergenic | RPS3A;SH3D19 | dist=12794;dist=2845 | . | . | . | 4q31.3 | rs6535764 |
| **chr4** | 152040304 | 152040304 | G | C | intergenic | RPS3A;SH3D19 | dist=14510;dist=1129 | . | . | . | 4q31.3 | rs76919918 |
| **chr4** | 152043997 | 152043997 | G | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs113768215 |
| **chr4** | 152044300 | 152044300 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs10009307 |
| **chr4** | 152044661 | 152044661 | C | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs78759735 |
| **chr4** | 152045368 | 152045368 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs28694456 |
| **chr4** | 152046288 | 152046288 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs76834528 |
| **chr4** | 152046979 | 152046979 | G | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs3762846 |
| **chr4** | 152047365 | 152047365 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs28593016 |
| **chr4** | 152049296 | 152049296 | G | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs77084366 |
| **chr4** | 152049743 | 152049743 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs76967763 |
| **chr4** | 152062317 | 152062317 | A | G | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs75298151 |
| **chr4** | 152063663 | 152063663 | T | - | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs113391399 |
| **chr4** | 152065197 | 152065197 | C | T | exonic | SH3D19 | . | synonymous SNV | SH3D19:NM_001128924:exon5:c.G1140A:p.G380G,SH3D19:NM_001128923:exon12:c.G1248A:p.G416G,SH3D19:NM_001243349:exon12:c.G1248A:p.G416G,SH3 | . | 4q31.3 | rs3736502 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | D19:NM_001009555:exon13:c.G1317A:p.G439G | | | |
| **chr4** | 152067460 | 152067460 | T | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs7699495 |
| **chr4** | 152071066 | 152071066 | G | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs12501643 |
| **chr4** | 152073536 | 152073536 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs76659656 |
| **chr4** | 152073557 | 152073557 | T | C | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs78718721 |
| **chr4** | 152074049 | 152074049 | A | G | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs112161281 |
| **chr4** | 152076989 | 152076989 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs112419089 |
| **chr4** | 152085914 | 152085914 | T | C | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs76755719 |
| **chr4** | 152086244 | 152086244 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs74976701 |
| **chr4** | 152086611 | 152086611 | A | G | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs74439947 |
| **chr4** | 152087065 | 152087065 | T | C | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs75271898 |
| **chr4** | 152091342 | 152091342 | G | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs79599251 |
| **chr4** | 152091609 | 152091609 | A | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs75832380 |
| **chr4** | 152094544 | 152094544 | T | C | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs75715047 |
| **chr4** | 152098353 | 152098353 | A | G | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs74768630 |
| **chr4** | 152101230 | 152101230 | T | C | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs6858302 |
| **chr4** | 152111196 | 152111196 | T | G | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs60953992 |
| **chr4** | 152115280 | 152115280 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs78296340 |
| **chr4** | 152117145 | 152117145 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs77852985 |
| **chr4** | 152119597 | 152119597 | A | - | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs11337469 |
| **chr4** | 152122773 | 152122773 | T | C | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs77657350 |
| **chr4** | 152125521 | 152125521 | G | A | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs60367453 |
| **chr4** | 152127159 | 152127159 | - | AC | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs10635205 |
| **chr4** | 152127924 | 152127924 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs74883557 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr4** | 152131851 | 152131852 | CC | - | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs141015944 |
| **chr4** | 152136896 | 152136896 | G | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs78264024 |
| **chr4** | 152139883 | 152139883 | T | - | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs74575144 |
| **chr4** | 152140506 | 152140506 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs58496896 |
| **chr4** | 152146844 | 152146844 | C | T | intronic | SH3D19 | . | . | . | . | 4q31.3 | rs78465919 |
| **chr4** | 152156929 | 152156930 | CT | - | intergenic | SH3D19;PRSS48 | dist=7868;dist=41395 | . | . | . | 4q31.3 | rs144857627 |
| **chr4** | 152158519 | 152158519 | T | C | intergenic | SH3D19;PRSS48 | dist=9458;dist=39806 | . | . | . | 4q31.3 | rs17633648 |
| **chr4** | 152159424 | 152159424 | T | - | intergenic | SH3D19;PRSS48 | dist=10363;dist=38901 | . | . | . | 4q31.3 | rs147152551 |
| **chr4** | 152159560 | 152159560 | G | T | intergenic | SH3D19;PRSS48 | dist=10499;dist=38765 | . | . | . | 4q31.3 | rs17633654 |
| **chr4** | 152159628 | 152159628 | T | C | intergenic | SH3D19;PRSS48 | dist=10567;dist=38697 | . | . | . | 4q31.3 | rs74664063 |
| **chr4** | 152159726 | 152159726 | T | C | intergenic | SH3D19;PRSS48 | dist=10665;dist=38599 | . | . | . | 4q31.3 | rs72967545 |
| **chr4** | 152163489 | 152163489 | T | C | intergenic | SH3D19;PRSS48 | dist=14428;dist=34836 | . | . | . | 4q31.3 | rs7671332 |
| **chr5** | 160582417 | 160582417 | G | A | intergenic | LINC02159;GABRB2 | dist=216784;dist=133019 | . | . | . | 5q34 | rs141984490 |
| **chr5** | 160684731 | 160684731 | T | C | intergenic | LINC02159;GABRB2 | dist=319098;dist=30705 | . | . | . | 5q34 | rs114811931 |
| **chr6** | 29740166 | 29740166 | A | C | intergenic | IFITM4P;HCG4 | dist=21241;dist=18642 | . | . | . | 6p22.1 | rs2844841 |
| **chr6** | 29742257 | 29742257 | T | G | intergenic | IFITM4P;HCG4 | dist=23332;dist=16551 | . | . | . | 6p22.1 | rs2743945 |
| **chr6** | 29744623 | 29744623 | T | A | intergenic | IFITM4P;HCG4 | dist=25698;dist=14185 | . | . | . | 6p22.1 | rs2975038 |
| **chr6** | 29745075 | 29745075 | G | T | intergenic | IFITM4P;HCG4 | dist=26150;dist=13733 | . | . | . | 6p22.1 | rs2517930 |
| **chr6** | 29745175 | 29745175 | A | G | intergenic | IFITM4P;HCG4 | dist=26250;dist=13633 | . | . | . | 6p22.1 | rs2517929 |
| **chr6** | 29745350 | 29745350 | C | T | intergenic | IFITM4P;HCG4 | dist=26425;dist=13458 | . | . | . | 6p22.1 | rs2517928 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 29749118 | 29749118 | G | A | intergenic | IFITM4P;HCG4 | dist=30193;dist=9690 | . | . | . | 6p22.1 | rs2735038 |
| **chr6** | 29755760 | 29755760 | T | C | intergenic | IFITM4P;HCG4 | dist=36835;dist=3048 | . | . | . | 6p22.1 | rs2743940 |
| **chr6** | 29757715 | 29757715 | C | T | intergenic | IFITM4P;HCG4 | dist=38790;dist=1093 | . | . | . | 6p22.1 | rs2735035 |
| **chr6** | 29762303 | 29762303 | C | T | ncRNA_intronic | HLA-V | . | . | . | . | 6p22.1 | rs9357085 |
| **chr6** | 29762390 | 29762390 | A | - | ncRNA_intronic | HLA-V | . | . | . | . | 6p22.1 | rs28724922 |
| **chr6** | 29772098 | 29772098 | C | T | intergenic | HLA-V;HLA-G | dist=6514;dist=22658 | . | . | . | 6p22.1 | rs2394688 |
| **chr6** | 29773158 | 29773158 | C | T | intergenic | HLA-V;HLA-G | dist=7574;dist=21598 | . | . | . | 6p22.1 | rs2394173 |
| **chr6** | 29784503 | 29784503 | A | G | intergenic | HLA-V;HLA-G | dist=18919;dist=10253 | . | . | . | 6p22.1 | . |
| **chr6** | 29784693 | 29784693 | C | T | intergenic | HLA-V;HLA-G | dist=19109;dist=10063 | . | . | . | 6p22.1 | rs6457058 |
| **chr6** | 29785101 | 29785101 | G | A | intergenic | HLA-V;HLA-G | dist=19517;dist=9655 | . | . | . | 6p22.1 | . |
| **chr6** | 29787202 | 29787202 | G | A | intergenic | HLA-V;HLA-G | dist=21618;dist=7554 | . | . | . | 6p22.1 | . |
| **chr6** | 29791033 | 29791033 | C | T | intergenic | HLA-V;HLA-G | dist=25449;dist=3723 | . | . | . | 6p22.1 | . |
| **chr6** | 29795747 | 29795747 | G | C | intronic | HLA-G | . | . | . | . | 6p22.1 | . |
| **chr6** | 29795751 | 29795751 | C | T | intronic | HLA-G | . | . | . | . | 6p22.1 | rs6932596 |
| **chr6** | 29798794 | 29798794 | A | G | UTR3 | HLA-G | NM_001363567:c.*278A>G;NM_002127:c.*278A>G | . | . | . | 6p22.1 | rs9380142 |
| **chr6** | 29799943 | 29799943 | A | - | intergenic | HLA-G;ZNRD1ASP | dist=1044;dist=1420 | . | . | . | 6p22.1 | rs28724925 |
| **chr6** | 29802045 | 29802045 | C | T | ncRNA_exonic | ZNRD1ASP | . | . | . | . | 6p22.1 | rs9380143 |
| **chr6** | 29802690 | 29802690 | G | C | ncRNA_exonic | ZNRD1ASP | . | . | . | . | 6p22.1 | rs9380145 |
| **chr6** | 29803526 | 29803526 | G | T | upstream | ZNRD1ASP | dist=696 | . | . | . | 6p22.1 | rs4713250 |
| **chr6** | 29803880 | 29803880 | A | G | intergenic | ZNRD1ASP;HCP5B | dist=1050;dist=35792 | . | . | . | 6p22.1 | rs9295821 |
| **chr6** | 29818123 | 29818123 | G | T | intergenic | ZNRD1ASP;HCP5B | dist=15293;dist=21549 | . | . | . | 6p22.1 | rs2523763 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 29818123 | 29818123 | G | A | intergenic | ZNRD1ASP;HCP5B | dist=15293;dist=21549 | . | . | . | 6p22.1 | rs2523763 |
| **chr6** | 29818159 | 29818159 | C | A | intergenic | ZNRD1ASP;HCP5B | dist=15329;dist=21513 | . | . | . | 6p22.1 | rs4711200 |
| **chr6** | 29818159 | 29818159 | C | T | intergenic | ZNRD1ASP;HCP5B | dist=15329;dist=21513 | . | . | . | 6p22.1 | rs4711200 |
| **chr6** | 29818229 | 29818229 | C | G | intergenic | ZNRD1ASP;HCP5B | dist=15399;dist=21443 | . | . | . | 6p22.1 | rs2734987 |
| **chr6** | 29818229 | 29818229 | C | A | intergenic | ZNRD1ASP;HCP5B | dist=15399;dist=21443 | . | . | . | 6p22.1 | rs2734987 |
| **chr6** | 29819202 | 29819202 | C | G | intergenic | ZNRD1ASP;HCP5B | dist=16372;dist=20470 | . | . | . | 6p22.1 | rs2975035 |
| **chr6** | 29820278 | 29820278 | A | G | intergenic | ZNRD1ASP;HCP5B | dist=17448;dist=19394 | . | . | . | 6p22.1 | rs3115627 |
| **chr6** | 29821268 | 29821268 | G | A | intergenic | ZNRD1ASP;HCP5B | dist=18438;dist=18404 | . | . | . | 6p22.1 | rs3094643 |
| **chr6** | 29821567 | 29821567 | G | T | intergenic | ZNRD1ASP;HCP5B | dist=18737;dist=18105 | . | . | . | 6p22.1 | rs2734982 |
| **chr6** | 29822139 | 29822139 | G | A | intergenic | ZNRD1ASP;HCP5B | dist=19309;dist=17533 | . | . | . | 6p22.1 | rs2975034 |
| **chr6** | 29822261 | 29822261 | G | A | intergenic | ZNRD1ASP;HCP5B | dist=19431;dist=17411 | . | . | . | 6p22.1 | rs2975033 |
| **chr6** | 29822413 | 29822413 | A | C | intergenic | ZNRD1ASP;HCP5B | dist=19583;dist=17259 | . | . | . | 6p22.1 | rs2905756 |
| **chr6** | 29825566 | 29825566 | A | C | intergenic | ZNRD1ASP;HCP5B | dist=22736;dist=14106 | . | . | . | 6p22.1 | . |
| **chr6** | 29825566 | 29825566 | A | G | intergenic | ZNRD1ASP;HCP5B | dist=22736;dist=14106 | . | . | . | 6p22.1 | . |
| **chr6** | 29826092 | 29826092 | C | T | intergenic | ZNRD1ASP;HCP5B | dist=23262;dist=13580 | . | . | . | 6p22.1 | . |
| **chr6** | 29828051 | 29828051 | A | C | intergenic | ZNRD1ASP;HCP5B | dist=25221;dist=11621 | . | . | . | 6p22.1 | rs1611702 |
| **chr6** | 29828051 | 29828051 | A | G | intergenic | ZNRD1ASP;HCP5B | dist=25221;dist=11621 | . | . | . | 6p22.1 | rs1611702 |
| **chr6** | 29828467 | 29828467 | C | T | intergenic | ZNRD1ASP;HCP5B | dist=25637;dist=11205 | . | . | . | 6p22.1 | rs1611704 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 29828660 | 29828660 | A | G | intergenic | ZNRD1ASP;HCP5B | dist=25830;dist=11012 | . | . | . | 6p22.1 | rs2523822 |
| **chr6** | 29829482 | 29829482 | A | C | intergenic | ZNRD1ASP;HCP5B | dist=26652;dist=10190 | . | . | . | 6p22.1 | rs1611715 |
| **chr6** | 29833840 | 29833840 | T | G | intergenic | ZNRD1ASP;HCP5B | dist=31010;dist=5832 | . | . | . | 6p22.1 | rs2523818 |
| **chr6** | 29834199 | 29834199 | T | C | intergenic | ZNRD1ASP;HCP5B | dist=31369;dist=5473 | . | . | . | 6p22.1 | rs2734972 |
| **chr6** | 29835026 | 29835026 | G | A | intergenic | ZNRD1ASP;HCP5B | dist=32196;dist=4646 | . | . | . | 6p22.1 | . |
| **chr6** | 29835518 | 29835518 | T | A | intergenic | ZNRD1ASP;HCP5B | dist=32688;dist=4154 | . | . | . | 6p22.1 | . |
| **chr6** | 29838642 | 29838642 | T | C | intergenic | ZNRD1ASP;HCP5B | dist=35812;dist=1030 | . | . | . | 6p22.1 | rs2844821 |
| **chr6** | 29841280 | 29841280 | T | C | ncRNA_exonic | HCP5B | . | . | . | . | 6p22.1 | rs2844820 |
| **chr6** | 29841990 | 29841990 | T | G | upstream | HCP5B | dist=430 | . | . | . | 6p22.1 | rs2734966 |
| **chr6** | 29842409 | 29842409 | T | C | upstream | HCP5B | dist=849 | . | . | . | 6p22.1 | rs2523812 |
| **chr6** | 29849333 | 29849333 | C | T | intergenic | HCP5B;HLA-H | dist=7773;dist=6017 | . | . | . | 6p22.1 | rs2523811 |
| **chr6** | 29849674 | 29849674 | T | C | intergenic | HCP5B;HLA-H | dist=8114;dist=5676 | . | . | . | 6p22.1 | rs2734960 |
| **chr6** | 29850335 | 29850335 | C | T | intergenic | HCP5B;HLA-H | dist=8775;dist=5015 | . | . | . | 6p22.1 | rs2523806 |
| **chr6** | 29871173 | 29871173 | - | ACACA | intergenic | HLA-H;HCG4B | dist=12317;dist=21196 | . | . | . | 6p22.1 | . |
| **chr6** | 29907835 | 29907835 | T | A | intergenic | HCG4B;HLA-A | dist=12843;dist=2412 | . | . | . | 6p22.1 | rs28749130 |
| **chr6** | 29908239 | 29908239 | C | T | intergenic | HCG4B;HLA-A | dist=13247;dist=2008 | . | . | . | 6p22.1 | rs1143147 |
| **chr6** | 29908501 | 29908501 | - | T | intergenic | HCG4B;HLA-A | dist=13509;dist=1746 | . | . | . | 6p22.1 | . |
| **chr6** | 29908525 | 29908525 | T | C | intergenic | HCG4B;HLA-A | dist=13533;dist=1722 | . | . | . | 6p22.1 | rs28749139 |
| **chr6** | 29908687 | 29908694 | TGATGAAT | - | intergenic | HCG4B;HLA-A | dist=13695;dist=1553 | . | . | . | 6p22.1 | . |
| **chr6** | 29909279 | 29909279 | T | A | upstream | HLA-A | dist=968 | . | . | . | 6p22.1 | rs28749142 |
| **chr6** | 29910378 | 29910378 | C | T | exonic | HLA-A | . | synonymous SNV | HLA-A:NM_001242758:exon1:c.C48T:p.A16A,HLA-A:NM_0 | . | 6p22.1 | rs41546314 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 02116:exon1:c.C48T:p.A16A | | | |
| **chr6** | 29910450 | 29910450 | G | C | intronic | HLA-A | . | . | . | . | 6p22.1 | rs17885299 |
| **chr6** | 29910752 | 29910752 | G | A | exonic | HLA-A | . | nonsynonymous SNV | HLA-A:NM_001242758:exon2:c.G292A:p.D98N,HLA-A:NM_002116:exon2:c.G292A:p.D98N | . | 6p22.1 | rs1136683 |
| **chr6** | 29910752 | 29910752 | G | C | exonic | HLA-A | . | nonsynonymous SNV | HLA-A:NM_001242758:exon2:c.G292C:p.D98H,HLA-A:NM_002116:exon2:c.G292C:p.D98H | . | 6p22.1 | rs1136683 |
| **chr6** | 29910986 | 29910986 | G | A | intronic | HLA-A | . | . | . | . | 6p22.1 | rs17882350 |
| **chr6** | 29911092 | 29911092 | G | A | exonic | HLA-A | . | nonsynonymous SNV | HLA-A:NM_001242758:exon3:c.G391A:p.G131R,HLA-A:NM_002116:exon3:c.G391A:p.G131R | . | 6p22.1 | rs1136702 |
| **chr6** | 29911092 | 29911092 | G | C | exonic | HLA-A | . | nonsynonymous SNV | HLA-A:NM_001242758:exon3:c.G391C:p.G131R,HLA-A:NM_002116:exon3:c.G391C:p.G131R | . | 6p22.1 | rs1136702 |
| **chr6** | 29911092 | 29911092 | G | T | exonic | HLA-A | . | nonsynonymous SNV | HLA-A:NM_001242758:exon3:c.G391T:p.G131W,HLA-A:NM_0 | . | 6p22.1 | rs1136702 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 02116:exon3:c.G391T:p.G131W | | | |
| **chr6** | 29914089 | 29914089 | G | A | downstream | HLA-A | dist=428 | . | . | . | 6p22.1 | rs12153924 |
| **chr6** | 29926301 | 29926315 | GGGACAGTACCCAGG | - | intergenic | HLA-A;HCG9 | dist=12640;dist=16577 | . | . | . | 6p22.1 | rs11267511 |
| **chr6** | 29928487 | 29928487 | T | C | intergenic | HLA-A;HCG9 | dist=14826;dist=14405 | . | . | . | 6p22.1 | rs2508035 |
| **chr6** | 29932865 | 29932865 | G | T | intergenic | HLA-A;HCG9 | dist=19204;dist=10027 | . | . | . | 6p22.1 | rs9357088 |
| **chr6** | 29932897 | 29932897 | G | A | intergenic | HLA-A;HCG9 | dist=19236;dist=9995 | . | . | . | 6p22.1 | rs3903160 |
| **chr6** | 29935576 | 29935576 | G | A | intergenic | HLA-A;HCG9 | dist=21915;dist=7316 | . | . | . | 6p22.1 | rs4538750 |
| **chr6** | 29935590 | 29935590 | C | T | intergenic | HLA-A;HCG9 | dist=21929;dist=7302 | . | . | . | 6p22.1 | rs4391295 |
| **chr6** | 29936307 | 29936307 | A | T | intergenic | HLA-A;HCG9 | dist=22646;dist=6585 | . | . | . | 6p22.1 | rs4959037 |
| **chr6** | 29936404 | 29936404 | T | C | intergenic | HLA-A;HCG9 | dist=22743;dist=6488 | . | . | . | 6p22.1 | rs6927487 |
| **chr6** | 29937104 | 29937104 | C | T | intergenic | HLA-A;HCG9 | dist=23443;dist=5788 | . | . | . | 6p22.1 | rs12193110 |
| **chr6** | 29937493 | 29937493 | G | C | intergenic | HLA-A;HCG9 | dist=23832;dist=5399 | . | . | . | 6p22.1 | rs4713274 |
| **chr6** | 29937541 | 29937541 | G | A | intergenic | HLA-A;HCG9 | dist=23880;dist=5351 | . | . | . | 6p22.1 | rs11752303 |
| **chr6** | 29937563 | 29937565 | GAA | - | intergenic | HLA-A;HCG9 | dist=23902;dist=5327 | . | . | . | 6p22.1 | rs66503418 |
| **chr6** | 29937580 | 29937580 | A | C | intergenic | HLA-A;HCG9 | dist=23919;dist=5312 | . | . | . | 6p22.1 | rs4713275 |
| **chr6** | 31321776 | 31321776 | T | C | UTR3 | HLA-B | NM_005514:c.*302A>G | . | . | . | 6p21.33 | rs1057151 |
| **chr6** | 31322108 | 31322108 | T | G | intronic | HLA-B | . | . | . | . | 6p21.33 | rs111301312 |
| **chr6** | 31322210 | 31322210 | C | G | intronic | HLA-B | . | . | . | . | 6p21.33 | rs151341415 |
| **chr6** | 31322224 | 31322224 | - | C | intronic | HLA-B | . | . | . | . | 6p21.33 | rs146647111 |
| **chr6** | 31322611 | 31322611 | C | T | intronic | HLA-B | . | . | . | . | 6p21.33 | rs41561016 |
| **chr6** | 31322690 | 31322690 | A | G | intronic | HLA-B | . | . | . | . | 6p21.33 | rs41543314 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 31322987 | 31322987 | C | T | exonic | HLA-B | . | synonymous SNV | HLA-B:NM_005514:exon5:c.G909A:p.Q303Q | . | 6p21.33 | rs1140487 |
| **chr6** | 31323455 | 31323455 | A | G | intronic | HLA-B | . | . | . | . | 6p21.33 | rs41557415 |
| **chr6** | 31324202 | 31324202 | T | C | exonic | HLA-B | . | nonsynonymous SNV | HLA-B:NM_005514:exon3:c.A361G:p.S121G | . | 6p21.33 | rs41556417 |
| **chr6** | 31324202 | 31324202 | T | A | exonic | HLA-B | . | nonsynonymous SNV | HLA-B:NM_005514:exon3:c.A361T:p.S121C | . | 6p21.33 | rs41556417 |
| **chr6** | 31325884 | 31325884 | C | T | upstream | HLA-B | dist=928 | . | . | . | 6p21.33 | rs147887806 |
| **chr6** | 31329100 | 31329100 | A | G | intergenic | HLA-B;MICA-AS1 | dist=4144;dist=32966 | . | . | . | 6p21.33 | rs112515516 |
| **chr6** | 31329447 | 31329447 | A | G | intergenic | HLA-B;MICA-AS1 | dist=4491;dist=32619 | . | . | . | 6p21.33 | rs113688927 |
| **chr6** | 31329628 | 31329628 | G | A | intergenic | HLA-B;MICA-AS1 | dist=4672;dist=32438 | . | . | . | 6p21.33 | rs111518019 |
| **chr6** | 31331255 | 31331255 | G | A | intergenic | HLA-B;MICA-AS1 | dist=6299;dist=30811 | . | . | . | 6p21.33 | rs115986568 |
| **chr6** | 31335189 | 31335189 | G | A | intergenic | HLA-B;MICA-AS1 | dist=10233;dist=26877 | . | . | . | 6p21.33 | rs112168410 |
| **chr6** | 31335791 | 31335791 | G | A | intergenic | HLA-B;MICA-AS1 | dist=10835;dist=26275 | . | . | . | 6p21.33 | rs57989216 |
| **chr6** | 31342005 | 31342005 | T | G | intergenic | HLA-B;MICA-AS1 | dist=17049;dist=20061 | . | . | . | 6p21.33 | rs114118665 |
| **chr6** | 31344183 | 31344183 | G | - | intergenic | HLA-B;MICA-AS1 | dist=19227;dist=17883 | . | . | . | 6p21.33 | rs149663102 |
| **chr6** | 31345421 | 31345421 | G | A | intergenic | HLA-B;MICA-AS1 | dist=20465;dist=16645 | . | . | . | 6p21.33 | rs114170382 |
| **chr6** | 31346655 | 31346655 | G | A | intergenic | HLA-B;MICA-AS1 | dist=21699;dist=15411 | . | . | . | 6p21.33 | rs58102217 |
| **chr6** | 31351299 | 31351303 | AAAAC | - | intergenic | HLA-B;MICA-AS1 | dist=26343;dist=10763 | . | . | . | 6p21.33 | rs113223977 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 31351940 | 31351940 | G | T | intergenic | HLA-B;MICA-AS1 | dist=26984;dist=10126 | . | . | . | 6p21.33 | rs114607072 |
| **chr6** | 31363555 | 31363557 | TTT | - | ncRNA_intronic | MICA-AS1 | . | . | . | . | 6p21.33 | rs201415628 |
| **chr6** | 31365110 | 31365110 | C | G | ncRNA_intronic | MICA-AS1 | . | . | . | . | 6p21.33 | rs138099588 |
| **chr6** | 31371071 | 31371071 | C | A | intronic | MICA | . | . | . | . | 6p21.33 | rs59440261 |
| **chr6** | 31373018 | 31373018 | G | A | intronic | MICA | . | . | . | . | 6p21.33 | rs202036493 |
| **chr6** | 31373445 | 31373445 | T | C | intronic | MICA | . | . | . | . | 6p21.33 | rs2923003 |
| **chr6** | 31378257 | 31378257 | G | A | intronic | MICA | . | . | . | . | 6p21.33 | rs144888775 |
| **chr6** | 31378864 | 31378864 | A | G | exonic | MICA | . | nonsynonymous SNV | MICA:NM_000247:exon3:c.A341G:p.Q114R,MICA:NM_001177519:exon3:c.A341G:p.Q114R,MICA:NM_001289152:exon3:c.A50G:p.Q17R,MICA:NM_001289153:exon3:c.A50G:p.Q17R | . | 6p21.33 | rs41558312 |
| **chr6** | 31380003 | 31380003 | G | - | splicing | MICA | NM_001289153:exon4:c.601+1G>-;NM_001289152:exon4:c.601+1G>-;NM_001289154:exon4:c.478+1G>-;NM_001177519:exon4:c.892+1G>-;NM_000247:exon4:c.892+1G>- | . | . | . | 6p21.33 | . |
| **chr6** | 31382495 | 31382495 | G | A | intronic | MICA | . | . | . | . | 6p21.33 | rs140810304 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 31383613 | 31383613 | A | G | downstream | MICA | dist=521 | . | . | . | 6p21.33 | rs115841246 |
| **chr6** | 31387448 | 31387448 | G | T | intergenic | MICA;LINC01149 | dist=4356;dist=21996 | . | . | . | 6p21.33 | rs79479695 |
| **chr6** | 31387667 | 31387667 | T | A | intergenic | MICA;LINC01149 | dist=4575;dist=21777 | . | . | . | 6p21.33 | rs75119533 |
| **chr6** | 31387810 | 31387810 | C | T | intergenic | MICA;LINC01149 | dist=4718;dist=21634 | . | . | . | 6p21.33 | rs79411911 |
| **chr6** | 31389740 | 31389740 | G | A | intergenic | MICA;LINC01149 | dist=6648;dist=19704 | . | . | . | 6p21.33 | rs116081995 |
| **chr6** | 31390266 | 31390266 | G | A | intergenic | MICA;LINC01149 | dist=7174;dist=19178 | . | . | . | 6p21.33 | rs116419909 |
| **chr6** | 31390361 | 31390361 | - | AC | intergenic | MICA;LINC01149 | dist=7269;dist=19083 | . | . | . | 6p21.33 | rs113265260 |
| **chr6** | 31390365 | 31390365 | T | C | intergenic | MICA;LINC01149 | dist=7273;dist=19079 | . | . | . | 6p21.33 | rs201023435 |
| **chr6** | 31390971 | 31390971 | G | A | intergenic | MICA;LINC01149 | dist=7879;dist=18473 | . | . | . | 6p21.33 | rs116339333 |
| **chr6** | 31400137 | 31400137 | A | G | intergenic | MICA;LINC01149 | dist=17045;dist=9307 | . | . | . | 6p21.33 | rs138130755 |
| **chr6** | 31400705 | 31400705 | G | A | intergenic | MICA;LINC01149 | dist=17613;dist=8739 | . | . | . | 6p21.33 | rs138117378 |
| **chr6** | 31402358 | 31402358 | A | G | intergenic | MICA;LINC01149 | dist=19266;dist=7086 | . | . | . | 6p21.33 | rs148792134 |
| **chr6** | 31405128 | 31405128 | T | A | intergenic | MICA;LINC01149 | dist=22036;dist=4316 | . | . | . | 6p21.33 | rs112689184 |
| **chr6** | 31409677 | 31409677 | A | C | ncRNA_exonic | LINC01149 | . | . | . | . | 6p21.33 | rs140991764 |
| **chr6** | 31411714 | 31411714 | T | G | ncRNA_intronic | LINC01149 | . | . | . | . | 6p21.33 | rs74655380 |
| **chr6** | 31412271 | 31412271 | G | T | ncRNA_intronic | LINC01149 | . | . | . | . | 6p21.33 | rs77311173 |
| **chr6** | 31421547 | 31421547 | A | G | intergenic | LINC01149;HCP5 | dist=6797;dist=9410 | . | . | . | 6p21.33 | rs144027808 |
| **chr6** | 31430060 | 31430060 | G | T | upstream | HCP5 | dist=897 | . | . | . | 6p21.33 | rs115846244 |
| **chr6** | 31431780 | 31431780 | T | G | ncRNA_exonic | HCP5 | . | . | . | . | 6p21.33 | rs2395029 |
| **chr6** | 31445771 | 31445771 | C | A | intergenic | HCG26;MICB-DT | dist=5586;dist=1924 | . | . | . | 6p21.33 | rs3021366 |
| **chr6** | 31452292 | 31452292 | G | A | ncRNA_intronic | MICB-DT | . | . | . | . | 6p21.33 | rs2905741 |
| **chr6** | 31453290 | 31453290 | G | A | ncRNA_intronic | MICB-DT | . | . | . | . | 6p21.33 | rs2905736 |
| **chr6** | 31453711 | 31453711 | A | C | ncRNA_intronic | MICB-DT | . | . | . | . | 6p21.33 | rs2905734 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 31458936 | 31458936 | A | T | ncRNA_intronic | MICB-DT | . | . | . | . | 6p21.33 | rs2905725 |
| **chr6** | 31462648 | 31462648 | G | T | upstream | MICB;MICB-DT | dist=10 | . | . | . | 6p21.33 | rs41293856 |
| **chr6** | 31464091 | 31464091 | G | A | intronic | MICB | . | . | . | . | 6p21.33 | rs41293860 |
| **chr6** | 31465917 | 31465917 | G | T | intronic | MICB | . | . | . | . | 6p21.33 | rs3828917 |
| **chr6** | 31468404 | 31468404 | C | A | intronic | MICB | . | . | . | . | 6p21.33 | rs41293879 |
| **chr6** | 31472339 | 31472342 | TGGA | - | intronic | MICB | . | . | . | . | 6p21.33 | rs201725498 |
| **chr6** | 31472346 | 31472349 | TAAT | - | intronic | MICB | . | . | . | . | 6p21.33 | rs199923645 |
| **chr6** | 31474820 | 31474820 | C | T | exonic | MICB | . | nonsynonymous SNV | MICB:NM_001289160:exon4:c.C539T:p.T180I,MICB:NM_001289161:exon4:c.C506T:p.T169I,MICB:NM_005931:exon4:c.C635T:p.T212I | . | 6p21.33 | rs41293883 |
| **chr6** | 31475628 | 31475628 | G | A | intronic | MICB | . | . | . | . | 6p21.33 | rs4959077 |
| **chr6** | 31478960 | 31478960 | G | A | downstream | MICB | dist=59 | . | . | . | 6p21.33 | rs7757162 |
| **chr6** | 31479223 | 31479223 | C | T | downstream | MICB | dist=322 | . | . | . | 6p21.33 | rs7757383 |
| **chr6** | 31479535 | 31479535 | T | A | downstream | MICB | dist=634 | . | . | . | 6p21.33 | rs41293887 |
| **chr6** | 31480005 | 31480005 | G | A | intergenic | MICB;MCCD1 | dist=1104;dist=16734 | . | . | . | 6p21.33 | rs41293891 |
| **chr6** | 31481133 | 31481133 | T | C | intergenic | MICB;MCCD1 | dist=2232;dist=15606 | . | . | . | 6p21.33 | rs41293895 |
| **chr6** | 31481753 | 31481753 | G | T | intergenic | MICB;MCCD1 | dist=2852;dist=14986 | . | . | . | 6p21.33 | rs41293899 |
| **chr6** | 31482626 | 31482626 | G | A | intergenic | MICB;MCCD1 | dist=3725;dist=14113 | . | . | . | 6p21.33 | rs4959078 |
| **chr6** | 31483892 | 31483892 | A | T | intergenic | MICB;MCCD1 | dist=4991;dist=12847 | . | . | . | 6p21.33 | rs41293907 |
| **chr6** | 31492279 | 31492279 | C | T | intergenic | MICB;MCCD1 | dist=13378;dist=4460 | . | . | . | 6p21.33 | rs41293911 |
| **chr6** | 31492870 | 31492870 | T | C | intergenic | MICB;MCCD1 | dist=13969;dist=3869 | . | . | . | 6p21.33 | rs41293915 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 31496381 | 31496381 | A | T | upstream | MCCD1 | dist=358 | . | . | . | 6p21.33 | rs2516484 |
| **chr6** | 31503426 | 31503426 | T | G | ncRNA_intronic | ATP6V1G2-DDX39B | . | . | . | . | 6p21.33 | rs2516474 |
| **chr6** | 31509980 | 31509980 | G | T | ncRNA_intronic | ATP6V1G2-DDX39B | . | . | . | . | 6p21.33 | rs41293919 |
| **chr6** | 31511579 | 31511579 | G | T | ncRNA_intronic | ATP6V1G2-DDX39B | . | . | . | . | 6p21.33 | rs41293923 |
| **chr6** | 31519365 | 31519365 | C | T | intronic | NFKBIL1 | . | . | . | . | 6p21.33 | rs28732141 |
| **chr6** | 31521095 | 31521095 | C | T | intronic | NFKBIL1 | . | . | . | . | 6p21.33 | rs28732142 |
| **chr6** | 31530810 | 31530810 | G | A | ncRNA_intronic | LOC100287329 | . | . | . | . | 6p21.33 | rs28732143 |
| **chr6** | 31532142 | 31532142 | C | - | ncRNA_intronic | LOC100287329 | . | . | . | . | 6p21.33 | . |
| **chr6** | 31546495 | 31546495 | G | C | downstream | TNF | dist=383 | . | . | . | 6p21.33 | rs3093668 |
| **chr6** | 31546789 | 31546789 | T | C | downstream | TNF | dist=677 | . | . | . | 6p21.33 | rs3093726 |
| **chr6** | 31547115 | 31547115 | G | A | intergenic | TNF;LTB | dist=1003; dist=1221 | . | . | . | 6p21.33 | rs3093727 |
| **chr6** | 31556205 | 31556205 | G | A | intronic | LST1 | . | . | . | . | 6p21.33 | rs28732144 |
| **chr6** | 31558264 | 31558264 | A | G | intronic | NCR3 | . | . | . | . | 6p21.33 | rs28732145 |
| **chr6** | 31562107 | 31562107 | C | T | intergenic | NCR3;AIF1 | dist=1324; dist=20904 | . | . | . | 6p21.33 | rs28895015 |
| **chr6** | 31566109 | 31566109 | - | ATAT | intergenic | NCR3;AIF1 | dist=5326; dist=16902 | . | . | . | 6p21.33 | rs199682092 |
| **chr6** | 31569520 | 31569520 | G | A | intergenic | NCR3;AIF1 | dist=8737; dist=13491 | . | . | . | 6p21.33 | rs17207190 |
| **chr6** | 32232652 | 32232652 | C | G | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268072 |
| **chr6** | 32237926 | 32237926 | T | C | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs6913182 |
| **chr6** | 32245370 | 32245370 | A | G | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268103 |
| **chr6** | 32253101 | 32253101 | - | T | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | . |
| **chr6** | 32255238 | 32255238 | - | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | . |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **chr6** | 32257444 | 32257444 | G | C | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268147 |
| **chr6** | 32259527 | 32259527 | G | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268148 |
| **chr6** | 32263099 | 32263099 | T | C | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268149 |
| **chr6** | 32266021 | 32266021 | T | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268154 |
| **chr6** | 32266310 | 32266310 | T | C | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268155 |
| **chr6** | 32279938 | 32279938 | A | G | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268205 |
| **chr6** | 32289390 | 32289390 | A | C | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs9268234 |
| **chr6** | 32300809 | 32300809 | A | G | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3129960 |
| **chr6** | 32305979 | 32305979 | G | T | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3129900 |
| **chr6** | 32316016 | 32316016 | T | G | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3117125 |
| **chr6** | 32318610 | 32318610 | G | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3117119 |
| **chr6** | 32320153 | 32320153 | G | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3132963 |
| **chr6** | 32336187 | 32336187 | T | C | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3129934 |
| **chr6** | 32336495 | 32336495 | T | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs3129938 |
| **chr6** | 32338621 | 32338621 | - | A | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | . |
| **chr6** | 32338632 | 32338632 | - | AC | ncRNA_intronic | TSBP1-AS1 | . | . | . | . | 6p21.32 | rs138019081 |

**Table S4 List of antibodies used**

Primary antibody list

| Target protein | Origin | Manufacturer | Dilution |
|---|---|---|---|
| ATP1A1 | Mouse | Santa Cruz (C464.6) | 1/250 |
| ZO-1 | Mouse | Life Technologies (339100) | 1/200 |
| ZO-1 | Rabbit | Thermo (61-7300) | 1/100 |
| BSEP | Rabbit | Sigma (HPA019035) | 1/500 |
| Beta-catenin | Mouse | Santa Cruz (E-5, sc7963) | 1/500 |
| Albumin | Goat | Bethyl (E80-129) | 1/500 |
| HNF4a | Rabbit | Santa Cruz (sc8987) | 1/500 |
| ACTIN | n/a | Thermo (A12379) | 1/250 |
| MRP2 | Rabbit | MERCK (M8316) | 1/200 |
| CK18 | Rabbit | Abcam (AB52948) | 1/500 |
| DAPI | n/a | Life Technologies (62248) | n/a |

Secondary antibody list

| | Manufacturer |
|---|---|
| Alexa Fluor 647 Goat anti-Rabbit | Thermo/Life Technologies |
| Alexa Fluor 488 Donkey anti-Rabbit | Jackson Immunoresearch |
| Alexa Fluor 488 Goat anti-Mouse | Jackson Immunoresearch |
| Alexa Fluor 647 Goat anti-Mouse | Thermo/Life Technologies |
| Alexa Fluor 488 Donkey anti-Goat | Abcam |