University of Nottingham
UK | CHINA | MALAYSIA

Deciphering the role of BMP4 signalling and gene regulation in the specification of human liver progenitor cells

Paulina Maria Durczak MSc, BSc

School of Medicine

Biodiscovery Institute

The University of Nottingham

Principal supervisor: Dr Nicholas R.F. Hannan

Secondary supervisor: Professor Guruprasad Aithal

Thesis submitted to the University of Nottingham for the Degree of Doctor of Philosophy

March 2023

To my Mum – in loving memory.

Acknowledgments

Firstly, I would like to thank my principal supervisor Dr Nicholas R.F. Hannan for his support, guidance, and encouragement throughout my PhD journey. His calm and optimism have been invaluable to help me get through some though moments, especially following the COVID-19 pandemic. I would also like to thank my secondary supervisor Professor Guruprasad P. Aithal and the technical and administrative staff at the Biodiscovery Institute who provided me with help and support.

I would like to thank Dr Lara Lewis for excellent training in sequencing library preparations which made one of the most important experiments of my thesis a success. I thank Dr Nadine Holmes and Matthew Carlile for their expertise with ChIP and RNA seq library preparations. Dr David Onion and Dr Nicola Croxall for help with flow cytometry. I would also like to thank Dr Simon Andrews for help with RNA sequencing data analysis. And finally, I thank Dr Alexander Kondrashov for his expertise in genetic engineering.

I would like to thank members of the Hannan group for their support and making the long hours in the lab so much more entertaining. Special thanks to Dr Peggy Cho Kiu Lo for training and running the thigh ship in A88 which made the work so much smoother, and for her friendship. I would like to thank Dr Sara Cuevas-Ocana for training and support during my CRISPR work and for teaching me how to best navigate the rules of the system to maximise your work outcomes.

I could not have made it through the last five years without the collegues who became my friends: Emma Collins, Sara Bodbin, Analy Serna, Liam Reed, Peggy Lo, Maria Eleftheriou, Kate Vo and Sara Pijuan. The many lunch time hours, coffee breaks and pub visits filled with laughter were the best remedy for any lab-related disaster. I have made many great memories with you and will forever be grateful for your friendship and support.

Finally, I would like to thank my parents and extended family for always believing in me, and supporting me during the particularly difficult time of thesis writing. Thank you for your love, wellbeing check-ups and home delivered meals.

Declaration

I hereby declare that this thesis has been composed by myself and has not been submitted for any other degree previously. This project was supervised by Dr Nicholas R.R. Hannan and Professor Guru Aithal. Acknowledgements of specific procedures not performed by myself are stated; otherwise, the work described is my own.

Signature:

Date: 31st March 2023

COVID statement

The COVID pandemic started about 5 months into the last year of my project. We have only just become operational after laboratory move that disrupted the third year of my research for about 3 months, so I already felt that I was behind with my experiments. Towards the end of March, we were told to stop all experiments and the laboratories were shut down due to the national lockdown. We were told to work from home as much as possible on things like thesis introduction or articles for publishing. Being at home, pretty much alone, as social contact was limited to absolute minimum, was hard. At that time, there was not much information on how dangerous the disease really is, treatment options were few and testing for the disease was limited. I was worried about my parents back in another country as well as my wider family and friends. I was also worried about myself, every headache or other symptom prompting a worry that maybe I have the new virus. Once we were asked to leave the laboratory, I was also worried if I would be able to complete my PhD thesis. I felt I did not have enough data to write up to the level acceptable by the university. In the first few months of the first strict lockdown, there was very little information about how it will be handled. Are we going to get more time in the lab? If yes, will it be funded or not? If not, how will I be able to maintain myself during those months when I try to generate the most interesting data for my write up. I also felt the social isolation quite strongly, with all these worries in my head I could not relax in my usual ways by meeting friends or having a break from it all and fly to see my family. Although meetings with friends were organised online, it was not the same as meeting up, going to the movies or out for dinner. All of that, has affected my mental state quite badly. I was extremely unproductive, I found it difficult to concentrate on reading papers, doing data analysis or even planning a workday. I missed my family and friends and I mainly focused on staying in touch with everyone as much as possible, reading books or watching TV to take my mind off everything that worried me. Workwise I achieved very little during those 5 months the laboratory was shut down.

Although coming back to the laboratory at the end of August was a much anticipated and welcome news, it turned out not as straightforward as we would like. The time we could spend in the lab was limited compared to pre-pandemic times. The number of people working in the laboratory was also severely limited. For example, in the molecular laboratory where previously 5 or 6 people would comfortably work in one bay, only 1 person was allowed. There was a substantial amount of extra cleaning we had to do before and after working in each section, we were asked not to come within 2 meters of each other and every minute of your time in the laboratory had to be booked somewhere, and permission for weekend work had to be asked for weekly, and a number of people that could be granted it, limited. There were also almost constant problems with supplies. Initially, it was mainly gloves and pipette tips but eventually there were delivery issues with almost anything. At some point, we were a few hours of running out of ethanol necessary for disinfecting hoods before, after and during cell culture, a major component of my experimental work. Having to navigate the new reality of limited access and extra admin and cleaning has impacted my productivity and I was not able to proceed with my experiments at the normal pace. Although, very slowly and gradually the restrictions and rules were relaxed and changed as the pandemic reality changed, for the remaining time of my laboratory time it was never the same as pre-pandemic time. Before the pandemic I have really enjoyed working on my PhD thesis and I have great memories of the first two years. The pandemic has affected the experience massively, and although I have managed to gather enough data to write my thesis, the process was much more stressful and do not remember it very fondly.

Abstract

Early hepatic specification and organogenesis can be modelled in *vitro* using human induced pluripotent stem cells (hiPSCs). These models apply differentiation protocols to direct hiPSCs through all the key developmental stages to accurately reflect in *vivo* development. Bone morphogenetic protein (BMP) and fibroblast growth factor (FGF) signalling are crucial for the specification of hepatic progenitors during early liver development. While the signalling cascades of these two morphogens are well characterized, the mechanisms by which they promote hepatic cell fate choice and hepatic gene expression in anterior foregut endoderm (FE) cells is not very well understood.

In this project, we characterize hiPSCs-based model of early liver development and apply it to understand the role of BMP signalling in hepatic specification. We confirm that BMP4 signalling is also necessary for liver progenitor cells (LPCs) specification from FE during hiPSCs differentiation. Using RNA sequencing (RNA seq.) we examine transcriptome changes induced by BMP4 during the transition from FE to LPC stage. Overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA) analysis revealed early activation of hepatocyte-specific functions such as lipid and protein homeostasis, haem metabolism or coagulation, while at the same time, cell adhesion and locomotion related genes are downregulated indicating preparation for cell migration out of the forming liver bud. We also notice upregulation of all four FGF receptors upon BMP signalling indicating at possible cross talk between the two pathways. The RNA seq. also detected a number of BMP4 upregulated transcription factors (TFs), several of these TFs are known for their roles in multiple developmental processes. Among them TBX3, previously reported to have a role in hepatic specification in mice, and two other TBX family members: TBX2 and TBX20. As a preliminary screen, we used a published, optimized protocol for creating inducible knockdown hiPSC lines to assess the importance of TBX and other TFs for the process of LPC specification. Double knockdown of TBX3 and TBX20 TFs significantly disrupted the hepatic induction process as shown by decreased expression of early hepatic genes such as TTR, AFP, AAT and ALB. Further studies

are necessary to confirm and further characterize the role of TBX TFs for hepatic specification.

Our study demonstrates that application of hiPSCs derived models for the study of development can aid the understanding of molecular mechanisms driving early liver specification and improve our understanding of human embryology and organogenesis. This knowledge can also be used to created more efficient differentiation platforms that can yield more mature, functional and clinically relevant populations of hiPSC-derived hepatocytes.

Table of contents

List of figures

List of tables

Abbreviations

Abbreviation	Description
%	percentage
°C	degree Celsius
A1AT	alpha 1 antitrypsin
AAVS	adeno-associated virus integration Sit 1
Ab	antibody
ActA	Activin A
ActRII	activin type II receptor
ActRIIB	activin type IIB receptor
AFP	alpha foetal protein
AFT2	activating transcription factor 2
ALB	albumin
ALK	serine/threonine kinases
ANOVA	analysis of variance
AVE	anterior ventral endoderm
BAMBI	BMP and activin membrane bound inhibitor
BMP	bone morphogenetic protein
BMPRI	BMP receptor I
BMPRII	BMP receptor II
bp	base pairs
BRA(T)	Brachyury (also TBXT)
C/EBPα	CCAAT enhancer binding protein alpha
CAR	coxsackie/adenovirus receptor
CARMEN	(CAR)diac (M)esoderm (E)nhancer-associated (N)oncoding RNA
CAS9	CRISPR associated protein 9
cDNA	complementary cDNA
CDX2	caudal type homeobox 2
c-FOS	Fos proto-oncogene

ChIP	chromatin Immunoprecipitation
c-JUN	Jun proto-oncogene
CK	cytokeratin
c-MYC	MYC proto-oncogene, BHLH transcription factor
CRISPR	clustered regularly interspaced short palindromic repeats
CXCR4	C-X-C chemokine receptor type 4
DAPI	4′,6-diamidino-2-phenylindole
DEx	differential expression
DE	definitive endoderm
DEANR1	definitive endoderm-associated lncRNA1
DExGs	differentially expressed genes
DIGIT	divergent to GSC, induced by TGF-β family signalling
DKK-1	Dickkopf related protein 1
DMEM	Dulbecco's modified eagle medium
DMH1	dorsomorphin homolog 1
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic Acid
dNTP	deoxyribonucleotide triphosphate
DOR	dorsomorphin
DS	double strand
DSB	double strand break
DTT	dithiothreitol
EC	endothelial cells
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EMT	glycogen synthase kinase-3 beta
EOMES	eomesodermin
ERK1/2	extracellular signal-regulated kinases 1/2
ES	enrichment score
ESC	embryonic stem cells
FBS	foetal bovine serum

FDR	false discovery rate
FG	foregut
FGF	fibroblast growth factor
FGF2	fibroblast growth factor 2
Foxa1	forkhead box protein A1
Foxa2	forkhead box protein A2
FRZB	Frizzled related protein
GATA	GATA-binding factor
gDNA	genomic DNA
GFP	green fluorescent protein
GO	gene ontology
gRNA	guide RNA
GSC	goosecoid Homeobox
GSEA	gene set enrichment analysis
GSK3β	glycogen synthase kinase-3 beta
HB	hepatoblast
HBr	homebrew
HDR	homology directed repair
HEY1	hairy/enhancer-of-split related with YRPW motif protein 1
HEY2	hairy/enhancer-of-split related with YRPW motif protein 2
HG	hindgut
HHEX	haematopoietically expressed homeobox protein
hiPSC	human induced pluripotent stem cells
HLC	hepatocyte like cell
HLXB9	homeobox HB9
HM	homozygous
HNF1a	hepatocyte nuclear factor 1 homeobox A
HNF1b	hepatocyte nuclear factor 1 homeobox B
HNF4a	hepatocyte nuclear factor 4 Alpha
HNF6/OC1	hepatocyte nuclear factor 6/one cut homeobox 1
HSV	herpes simplex virus

HT	heterozygous
HULC	highly upregulated in liver cancer
JNK	c-Jun N-terminal kinases
kb	kilobases
KCP	Kielin/Chordin-like protein
KD	knockdown
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLF-4	kruppel-like factor 4
KO	knockout
LB	lysogeny broth
LGR5	leucine-rich repeat-containing G-protein coupled receptor 5
LHA	left homology arm
LIV	liver
LPC	liver progenitor cell
LSEC	liver sinusoidal endothelial cells
MAF	MAF BZIP Transcription Factor
MAPK	mitogen-activated protein kinase
MIXL1	mix paired-like homeobox
MM	master mix
mRNA	messenger RNA
MSC	multiple cloning site
MSX1	msh homeobox 1
MSX2	msh homeobox 2
NEAA	non-essential amino acids
NES	normalized enrichment score
NFW	nuclease free water
NHEJ	non homologous end joining
NKD1	Naked cuticle 1
NKD1	Naked cuticle 1
NKX2-1	NK2 homeobox 1

NOG	Noggin
OC2	one cut homeobox 2
OCT4/POU5F1	octamer-binding transcription factor 4)/POU domain, class 5, transcription factor 1
OEP	overexpression plasmid
ORA	overrepresentation analysis
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCA	principal component analysis
PCGs	protein coding genes
PCR	polymerase chain reaction
PDX1	pancreatic and duodenal homeobox 1
Pen/Strep	penicillin/streptomycin
PHH	primary human hepatocytes
PI	propidium iodide
PPC	pancreatic progenitor cell
Prox1	Prospero Homeobox 1
PSC	pluripotent stem cells
puro	puromycin
qPCR	quantitative PCR
QQ	quantile-quantile (plot)
RA	retinoic acid
RGM	repulsive guidance molecules
RHA	right homology arm
RIN	RNA integrity value
RNA	ribonucleic acid
RNAi	RNA interference
ROCKi	rho-associated protein kinase inhibitor
R-PAT M	REBLPAT monoclonal hiPSCs cell line
RPMI	Roswell Park Memorial Institute Medium
rRNA	ribosomal RNA

RT	room temperature
rt-PCR	reverse transcriptase PCR
rtRA	reverse tetracycline-controlled transactivator
RVD	repeat variable diresidues
SB	SB431542
SCR	scramble
SD	standard deviation
sgRNA	single guide RNA
shRNA	short hairpin RNA
siRNA	short interfering RNA
SLUG	aka SNAI2 (Snail Family Transcriptional Repressor 2)
SM	small molecule
SNP	single nucleotide polymorphism
SOC	super optimal broth with catabolite repression
SOX17	SRY-Box transcription factor 17
SOX2	SRY-Box transcription factor 2
SS	single strand
SSB	single strand break
SSEA-4	stage-specific embryonic antigen 4
STM	septum transversum mesenchyme
TAK1	mitogen-activated protein kinase kinase kinase 7
TALEN	transcription activator-like effector nucleases
TBX2	T-box transcription factor 2
TBX20	T-box transcription factor 20
TBX3	T-box transcription factor 3
TET	tetracycline
tetO	tet operon
TetR	tetracycline repressor
TF	transcription factor
TGF-b	transforming growth factor beta
TRA-1	podocalyxin

TRE	TET responsive element
TTR	transthyretin
TWIST	twist family BHLH transcription factor 1
VFE	ventral foregut endoderm
VLDL	very low density lipoprotein
WT	wild type
XIAP	X-linked inhibitor of apoptosis
ZFN	zinc finger nuclease

Chapter 1. Introduction

1.1. The liver

The liver is the second largest organ of the human body and its largest gland. It performs over five hundred functions including metabolism of carbohydrates, lipids and protein, as wells as storage of glycogen, triglycerides, vitamins (A, D, E, K and B_{12}) and minerals (iron and copper). It produces bile salts, which assist in the process of lipid absorption form the small intestine, proteins (e.g.: clotting factors, albumin, beta globulins) and excretes bilirubin, a metabolite of the haem group. It processes drugs and toxins, participates in the metabolism of vitamin D and urea (Tortora, 2011, Si-Tayeb et al., 2010a).

1.1.1. The gross anatomy

Most of the liver is located in the right upper quadrant of the abdomen, with part of it extending into the left upper quadrant. It is grossly divided into two main lobes separated by the falciform ligament: larger right lobe and smaller left lobe. The right lobe appears to be further divided into the quadrate and caudate lobes, but each lobe is functionally different. The gallbladder is a pear-shaped organ located between the right lobe and the quadrate lobe. The oxygenated blood is delivered to the liver by the hepatic artery. The liver also receives blood from the gastrointestinal tract via the hepatic portal vein. This blood is rich in nutrients from the gastrointestinal tract. It can also potentially contain drugs, toxins and microbes ingested with the food. Deoxygenated blood with substances processed by the liver or nutrients needed in other cells is collected via the central vein into the hepatic vein and transported to the heart (Drake, 2010).

The basic structural unit of the liver is the liver lobule (Figure 1-1). It is hexagonal in shape, with central vein in its middle and portal triad at each corner. The portal triad is made of hepatic artery, portal vein and bile duct. Hepatocytes radiate from the central vein arranged in one to two-cell thick plates interspaced with hepatic sinusoids: fenestrated, thin-walled spaces consisting of endothelial cells interspaced with Kupffer cells. Hepatic sinusoids receive a mix of

oxygenated blood delivered by arterioles of the hepatic artery and nutrient-rich blood delivered by branches of the portal vein. As the blood moves from the portal triad towards the central vein, there is an exchange of oxygen, nutrients, excreted factors and waste products, with the hepatocytes on their basolateral surface. Liver-produced proteins, hormones or nutrients can then be distributed around the body. Bile is secreted via the apical surface of hepatocytes into the bile canaliculi and carried to bile ducts within the portal triad. The blood and the bile flow in the opposite directions within the hepatic lobule (Gordillo et al., 2015, Ovalle, 2021)

1.1.2. Cellular composition

The hepatocyte is the most dominant cell type of the liver (~78% of cell population) and is responsible for performing the overwhelming majority of liver functions mentioned earlier. The next most abundant cell type in the liver are cholangiocytes, around 3% of the liver cell population. Cholangiocytes form the walls of bile ducts and control the flow of the bile and its pH. They secrete water and bicarbonate. Hepatocytes and cholangiocytes derive from a common precursor cell, hepatoblast, in the early liver development. Endothelial cells form the arteries, veins, arterioles and venuoles of the liver and help to control the blood flow within the organ. There are also endothelial cells within the liver sinusoids: liver sinusoidal endothelial cells (LSEC). These are highly specialised cells that allow the transfer of molecules between the serum and hepatocytes. LSECs also secrete cytokines, participate in blood clotting and antigen presentation. Kupffer cells are liver resident macrophages that are located within the liver sinusoids. These cells are the first line of defence against any bacteria or bacterial endotoxins that can be transported from the GI tract via the portal vein. They are part of the innate immune system response and can neutralise particles by phagocytosis or pinocytosis.

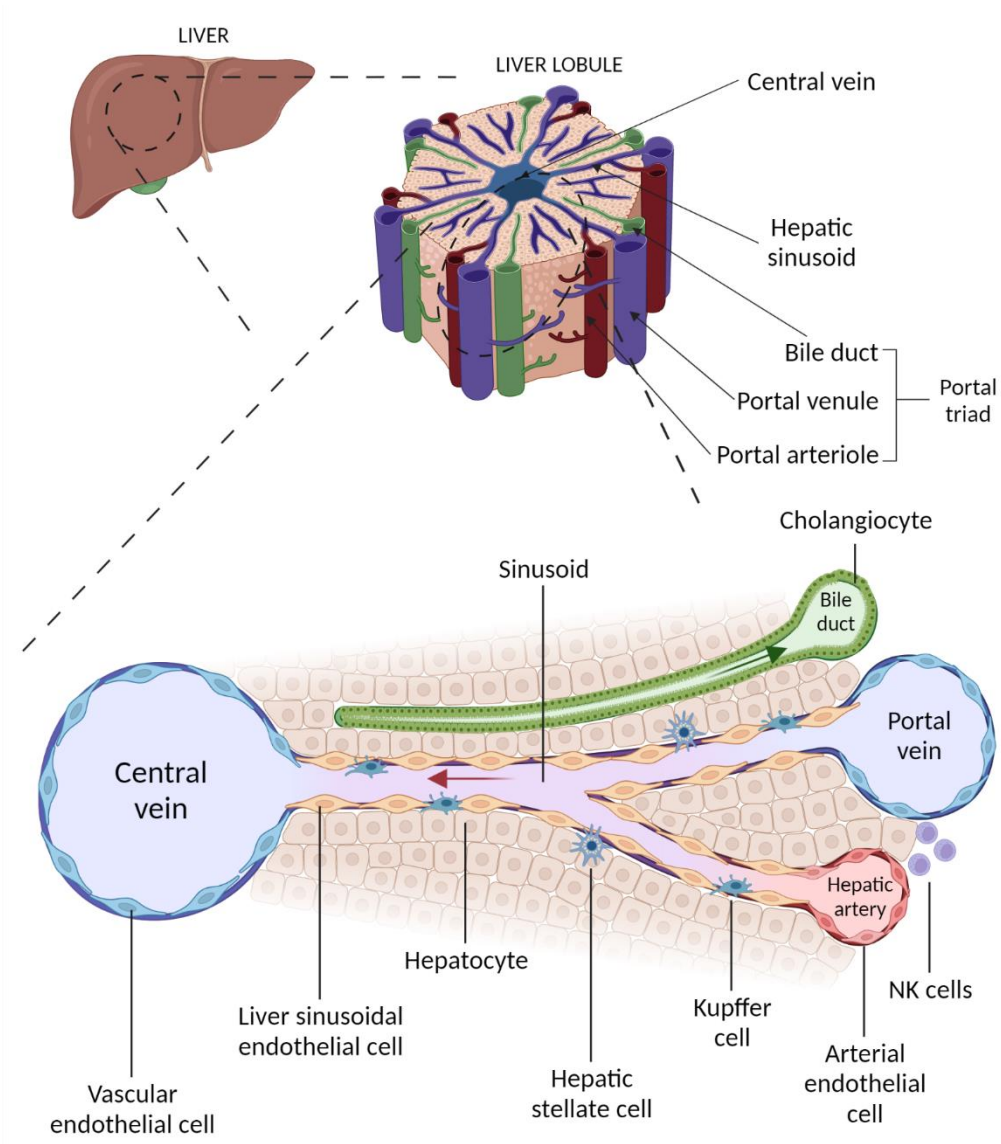

Figure 1-1 Structure of the hepatic lobule and hepatic sinusoid.
Image created using Biorender templates.

They can also secrete pro-inflammatory or anti-inflammatory cytokines and proteinases as a defence mechanism. Pit cells are another part of the innate immune system resident in the liver. These are natural killer cells that respond to intracellular pathogens and tumour formation and have cytotoxic properties. Hepatic stellate cells reside around the liver sinusoids and help in the maintenance of the extracellular matrix (ECM). They store vitamin A and its related forms, control muscular tone and contribute towards the regenerative response to tissue injury (Si-Tayeb et al., 2010a, Gordillo et al., 2015).

The intricate structure of the liver is essential for the performance of the multitude of its functions. The liver is exposed to many factors that can damage its tissue (e.g.: toxins, infectious agents). Despite liver's substantial regenerative abilities this can lead to chronic or acute liver failure.

1.1.3. Liver disease and available therapies

Diseases affecting the liver are a major burden on the health of the human population. Worldwide, around 2 million people die each year of liver diseases caused by abuse of alcohol or drugs, obesity and viral hepatitis (Asrani et al., 2019). In England alone, in 2020 10 127 people died prematurely due to liver disease (GOV.UK, 2021). Currently, the primary treatment for liver failure is transplantation. However, the need for organs far exceeds the available donated organs. Introduction of split liver transplants and partial organ transplants increased the number of the procedures. However, the need still far outnumbers the supply. Additionally, the incidence of liver disease is on the rise. In England, in the last ten years the number of patients admitted to hospital with alcohol-related liver disease increased two fold (GOV.UK, 2021). Therefore, it can be expected that the number of people needing a new liver will grow in the coming years. Cell therapy using primary human hepatocytes (PHH) is an alternative method that has already showed some promising results (Dhawan et al., 2020). Using cells instead of a whole or partial organ increases the number of people who could be treated from one donation. However, PHHs also come with some limitations. They do not proliferate in *vitro* and cannot be maintained in culture for more than 10 to 14 days (Mitry et al., 2002). Additionally, cryopreservation decreases viability and enzymatic activity of the cells (Terry et al., 2005). Immortalized hepatocyte cell lines or cancer-derived cell lines have been suggested as an alternative cell source, however, they exhibit poor function, karyotypic instability and increased resilience to toxicological insult (Szkolnicka and Hay, 2016). Therefore, production of hepatocyte from human pluripotent stem cells (PSC) offers an attractive source of virtually limitless cells available for transplantation. The applications and limitations of PSC will be discussed in more detail in section 1.4 of the introduction.

1.2. Early liver development

Animal studies on zebrafish, xenopus and mice provided most of the information we currently hold on the early liver development. The liver is derived from the definitive endoderm (DE), a cell layer that emerges from the anterior primitive streak during gastrulation. As the development proceeds, the DE layer forms a primitive gut tube that is further patterned along the anterior-posterior axis into foregut (FG), midgut and hindgut (HG) (Zaret, 2016). Studies on mice showed that the liver originates from the ventral part of the FG. That region of FG also gives rise to ventral pancreas, lungs, thyroid and stomach (Tremblay and Zaret, 2005). The emergence of the liver begins when signals from the surrounding cardiac mesoderm and septum transversum mesenchyme (STM) specify foregut endoderm (FE) to express hepatic genes such as ALB, HNF4α, or TTR (Jung et al., 1999, Rossi et al., 2001). The cells begin to thicken forming liver diverticulum surrounded by a laminin-rich basement membrane. As the process continues, the cells of the diverticulum change their morphology from a monolayer of cuboidal cells to a multilayer of pseudostratified cells called hepatoblasts (also referred to as liver progenitor cells (LPC)), expressing albumin (ALB) and alpha-fetoprotein (AFP). The basal membrane breaks down and LPCs proliferate and invade the surrounding STM forming the liver bud (Nava et al., 2005, Bort et al., 2006). Endothelial cells found in the STM surrounding liver diverticulum contribute towards the hepatoblast expansion (Matsumoto et al., 2001). Once hepatoblasts invade the STM they proliferate and differentiate into hepatocytes and cholangiocytes, cells forming the majority of liver parenchyma and biliary tract, respectively.

1.2.1. Signalling during liver progenitor cell specification

Animal studies on chick embryos have helped to establish that first liver progenitor cells derive from the ventral part of FE (Le Douarin, 1968). Studies on mouse and chick embryos showed that the surrounding mesoderm is necessary for the formation of hepatic progenitors (Houssaint, 1980). The identification of the factors released by tissues surrounding FE was possible much later. FGFs secreted by cardiac mesoderm were the first factors found to be crucial for

induction of the liver fate in the endoderm. Replacing cardiac mesoderm with either FGF1 or FGF2 signalling in explant cultures was sufficient to induce the expression of hepatic genes, such as AFP and TTR (Jung et al., 1999). FGFs pattern the endodermal tissue in gradient-dependant manner with low concentrations specifying the liver and high concentrations specifying the lung (Serls et al., 2005). Although those studies were done on mice and chicks, their findings can be translated to human studies as FGF signalling has been shown to be necessary for the differentiation of hiPSCs towards the hepatic fate (Twaroski et al., 2015). BMPs secreted from the septum transversum mesenchyme (STM) were the next factors identified as crucial for the induction of liver progenitor cells from the FE in mice. BMP signalling acts in synergy with FGF signalling to bring about the expression of hepatic genes, and both signals are necessary for the specification of liver progenitor cells (Rossi et al., 2001). Those two signals are crucial for liver development in mice, chicks, Xenopus and zebrafish (Shin et al., 2007, Chen et al., 2003).

Wnt signalling is also implicated in the hepatic specification although not, as is the case in BMP4 and FGFs, as an inductor but as a repressor. Canonical Wnt signalling in the posterior endoderm blocks the expression of an important hepatic TF Hhex. When Wnt signalling is blocked, there is an ectopic liver development in the posterior endoderm. Therefore, expression of Wnt antagonists in the anterior endoderm is necessary for its ability to commit to hepatic fate (McLin et al., 2007).

1.2.2. Transcription factors in hepatic specification

Transcription factors (TF) are DNA-binding proteins that have a crucial role in the regulation of gene expression. While the signalling pathways governing the specification of the liver have been identified, and the signalling cascades that are activated upon binding of signalling molecules to their respective receptors are well described, the molecular events that bring about the changes in FE initiated by FGF and BMP signalling are not well known. Many studies attempted to understand the molecular events behind hepatocyte and biliary epithelial cell (BEC) development by identifying factors enriched in those cells (Cereghini,

1996). Building on that work, later knockout studies in mice and zebrafish have reported a role for some of them in the early specification of the liver. At the earliest stages of establishing the hepatic competence within the FE, Foxa1 and Foxa2 were identified as crucial factors. There is a redundancy between them as single knockout of either of these TFs does not affect hepatic development, but double knockout completely prevents hepatic specification (Lee et al., 2005). Similarly, knockout of HNF1β prevents the acquisition of hepatic competency by the FE. The removal of this factors prevents the formation of liver bud and albumin expression in the FE predicted to develop into the liver. Pancreatic development is also affected, as the ventral part of the pancreas fails to be specified (Lokmane et al., 2008). Additionally, knockout of this factor after hepatic specification severely affects the development of intra hepatic biliary ducts (IHBD) (Coffinier et al., 2002). These three TFs are so far the only ones with major roles at the very beginning of hepatic development. Other TFs identified in hepatocytes are important in the later stages of liver development such as hepatoblast proliferation, migration and differentiation (Table 1-1). Interestingly, HNF4α has been shown in mice studies to be an important regulator of hepatoblast differentiation but not necessary for hepatic specification (Li et al., 2000). However, human embryonic stem cell (hESCs) differentiation to hepatocyte-like cells (HLCs) seems to require this TF. Knocking down of HNF4α in hESC completely prevented their differentiation to hepatocyte-like cells at the hepatic specification stage (Delaforest et al., 2011). It is unclear whether the reason for this discrepancy is due to species differences or due to limitations of 2D, in *vitro* modelling of the development. However, the use of simplified models allows for a more detailed study of the molecular mechanisms governed by the TFs of interest. For example, further exploration of the role of HNF4α in the differentiation of hESCs to HLCs revealed that this factor is responsible for recruitment of RNA polymerase II to the promoters of multiple genes activated during hepatoblast specification (Delaforest et al., 2018). Therefore, although animal studies were invaluable in identifying many genes crucial for early development, it is still important to establish, on human

models of development, if the findings translate between the species. Human PSCs provide species relevant model and an accessible system for the study of molecular mechanisms.

Transcription factor	Development model	Phenotype	Reference
Gata6	Tetraploid embryo complementation with Gata6 null ESCs;	• Normal hepatic specification assessed by expression of Afp, Alb, Hnf4,Rbp4, Ttr at E8.0; • Arrested liver bud development;	(Zhao et al., 2005)
Gata4	Tetraploid embryo complementation with Gata6 null ESCs;	• Normal hepatic specification assessed by expression of Afp, Alb, Hnf4,Rbp4, Ttr at E8.0; • Arrested liver bud development	(Watt et al., 2007)
Foxa1 and Foxa2	Mice with Foxa1 null allele and conditionally deleted Foxa2 using Cre-LoxP system with Cre under the control of the Foxa3 promoter;	• No liver bud formation; • Failure of hepatoblast specification (no competence of ventral foregut for induction of hepatic genes);	(Lee et al., 2005)
Hnf1β	Conditional deletion of first exon of HNF1β gene using Cre-LoxP system with Cre under the control of AlfpCre transgene;	• Severe defect in development of small and large intra hepatic biliary ducts; • Decreased expression in genes involved in fatty acid oxidation;	(Coffinier et al., 2002)
Hnf1β	Tetraploid embryo complementation with null Hnf1β ESCs,	• No Alb expression at E8.5 • No pancreatic marker expression • No liver bud formation • Reduced liver size with cells lacking hepatoblast characteristics; • No hepatic markers expression;	(Lokmane et al., 2008)
Prox1	Functional inactivation of Prox1 gene by in frame insertion of the β-galactosidase gene;	• Defect of hepatocyte proliferation; • Failure of hepatocyte migration from the hepatic bud into STM; • Failure to degrade the membrane surrounding the hepatic bud; • Persistently high E-Cad expression;	(Sosa-Pineda et al., 2000)
Prox1	Conditional deletion of Prox1 using Cre/LoxP system with Cre expressed from Foxa3 promoter; Prox1 deleted after hepatoblasts migrate out of the liver bud;	• Impaired hepatocyte differentiation (reduced levels of HNF4α and increased levels of HNF6 and HNF1b); • Increase in expression of biliary transcripts (Sox9, Lamb1 and Krt19); • Hepatoblast differentiation skewed towards biliary fate;	(Seth et al., 2014)

Hhex	Transgenic mice created from Hhex null ESCs;	• Hepatoblast specification and proliferation not affected; • Failure of migration into STM and formation of the liver bud;	(Martinez Barbera et al., 2000)
Hhex	Transgenic mice with null mutation in the Hex gene;	• Specification and proliferation of liver progenitors is initiated but the cells fail to migrate into the STM; • Hepatocyte differentiation is disrupted (no AFP or HNF3b expression by E10.5;	(Bort et al., 2006)
Hhex	Conditional Hhex knockout using Cre-LoxP system with Cre expressed from AlfpCre transgene;	• Abnormal hepatoblast differentiation and disruption of liver architecture; • Abnormal development of extra-hepatic and intrahepatic biliary ducts;	(Hunter et al., 2007)
Hnf4α	Tetraploid embryo complementation with Hnf4α null ESCs,	• Hepatoblast specification not affected; • Failure of hepatocyte differentiation: expression of genes associated with mature hepatocyte function was undetectable (apoAI, apoAII, apoB, apoCIII, apoCII, aldolase B, pAH, LFABP, transferrin, RBP, Epo)	(Li et al., 2000)
Hnf4α	Transgenic mice: Hnf4α knockout conditional on activation of Alb promoter and Afp enhancer via Cre-LoxP system	• Hepatoblast specification not affected; • Small, round and loosely associated hepatocytes with impaired glycogen storage capacity (barely detectable levels of glycogen synthase enzyme) • Impaired cell-cell contact (low levels of E-cadherin, and Ceacem1)	(Parviz et al., 2003)
HNF4α	Human ESCs expressing siRNAs against HNF4α; hESC differentiation to hepatocyte-like cells;	• Failure of hepatocyte specification during hESC differentiation; • Loss of hepatic identity of differentiated cells;	(Delaforest et al., 2011)
Tbx3	Tbx3 null mouse embryos;	• Defect of hepatoblast proliferation • Hepatoblast differentiation skewed towards biliary fate;	(Suzuki et al., 2008)
Tbx3	Tbx3 null mouse embryos;	• Reduced hepatoblast proliferation and failure to delaminated from the liver bud; • Decrease in expression of hepatic markers (Hnf4α and Cebpα) • Hepatoblast differentiation skewed towards biliary fate;	Ludtke et al., 2009
C/EBPα	C/EBPα null mice	• Impaired hepatocyte maturation; hepatoblast differentiation skewed towards biliary fate;	(Akai et al., 2014)

		• Abnormal bile duct morphogenesis • Decrease in HNF1α and HNF4α and increase in SOX9, E-Cad Hnf6 and HNF1β mRNA expression;	
HNF1α	Conditional HNF1α knockout using the Cre-LoxP system; Cre under the control of Ella promoter expressed during the early mouse embryo;	• Enlarged liver with degenerating hepatocytes at 12 weeks of age; • Disrupted hepatocyte function;	(Lee et al., 1998)

Table 1-1 Summary of the major studies of TFs involved in the early stages of liver development.

1.3. BMP signalling pathway

1.3.1. Ligands, receptors and mediators of the BMP pathway

Bone morphogenetic proteins (BMPs) are members of the transforming growth factor β (TGF-β) family of signalling molecules. They were named for the ability to induce ectopic bone formation, first observed in the 19th century. Since their isolation and cloning in the 1980s, they have been shown to have many diverse biological functions, e.g.: in the development of the kidneys, liver, skeletal system, hair follicles or teeth, differentiation of osteoblasts and chondrocytes, iron metabolism and cancer (Katagiri and Watabe, 2016). They are divided into subgroups according to structural homology (Figure 1-2).

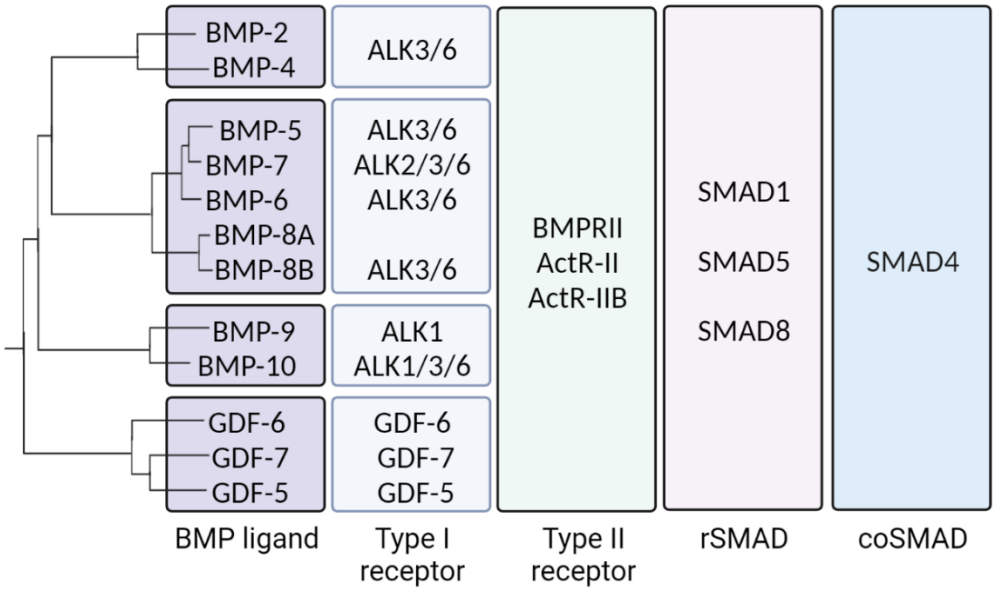

Figure 1-2 BMP family ligands, receptors and SMAD mediators.
Image adapted from (Nickel and Mueller, 2019)

BMPs are synthetised as inactive, pro-polypeptides. The mature form is at the C-terminal of the molecule, and a signal peptide is located at the N-terminal end. The two are separated by a pro-domain. The mature part of the BMP is enzymatically cleaved by a proteinase, e.g.: furin cleaves BMP4 (Nelsen and Christian, 2009). Mature BMP proteins have seven cysteine residues: six form

intramolecular disulfide bridges, and the 7th residue covalently binds via a disulfide bond to another molecule of mature BMP to form a biologically active, homodimeric or heterodimeric, ligand that can activate BMP receptors (Bragdon et al., 2011).

In canonical pathway, BMPs induce their effects by binding to type I and type II serine-threonine kinase transmembrane receptors (Figure 1-3). Although BMP ligands can bind to and activate type I receptors directly, their potency is much increased in the presence of type II receptors. There are three type II receptors that BMPs bind to: BMP type II receptor (BMPRII), which is specific for BMP molecules, activin type II receptor (ActRII) and activin type IIB receptor (ActRIIB), which are shared with activins and myostatin. From the seven type I receptors that are bound by the TGF-β signalling family, four are used by the BMP subgroup: ALK-1, ALK-2, ALK-3 and ALK-6 (Katagiri and Watabe, 2016).

To elicit a cellular response, BMPs bind to type II receptors on the surface of the cell. Type II receptors, which are constitutively active, phosphorylate type I receptors at the glycine-serine rich domain of the intracellular part of the receptor. Activated type I receptors phosphorylate SMAD proteins.

There are eight SMAD proteins (SMAD 1-8) identified in mammals. In BMP signalling, receptor related SMADS (rSMADS): SMAD1, SMAD5 and SMAD8 are phosphorylated by type I receptors activated by BMPs. Once activated, rSMADs form a heterotrimeric complex with SMAD4. SMAD4 is a common partner SMAD (coSMAD) utilized also in the TGF-β signalling. rSMAD-coSMAD complexes relocate to the nucleus where they associate with transcriptional activators (e.g.: CBP, GATA4/5/6, TCF4) or repressors (Gli3, ZEB2, DACH1, KLF4) and regulate transcription of target genes by binding to their regulatory elements (Ampuja and Kallioniemi, 2018).

BMPs can also signal via a non-canonical route that it not SMAD mediated. Instead, BMP type I receptors can be linked to protein ligase X-linked inhibitor of apoptosis (XIAP) that activates TAK1. TAK1 is a member of the MAPK family. It phosphorylates downstream MAP kinases such as p38, ERK1/2 and JNK.

Activated kinases re-locate to the nucleus where they activate TFs ATF2, c-JUN and c-FOS and induce changes in transcriptional activity (Zhang and Que, 2020).

1.3.2. Regulators of BMP signalling

BMP signalling can be controlled at several levels. There are inhibitors and potentiators that can influence the pathway at extracellular, receptor and intracellular level.

BMP ligands can be directly bound by proteins and prevented from interacting with their receptors. Noggin, Chordin, Gremlin and Cerberus are just some examples of extracellular BMP antagonists (Brazil et al., 2015). At the receptor levels, BMP signalling can be limited by expression of BAMBI, a pseudoreceptor for the TGFβ family. BAMBI competes with BMP receptors for BMP ligands but does not have the intracellular domain that can phosphorylate type I receptors and activate SMAD proteins (Onichtchouk et al., 1999). Further along the signalling pathways, BMP activated SMADs can be antagonised by inhibitory SMADS (iSMAD): SMAD6 and SMAD7. Additionally, SMURF1 and SMURF2 ligases can induce ubiquitination and degradation of SMAD1 and SMAD5. And finally, BMP-mediated gene expression can be negatively controlled by miRNAs and methylation (Brazil et al., 2015).

There are also several potentiators of BMP signalling. BMP1 can cleave Chordin bound BMP ligands and as such act as an activator of BMP pathway. Sulfated polysaccharides, such as heparin, have been reported to potentiate BMP2, BMP4 and BMP7 signalling in osteoblast differentiation. Kielin/Chordin-like protein (KCP) and proteins of the repulsive guidance molecule (RGMa, RGMb, RGMc) have also been reported to positively control BMP signalling (Katagiri and Watabe, 2016).

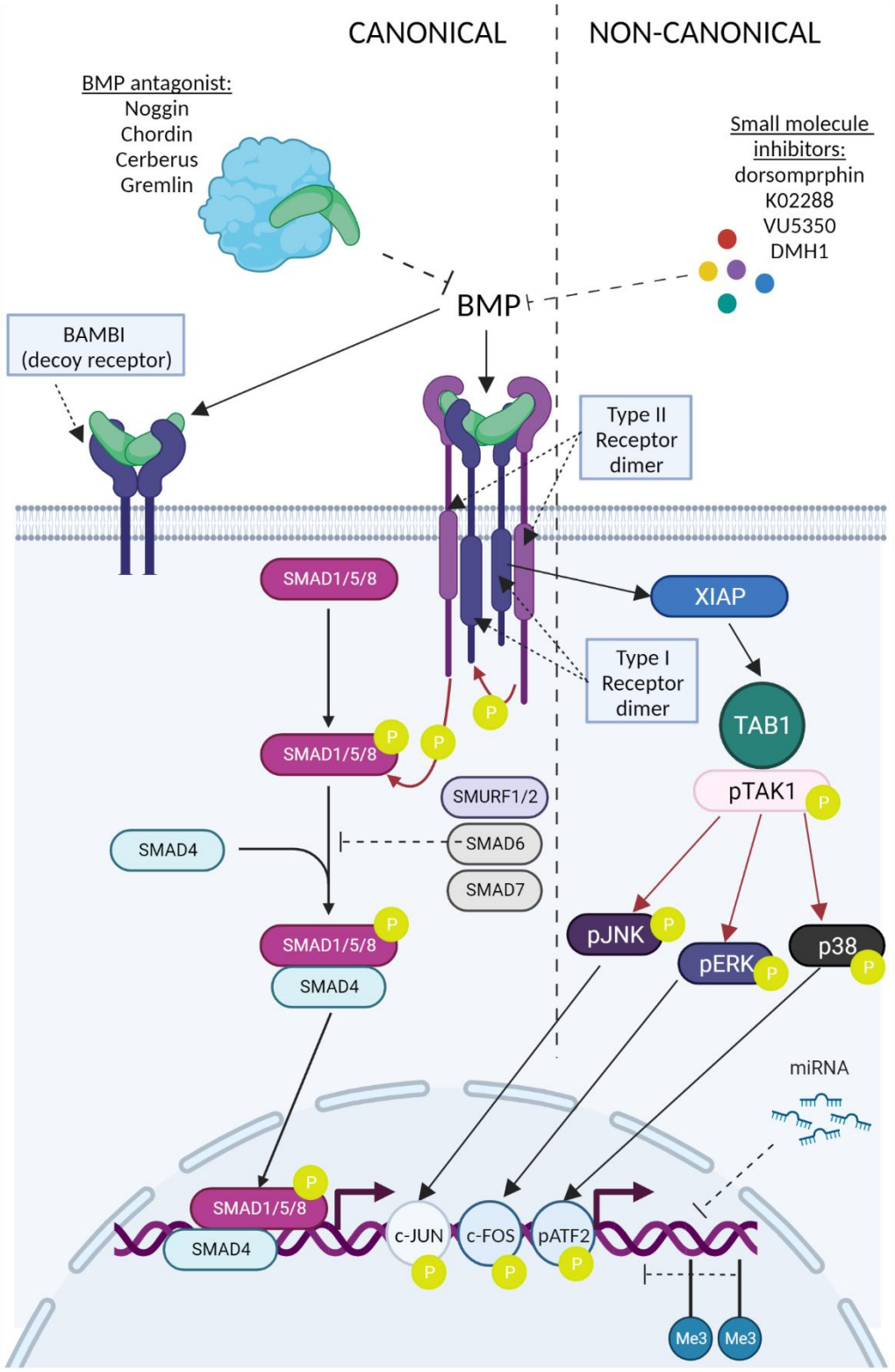

Figure 1-3 BMP signalling cascade.
Image created using Biorender templates.

As BMPs are involved in multiple physiological functions, that can be dysregulated in disease states, it was desirable to identify small molecule inhibitors of this signalling pathway, to potentially use them in clinical settings.

Several selective inhibitors of BMP receptors have been identified such as dorsomorphin, K02288, VU5350 or DMH1 (Sanchez-Duffhues et al., 2020). These can also be applied in in *vitro* experiments that study the role of BMP molecules in various biological processes.

As mentioned earlier, BMP signalling plays a vital role in the specification of LPCs during liver development, but this pathway is also involved in many other developmental processes. How the specific genetic program is activated by the same SMAD molecules is not clear.

1.4. Pluripotent stem cells

Pluripotent stem cells (PSCs) are characterized by the ability to differentiate to representative cell types of all three germ layers: mesoderm, endoderm and ectoderm but typically do not have the capacity to form extra-embryonic tissue. They are also self-renewing, meaning they can be propagated in vitro virtually indefinitely. They express a panel of pluripotency-related TFs such as OCT4, NANOG or SOX2, and surface markers e.g.: TRA-1-60, SSEA-4 and SSEA-3 (De Los Angeles et al., 2015). Although several types of PSCs have been derived and characterised, embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) are the most common types of PSCs.

Human embryonic stem cells (hESCs) were first derived in 1998 from a pre-implantation embryo (Thomson et al., 1998). Despite their great potential for application in cell therapy or drug testing, some countries completely banned their use on the ethical basis as hESCs creation requires destruction of an early human embryo (Walters, 2004). However, several years later human PSC cells were obtained by reprogramming of differentiated somatic cells into PSCs by introduction of four transcription factors: OCT4, SOX2, Klf-4 and c-Myc (Takahashi et al., 2007). Named 'induced' PSCs, these cells show the same characteristics as ESCs: ability to differentiate to cell types of all three germ layers and self-renewal but lack the ethical issues and legislative barriers that hinder research on human PSCs. Additionally, they present a possibility of personalised cell therapy or ability to study genetically inherited disorders. Since

their first creation, research on this type of PSCs has increased significantly (Figure 1-4).

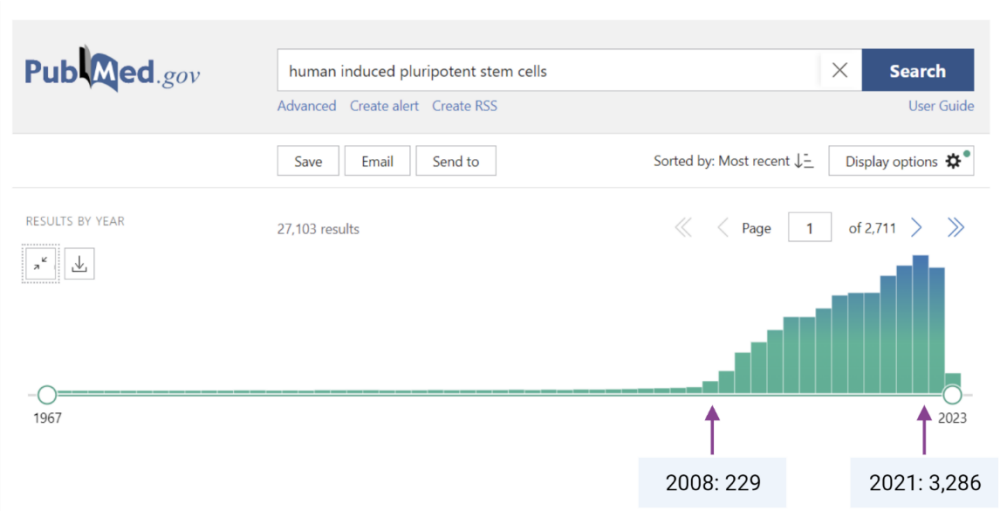

Figure 1-4 Number of publications relating to hiPSCs.
Pub med search using 'human induced pluripotent stem cells' demonstrates rapid increase in research interest in this field.

1.4.1. Application of pluripotent stem cells

Pluripotent stem cells (both embryonic and induced) are a possible cell source for many applications. Due to their self-renewal and a potential to differentiate to virtually any cell type of the body, they present an ideal source of almost limitless cells for basic research or cell-based therapies. The 'holy grail' of PSC research is the use of PSCs or PSC-derived cells for therapy. Intense efforts to bring PSCs to the clinic resulted in several early phase clinical trials that used PSCs in the treatment of cardiovascular and neurological disease, malignancy or viral infections with COVID-19 (Kim et al., 2022).

Although widespread use of PSC-derived therapies is not here just yet, hPSCs have found other applications. In the year following reprogramming of human somatic cells to pluripotency, several hiPSCs lines derived from patients with monogenic, complex and chromosomal genetic disorders were created for the purpose of disease modelling (Park et al., 2008). Since then, hiPSCs were used in the study of many genetic disorders, such as Duchenne muscular dystrophy, cystic fibrosis, cardiomyopathies or metabolic disorders (Morera et al., 2022,

Wong et al., 2012, Lee et al., 2014). PSCs are also a great source of healthy human cells for the study of viral disease, especially species-specific viruses where animal models cannot be applied. Primary tissues have been used but donor shortages, variability and difficulty in the long-term cell culture are limiting factors. PSCs-derived hepatocytes were used in the study of hepatitis B and C (Xia et al., 2017, Yoshida et al., 2011). Other cell types were also used, e.g.: cardiomyocytes in the study of coxsackievirus B3-induced myocarditis (Sharma et al., 2014) and sensory neurons for the study of infection with herpes simplex virus 1 (D'Aiuto et al., 2015). PSCs-based models allow to examine the molecular mechanisms of viral entry and spreading, immune sensing of the viral infection or signalling pathways that may lead to virus-associated long-term complications, such as hepatocellular carcinoma (Xia et al., 2017). Additionally, those cells can also be used for identification of novel treatments for viral infections, which brings us to the next application of PSCs: drug discovery and safety testing.

The availability of disease models not only allows the study of the pathology of the disease but also the identification of novel drug therapies. The accessibility of vast numbers of cells, that are affected by a condition or infection, enables high throughput screening of drug libraries containing thousands of candidate compounds and identification of potential treatments (Kaufmann et al., 2015). Furthermore, PSCs-derived cells can assist in the process of drug development. Traditional methods for assessing drug toxicity rely on animal models, which are expensive, low throughput, do not always accurately predict toxicity due to inter-species differences and come with ethical issues about animal treatment (Daston et al., 2022). The use of primary human tissue is limited for the same reasons as its use for disease modelling: scarcity, variability and difficulty in cell culture. Immortalized cell lines do not keep their exact phenotype and are affected by epigenetic alternations (Maqsood et al., 2013). Human PSC-derived cells pose a great alternative for supply of healthy cells, especially hepatocytes and cardiomyocytes, which are among the most sensitive cell types affected by drug toxicity (Weaver and Valentin, 2019). Hepatotoxicity or cardiotoxicity are

the most common reasons for drug failure in clinical trials or their withdrawal from the market (Solotke et al., 2018). Therefore, more accurate models for assessing toxicity could decrease animal use in the drug development process, reduce the risk for clinical trial participants, shorten the time and cut the cost of the process and limit the number of drugs that fail the trials.

Lastly, PSCs have also allowed the study of development. The use of human embryos for the study of early human development is strictly controlled and limited due to the ethically sensitive nature of such research. Currently, the UK law limits the timeframe for which human embryos can be kept alive using available technologies to 14 days. This limits in vitro research to pre-implantation stage before any significant organogenesis occurs (Carlson, 2019). Although recently, the ISSCR has recommended an extension of the 14-day rule, provided robust review process is in place, this area of research can still be highly controversial and poses ethical questions. The emergence of pluripotent stem cells, both embryonic and induced, delivered and alternative method for understanding the molecular mechanisms of early human development. They have been applied to the study of skin (Oceguera-Yanez et al., 2022), neurological (Knock and Julian, 2021), kidney (Khoshdel Rad et al., 2020), haematopoietic (Jung et al., 2018) or cardiac development (Ramirez-Calderon et al., 2022).

1.4.2. Current limitations of pluripotent stem cells

PSCs have found multiple applications since their derivation. However, some limitations to the application of the PSC-derived cells exist. One of the main issues of PSCs-derived cells is their immaturity. Phenotypical and functional analysis showed that PSC-derived hepatocytes resemble foetal hepatocytes rather than adult ones. They express lower levels of CYP enzymes involved in drug metabolism and produce less albumin and urea compared to PHH. They also express AFP, which is absent in PHH (Baxter et al., 2015). Transcriptomic analysis of PSC-derived hepatocytes using various differentiation protocols confirmed that they are more related to foetal hepatocytes rather than the adult ones. There are significant differences in the expression profiles of fatty acid and

drug metabolism genes or gluconeogenesis-related genes. Additionally, the analysis showed incomplete loss of original cell gene expression profile (lingering of pluripotency-related genes) and undesired gain of other identities such as lung or intestine/colon. Significant variability in transcriptome between hepatocyte-like cells differentiated by various protocols was also revealed (Ardisasmita et al., 2022). Similar issues with phenotypical and functional immaturity has been reported in PSC-derived cardiomyocytes (Wang et al., 2022) or neurons (Imaizumi and Okano, 2021). The use of immature cells can be problematic for drug discovery and toxicity studies. Although some reports showed that PSC-derived hepatocytes accurately predict drug toxicity (Szkolnicka et al., 2014), the fact that they predominantly express CYP enzymes of immature hepatocytes can be problematic. The immature phenotype may also prevent the use of PSC-derived cell types in the clinic as has been shown by a study that transplanted Macaque monkeys with PSC-derived cardiomyocytes for the treatment of myocardial infarction. Despite some promising signs of infarct remuscularisation and cardiomyocyte maturation, the monkeys also presented with arrythmias that can have life threatening consequences (Chong et al., 2014).

Another issue with the use of PSC-derived cell for clinical applications is the possible presence of undifferentiated PSCs in the transplanted cells. PSCs bring the risk of teratoma or tumour formation due to their enormous ability to proliferate, presence of reprogramming factors or genetic mutations acquired during in *vitro* manipulation. Although detection of chromosomal abnormalities is quite straightforward, detection and interpretation of single nucleotide variation poses greater difficulty (Yamanaka, 2020). Highly efficient differentiation protocols and stringent purification methods are necessary to ensure that no undifferentiated PSCs or proliferative progenitors persist in the cell population given to a patient as cell therapy. Another challenge to efficient production of PSC-derived therapeutic cells is heterogeneity between the PSC lines due to genetic background or epigenetic status. This results in significant differences in the ability of those PSC lines to differentiate into a desired cell

type. (Choi et al., 2015, Koyanagi-Aoi et al., 2013). This may necessitate the adjustment of differentiation protocols for individual patients increasing the cost and extending the time of the therapy.

Better understanding of early human development is necessary for improving differentiation protocols that can render pure populations of mature cell types, without any contaminating undifferentiated PSCs or proliferative progenitor stages.

1.4.3. Genetic modification of stem cells with programmable nucleases

The application of targeted genome editing techniques to pluripotent stem cell research has opened new avenues in cell therapy, disease and development modelling and drug screening. Precise and efficient modification of the DNA has become achievable at the break of this century with the development of programmable nucleases. Zinc-finger nucleases (ZFN), transcription-activator-like effector nucleases (TALENS) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease have become major tools for genome editing. Each of these methods can introduce a double strand break (DSB) at the precise location in the genome. The DSB can be repaired by two different routes: non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ has a high mutation rate as it frequently results in point mutations or deletions/insertions. If it can be targeted to a specific gene or its regulatory element it can disrupt the expression of that gene. HDR repairs the DSB by recombination with template DNA that has homology regions with the targeted site. The DNA template can be specifically designed with specific changes to the DNA sequence. The changes can be small, like single nucleotide alternations, or very long sequences, containing coding sequences for whole genes (Baumgart and Beyer, 2017). Each of the nuclease systems can cut the DNA at a specific site, but their mechanism, specificity and efficiency differ and need to be considered when selecting one for genome editing.

ZNFs are made by fusion of two domains from two different proteins. The cleavage domain, that introduces the DSB in the DNA, is derived from a restriction enzyme FokI. The enzyme's DNA binding domain is replaced by a zinc-

finger protein that interacts with a triplet within the DNA. Many zinc fingers exist that differ in their amino acid composition and affinity for different DNA triplets. As they function as independent modules, selected zinc fingers can be brought together designed in a way to target a specific DNA sequence of interest. Usually, 3-6 zinc finger proteins are used that target DNA sequences 9 to 18 bs in length (Kim et al., 1996). It has been shown that FokI cleavage domain has one catalytic centre and, for it to cut DNA, it needs to dimerize. Therefore, two ZFN monomers must be constructed to create the active genome engineering tool. Each monomer recognizes adjacent DNA sequences on opposing DNA strands with a spacer between them of 5-7bp (Bitinaite et al., 1998). This doubles the length of recognition site increasing the specificity of the ZFNs. Improvements to the original methods have substantially increased the specificity of ZFNs. The ability of wild-type FokI to form homodimers was responsible for high levels of off-target DNA cuts. Modification of the domain to impose heterodimer formation has reduced the issue substantially (Szczepek et al., 2007). Despite the improvements, cleavage with ZFNs has variable efficiency and is often greatest in G-rich regions, which limits the number of appropriate sites for targeting. Additionally, non-commercial ZNFs frequently have high toxicity most probably due to high rate of off-target cleavage (Kim and Kim, 2014).

TALENs are also a fusion of domains from two different proteins. Like ZNFs, their cleavage domain is derived from the FokI enzyme, but they use a different class of DNA-binding domain. Transcription activator-like effectors (TALEs) come from a Xanthomonas bacterium and are composed of 33-35 amino acids repeats that recognise a single nucleotide in the major grove of the double helix. Their specificity is mediated by repeat variable diresidues (RVD): amino acids in the 12^{th} and 13^{th} location of each repeat. RVDs recognising each of the four DNA nucleotides have been identified. As ZFNs, TALENs need to dimerize to be able to cleave DNA which increases the specificity of targeted site (Kim and Kim, 2014). Additionally, the recognition of single base by each TALEN offers more flexibility in the design compared to ZNFs. The only limitation to the design of

TALENS comes from the requirement of a T residue at the beginning of the binding site. However, even though the design is more straightforward, the actual cloning of repat TALEN arrays is challenging due to the repetitive nature of the TALENs. This issue has been mostly addressed by development of cloning methods like 'Golden Gate' cloning or ligation-independent cloning and a construction of a library targeting 18,740 human protein coding genes (Gaj et al., 2013).

CRISPR/Cas9 editing tool was developed on the basis of the adaptive immune system identified in bacteria and archaea. In those organisms, regions of highly repetitive sequences separated by non-repetitive spacer DNA were identified. Later, it was discovered that the spacer DNA belonged to viruses and mobile genetic elements. These small DNA fragments (~20bp) serve as an 'address' labels for the Cas9 nucleases that provide defence against invading pathogens (Adli, 2018). Like ZFNs and TALENs, CRISPR/Cas9 system is composed of two elements: Cas9 nuclease responsible for the cutting of DNA and a single guide RNA (sgRNA). SgRNA is an artificial, simplified version of the guiding CRISPR system. In bacteria, Cas9 is guided by CRISPR RNA (crRNA) transcribed from the non-repetitive protospacer elements within the CRISPR cluster and trans-activating crRNA (tracrRNA). When Cas9 complexes with guide RNA it forms an active nuclease complex that targets specific regions of the DNA and induces DSB (Figure 1-5). Over time, the guiding RNA has been simplified into one sgRNA (Jinek et al., 2012). The specificity of each complex is determined by the short 20nt sequence in the sgRNA ('protospacer' RNA) and a PAM sequence (protospacer-adjacent motif) that is recognised by the Cas9 enzyme. Protospacer RNA can be designed to target any area of the genome. It binds to its complimentary sequence within DNA and, if it is followed by the PAM sequence, Cas9 cleaves the DNA three nucleotides into the protospacer (Jinek et al., 2012).

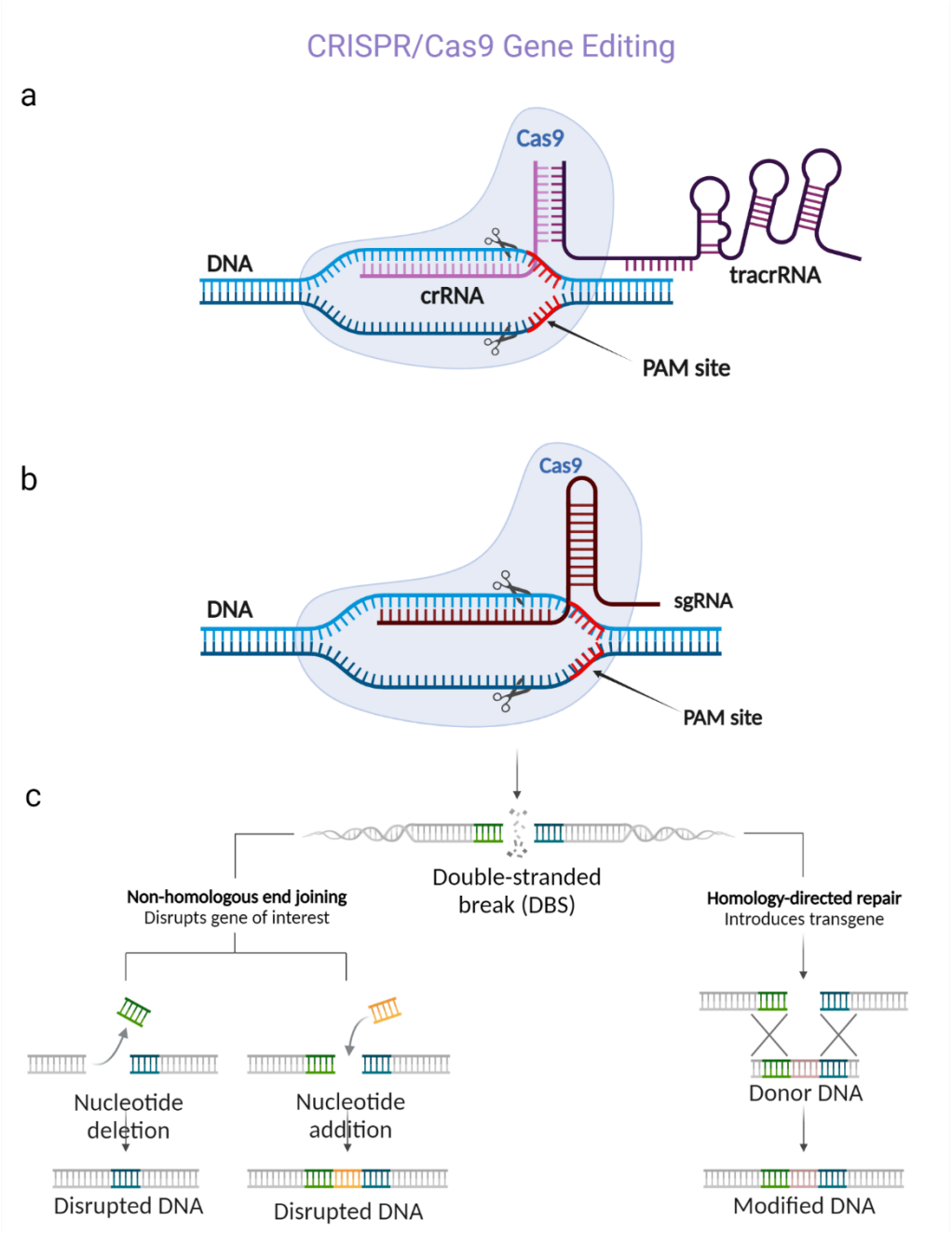

Figure 1-5 CRISPR/Cas9 gene editing tool.
a) CRISPR/Cas9 system with the initial tracrRNA and crRNA; b)CRISPR/Cas9 with simplified single guide RNA (sgRNA); c) How endogenous cell repair mechanisms of double strand break help genome editing. Image created using modified Biorender templates.

The requirement for the PAM sequence is a small limitation of this genome editing technique. The NGG (where N is any of the four nucleotides) PAM sequence recognised by the most commonly used Type II CRISPR/Cas9 appears in the genome every 8-12 bp restricting somehow the selection of target DNA.

However, new variants of the Cas9 enzymes have been identified that are less stringent about the PAM sequence (Kleinstiver et al., 2015) or have been reengineered to recognise shorter PAM (Hirano et al., 2016).

Unlike ZFNs and TALENs, CRISPR/Cas9 functions as a monomer and has been shown to induce off target effects (Cho et al., 2014). Cas9 can tolerate up to 5 mismatches between the guide RNA and targeted sequence (Fu et al., 2013). To reduce the frequency of off target effects, Cas9 with nickase activity has been developed. Cas9 has two catalytic domains: HNH and RuvC that cut complementary and non-complementary strand, respectively (Jinek et al., 2012). Inactivation of one of the domains leaves the Cas9 enzyme with nickase activity capable of introducing single strand breaks (SSB) only. It has been shown that a paired nickase approach using two guide RNAs and introducing off set SSB in close proximity on the genome significantly reduces the occurrence of off target effects increasing CRISPR/Cas9 specificity (Shen et al., 2014, Cho et al., 2014).

Despite the limitations of CRISPR/Cas9, it has become a major tool of genetic engineering due to extreme flexibility, ease of design and construction and editing efficiency. Additionally, CRISPR/Cas9 application has moved beyond the genome editing into gene expression regulation, epigenome editing or manipulation of chromatin topology (Adli, 2018).

1.4.4. Inducible gene expression systems

The study of early human development with the use of PSCs offers an invaluable opportunity for understanding the highly intricate genetic networks that govern cell differentiation and specialization. However, some TFs have dual roles in development and their knockout can result in embryonic lethality making the study of their role in later events impossible (Li et al., 2000, Zhao et al., 2005, Watt et al., 2007). Similarly, the pluripotency and self-renewal of PSCs can be affected by knockout of genes with roles in later development (Masui et al., 2007). Therefore, inducible gene expression systems can be very useful for the study of genes involved in various stages of the development. In PSCs, these systems allow temporal control of gene expression by inducing knockdowns,

knockouts or overexpression of candidate genes at the desired point in differentiation. Furthermore, gene induction or suppression can be reversible and dose-controlled (Kallunki et al., 2019).

Tetracycline-controlled operator system is a popular tool to control mammalian gene expression. It is based on antibiotic resistance mechanism of E.coli bacteria, where tetracycline repressor (TetR) protein binds to tet operons (tetO) located before promoters of resistance genes (e.g.: drug efflux genes) and blocks their expression. When tetracycline is present, it binds to TetR inducing conformational change within the protein. This change prevents TetR binding to tetO, therefore transcription of drug resistance genes can begin. The TET system adapted the tetO operons and TetR proteins to gene control in mammalian setting (Kallunki et al., 2019) Further modifications to the system created TET-OFF configuration where TetR is fused with a transactivator derived from herpes simplex virus 1, VP16. When TetR/VP16 hybrid is bound to tetO, there is activation of genes downstream of the operon, and addition of tetracycline switches gene expression off (Gossen and Bujard, 1992). TET-ON system was created by random mutagenesis of the TetR protein that identified a variant binding to TetO on addition of tetracycline (Gossen et al., 1995). This removes the need for continuous addition of tetracycline to the cell culture media, which is beneficial as tetracycline and its derivative, doxycycline, have been reported to alter the metabolism and proliferation rate of human cell lines (Ahler et al., 2013). The TET system has been successfully used for conditional expression of genes in haematopoietic differentiation (Zeng et al., 2021), as well as inducible knockdowns and knockouts in mesoderm, endoderm and ectoderm differentiation of the human PSCs (Figure 1-6) (Bertero et al., 2016).

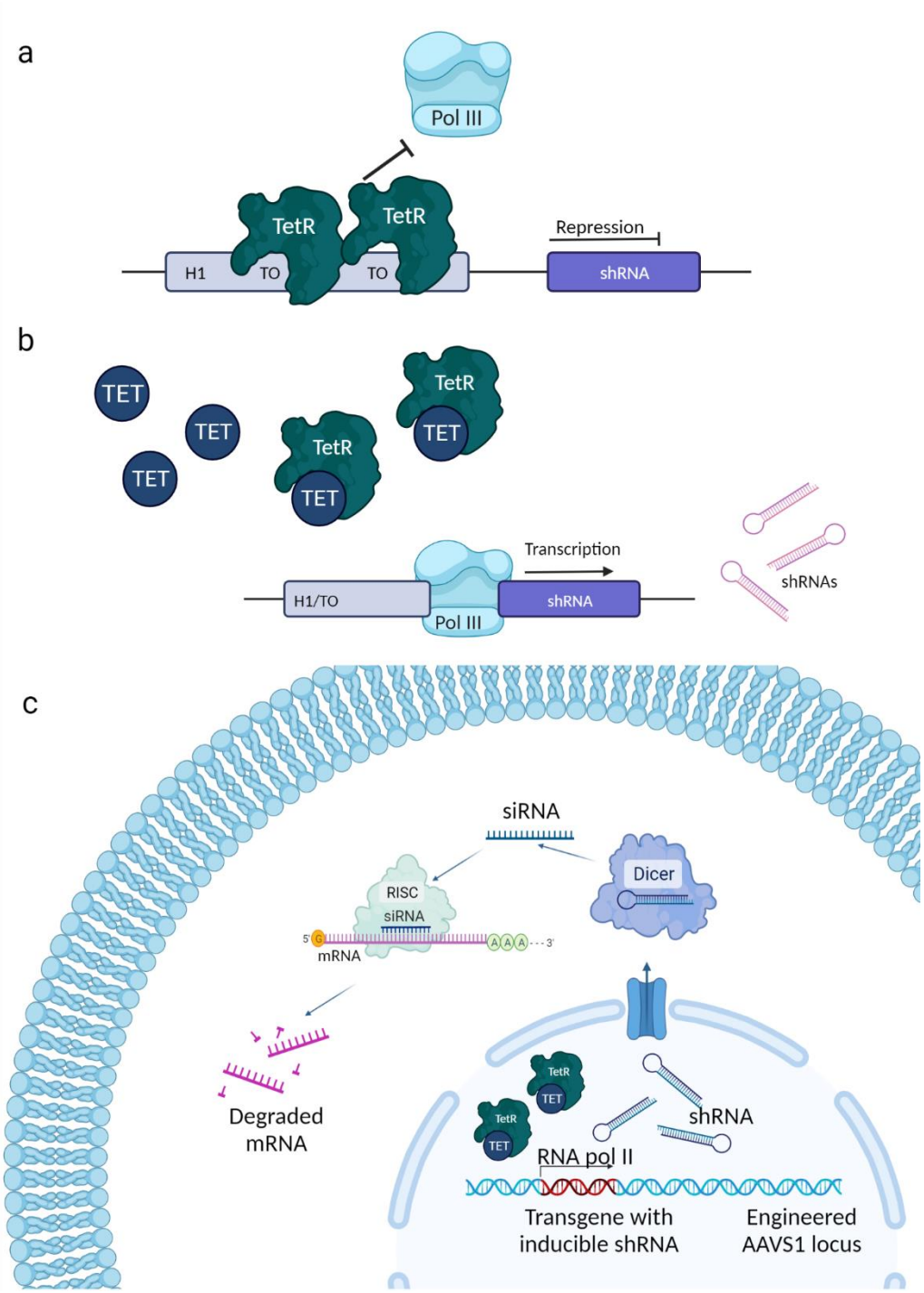

Figure 1-6 Combining the TET inducible system with RNAi interference

a) In the absence of tetracycline (TET), the tetracycline-inducible protein (TetR) binds to TO sequence within the H1 promoter. Access of Pol III to the promoter is blocked due to steric hindrance and there is no transcription of shRNAs. b) Added TET binds to TetR and induces conformational change in the protein. This change reduces TetR's ability to bind TO sequence. Pol III is able to access H1 promoter and shRNAs are transcribed c) Produced shRNAs are processed by internal cell RNA interfering pathways. Produced shRNAs knock down mRNA of the gene of interest. Images created using Biorender templates.

1.5. Aims of the thesis

BMP4 ligands secreted from the septum transversum mesenchyme have been shown to be crucial for LPC specification from foregut endoderm during mouse liver development (Rossi et al., 2001). However, the molecular mechanisms by which BMP4 induced transcriptional changes in the FE cells, as well as the identity of genes involved in downstream effects of BMP4/SMAD signalling, have not been identified. The overarching aim of this project was to gain a better understanding of this process in human development. The major stages of the project are:

I. Modelling of human early liver development with the use of human induced pluripotent stem cells differentiated to LPC using previously established differentiation protocols. Validation of the model using cell morphology assessment (microscopy), gene expression analysis (qPCR) and protein expression analysis (immunocytochemistry).
II. Examination of transcriptome changes induced by BMP4 signalling during LPC specification from foregut endoderm (RNA sequencing) and identification of direct targets of SMAD proteins during LPC specification (ChIP sequencing). Selection of candidate genes for further examination of molecular mechanisms.
III. Investigation of molecular mechanisms of BMP4 signalling. Application of CRISPR/Cas9 in the gene knockdown/knockout and overexpression to understand the role of BMP4 signalling mediators in the specification of LPCs.

Chapter 2. Materials and methods

2.1. Cell culture

2.1.1. HiPSCs

REBL-PAT Monoclonal (R-PAT M) hiPSC line was created by Dr Gary Duncan at the University of Nottingham. The cells were reprogrammed from human dermal fibroblasts using Sendai virus. The cell line was used for all cell-based experiments.

All cell culture was performed in type II Biological Safety Cabinets. Cells were maintained in humidified incubators at 37^{o}C and 5% CO_2.

2.1.2. Cell culture media

The composition of cell culture media used in this thesis is listed in Table 2-1. Once the media was made, it was stored at 4^{o}C for up to two weeks and warmed up to 37^{o}C before addition to the cells.

2.1.3. Matrigel coating

All cell culture dishes were prepared prior to cell seeding by coating with Matrigel^{TM} (Corning, #354230) at a constant concentration of 0.035mg/cm^2 and kept overnight at 37^{o}C. Before cells were added to a coated dish, excess Matrigel was removed by washing with PBS.

Medium	Composition	%(vol/vol)/Final conc.	Supplier
Homebrew E8	DMEM/F12	99.85%	Corning, cat# 10-092-CMR
	L-ascorbic acid 2- phosphate trisodium salt	0.64mg/ml	Sigma, cat# 49752
	Heparin sodium salt	100ng/ml	Sigma-Aldrich H3149
	Sodium selenite	14ng/ml	Sigma-Aldrich, cat# S5261
	Recombinant human insulin	20mg/ml	Sigma-Aldrich, cat# 1137649700
	Recombinant human transferrin	5mg/ml	Sigma-Aldrich , #T3705
	Recombinant human FGF2	100ng/ml	Peprotech, cat# 100-18B
	Recombinant human TGFb1	2ng/ml	Peprotech, cat# 100-21
Commercial TeSR E8 medium	TeSR (Basal medium)	96%	Stem Cell Technologies #05991
	TeSR E8 (25X Supplement)	4%	Stem Cell Technologies #05992
DE medium	RMPI 1640	96%	Gibco, #21875034
	B-27TM Supplement (50X)	2%	Gibco, #17504044
	NEAA (100X)	1%	Gibco, #11140050
	Pen/Strep (10,000 U/ml)	1%	Gibco, #15140-122
	Activin A	**100ng/ml**	Peprotech, #120-14P
	Wnt-3a	**50ng/ml**	R&D, #5036-WN
FG medium	RMPI 1640	96%	Gibco, #21875034
	B-27TM Supplement (50X)	2%	Gibco, #17504044
	NEAA (100X)	1%	Gibco, #11140050
	Pen/Strep (10,000 U/ml)	1%	Gibco, #15140-122
	Activin A	**50ng/ml**	Peprotech, #120-14P
HPC medium	RMPI 1640	96%	Gibco, #21875034
	B-27TM Supplement (50X)	2%	Gibco, #17504044
	NEAA (100X)	1%	Gibco, #11140050
	Pen/Strep (10,000 U/ml)	1%	Gibco, #15140-122
	BMP4	**10ng/ml**	R&D, #314-BP
	FGF10	**20ng/ml**	Peprotech, #100-26
	SB431542	**10mM**	Selleckchem, #S1067
PPC	RMPI 1640	96%	Gibco, #21875034
	B-27TM Supplement (50X)	2%	Gibco, #17504044
	NEAA (100X)	1%	Gibco, #11140050
	Pen/Strep (10,000 U/ml)	1%	Gibco, #15140-122
	Retinoic Acid	**3mM**	Sigma-Aldrich, #R2625
	SB431542	**10mM**	Selleckchem, #S1067
	Noggin	**100ng/ml**	Peprotech, #120-10C
	FGF10	**100ng/ml**	Peprotech, #100-26
HG media	RMPI 1640	96%	Gibco, #21875034
	B-27TM Supplement (50X)	2%	Gibco, #17504044
	NEAA (100X)	1%	Gibco, #11140050
	Pen/Strep (10,000 U/ml)	1%	Gibco, #15140-122
	CHIR99021	**3mM**	Tocris, #4423
	Retinoic Acid	**1mM**	Sigma-Aldrich, #R2625
Transfection media for lipid reagents	OptiMem	100%	ThermoFisher #31985070

Table 2-1 Composition of cell culture media.

2.1.4. hiPSC cell maintenance and differentiation

For majority of experiments, R-PAT M cells were maintained in T25 flasks in homebrew essential 8 (HB E8) media. Cells were split in 1:10 ratio when 80% confluency was reached. To split, cells were first washed with 5ml of PBS (Gibco, #14190094) and dissociated using 2.5ml of TrypLE (Gibco, #12604021) for 4min at room temperature (RT). After removal of TrypLE, the flask was gently tapped

to detach cells from the bottom and 5ml of DMEM was used to wash the flask and collect the cells. 500μl of cell suspension was added to a new Matrigel coated T25 flask containing 5ml of HB E8. For the first 24hrs the media was supplemented with 10μM ROCK inhibitor (ROCKi) (Tocris, #1254). Media changes were performed every 24hrs.

For genetic engineering experiments, R-PAT M cells were transitioned to commercial E8 media by gradual increase of the ratio between HB E8 and commercial E8 media (75%:25%; 50%:50%; 25%:75%; 100%:0) every 24hrs.

For differentiation, hiPSCs were seeded into multi-well Matrigel™ coated plates at constant density of 20k/cm^2 and maintained in HB E8 media for 48 hrs with daily media change. Differentiation was started by addition of DE media for 3 days with daily media change. Following DE specification, the cells could either be taken towards HG fate by addition of HG media for 4 days (with daily media change) or FG fate by addition of FG media for 4 days (with daily media change). For differentiation towards HCP or PPC, FG media was replaced with HCP media or PPC media, respectively, for 4 days with daily media change.

2.1.5. Cryopreservation

For cryopreservation cells were dissociated as in the protocol 2.1.4. Detached cells were collected using 5ml of RPMI media and centrifuged at 300 g for 3 min. RMPI media was aspirated and cell pellet was resuspended in 1ml of 10% DMSO (Sigma; #2650) in heat inactivated FBS (Gibco; #10500-064). Cells were quickly but gently resuspended and 250μl of cell suspension was placed in per cryotube. Tube were initially frozen using Mr Frosty™ and moved to liquid nitrogen storage 24-48h later.

2.1.6. Transfection of FG monolayer

Lipofectamine protocol: HiPSCs were seeded in 12 well plates and differentiated towards FG D2, FG D3 and FG D4. 24hrs prior to transfection, cells were fed with antibiotic free differentiation media. Immediately before transfection the media were changed as usual. 2μl, 3μl, 4μl or 5μl of RT Lipofectamine (Invitrogen, #11668030) reagent was diluted in 50μl of OptiMEM media and mixed with 1μg of

GFP plasmid (Lonza, #V4XP-3032) diluted in 50μl of OptiMEM media, mixed gently and incubated for 5min at RT. 100μl of lipid-DNA complex was added per well of 12 well plate at FG D2, FGD3 or FG D2 and FG D4 of differentiation. The plate was gently rocked to mix the transfection components with the cell culture media. Transfection efficiency was checked 24hrs post transfection by fluorescent microscopy and flow cytometry.

Promega protocol: HiPSCs were seeded in 12 well plates and differentiated towards FG D3. 24hrs prior to transfection, cells were fed with antibiotic free differentiation media. Immediately before transfection the media were changed as usual. FuGene HD (Promega, #E2311) was allowed to reach RT. GFP plasmid and FuGene HD reagent were diluted in OptiMEM media at two ratios: 2:1 and 4:1, gently mixed and incubated for 15min at RT. 100μl of FuGene/GFP mixture was added per well of 12 well plate and mixed by gently swirling of the plate. Transfection efficiency was checked by fluorescent microscopy at 24, 48 and 72hrs post transfection.

Biontex protocol: Two different transfection reagents were tested from this company: K4 (Biontex, #T080-1.0) and K2 (Biontex, T060-0.75) using the same protocol. HiPSCs were seeded in 12 well plates and differentiated towards FG D3. 24hrs prior to transfection, cells were fed with antibiotic free differentiation media. On the day of transfection, the media were changed as normal. 2hrs before transfection, a multiplier reagent was added to cell culture media to a total volume of 1% of the media. Transfection reagents were allowed to reach RT and diluted in OptiMEM media. 1μg of GFP plasmid was diluted in OptiMEM media and then diluted reagent and plasmid were mixed together at two ratios: 2:1 and 4:1, reagent (μl) to plasmid (μg). The solutions were gently mixed by pipetting up and down and incubated at RT for 20min. 100μl of the mixed solutions was added per well of cells and gently mixed by swirling the plate. Transfection efficiency was checked by fluorescent microscopy at 24, 48 and 72hrs post transfection.

Mirus protocol: TransIT-X2 (Mirus, #MIR6003) transfection reagent was chosen to be tested. . HiPSCs were seeded in 12 well plates and differentiated towards FG D3. 24hrs prior to transfection, cells were fed with antibiotic free differentiation media. On the day of transfection, the media were changed as normal. Transfection reagents were allowed to reach RT and diluted in OptiMEM media (either 2μl or 4μl per 50μl of the media). 1μg of GFP plasmid was diluted in 50μl of OptiMEM media. Diluted transfection reagent was mixed with the GFP plasmid at two ratios: 2:1 and 4:1 of reagent (μl) to DNA (μg), respectively, gently mixed and incubated for 15min at RT. 100μl of the X2 reagent/GFP mixture was added per well of the 12 well plate and mixed with cell culture media by gentle ricking of the plate. Transfection efficiency was checked by fluorescent microscopy at 24, 48 and 72hrs post transfection.

2.1.7. Nucleofection

R-PAT M cells were nucleofected using Amaxa™ 4D-Nucleofactor and P3 primary cell kit (Lonza; #V4XP-3024). On the day of nucleofection, R-PAT M cells dissociated as in protocol 2.1.4. and resuspended in HBr E8 for cell counting. The required number of cells (100k-320k) were centrifuged at 300g for 3min. Media was aspirated and the cells were resuspended in P3 buffer containing appropriate plasmids [either 1.2μg of CRISPR plasmids (300ng each of guide 2 and guide 3 RNA + 600ng of Cas9 plasmid) and 600ng of pAAV_puro_MsiKD plasmid; or 1μg GFP plasmid + 1.2μg of CRISPR plasmids]. Each nucleofection mixture was placed transferred to Nucleocuvette™ and placed in the 4D-Nucleofactor™ X unit. DN-100 programme was applied to the cuvettes. Post nucleofection cells were placed in the incubator at 37°C for 5min to recover. After incubation cells were seeded onto an appropriate cell culture plate previously coated with Matrigel™ in E8 TeSR media with ROCKi. Transfection efficiency was monitored by fluorescent microscopy at 24h and 48h post nucleofection and by flow cytometry at 48h post nucleofection.

2.1.8. Flow cytometry

For flow cytometry, cells were dissociated with TrypLE as described in the HiPSCs protocol. Dissociated cells were collected into a 15ml falcon tube, centrifuged

at 300g for 3 minutes and the supernatant was removed. Cells were resuspended in 500μl of PBS. For viability testing, Propidium Iodide (PI) dye was added at the final concentration of 2.5μg/ml. Flow cytometry data analysis was performed on Beckman Coulter Kaluza Analysis Software.

2.1.9. Puromycin kill curve

R-PAT M cells were seeded at 20k/cm2 in 12 well plate format in commercial E8 media with ROCKi and allowed to proliferate for 48hrs with daily media changes (commercial E8 only). At 48hrs, varying doses of puromycin were added to wells (0; 0.05μg/ml; 0.1μg/ml; 0.15μg/ml; 0.2μg/ml; 0.3μg/ml and 0.4μg/ml). Puromycin was added to the commercial E8 media with daily media changes for 72hrs. Microscopy images of cells at all concentrations of puromycin were collected before each media change. Optimal puromycin concentration for selection was determined by virtual elimination of all cells after 48hrs.

2.1.10. Manual cell dissection

Following puromycin selection, targeted cells were allowed to recover and from colonies. Once the colonies reached optimal size (~500μm), they were manually dissected under light microscope contained within a cell culture hood using a stem cell cutting tool (Invivogen, #14601). The stem cell colony was first cut into several small pieces, which were scraped off the bottom of the dish using the cutting tool. Once detached, the stem cell cutting tool was used to aspirate the fragments and transfer them to a well of a 24wp containing commercial E8 media with ROCKi. A small fragment was also collected and placed in a PCR strip for direct genomic DNA extraction.

2.1.11. Microscopy

Fixed fluorescence microscopy was performed using Operetta® High content image analysis system (PerkinElmer). Columbus™ analysis software was used to quantify fluorescence intensity. For each experiment, three technical experiments were performed. A technical replicate represents a well of a plate, with 7-10 fields of each well captured and analysed.

2.2. Molecular techniques

2.2.1. RNA extraction and cDNA synthesis

For RNA extraction, RNeasy Mini Kit (Qiagen, #75162) with on-column DNase (Sigma-Aldrich, #DNASE70) digestion were used, following manufacturer's instructions. Briefly, cells were washed once with PBS and lysed with 350µl of RLT buffer. One volume of 70% ethanol was added to the lysed cells and mixed. The reaction was moved onto provided the provided RNA-binding column and centrifuged. Bound RNA was washed with 500µl of RW1 and then DNase solution was applied to the column and left at RT for 15min. Following incubation with DNase, the column was washed with 500µl of RW1 and then once with 700 µl of RPE buffer. The column was dried by 1min centrifugation. RNA was eluted in 30µl of nuclease free water (NFW). Concentration of the eluted RNA was measured using NanoDrop-1000 spectrophotometer. For cDNA synthesis, 500 ng RNA and 0.5 µL random primers (Promega, 430 #C1181) with 1 µL of dNTPs (Promega, #U1511) per reaction were first denatured for 5 minutes at 65^{o}C and snap cooled to prevent re-formation of secondary structures. 4 µL of 1st strand buffer, 0.1M dithiothreitol (DTT), 0.5 µL RNase out and 0.125 µL Superscript II (Invitrogen, 18054071) were added to each reaction and samples were placed in a thermocycler using settings: 10 minutes at 25^{o}C, 50 minutes at 42^{o}C and 15 minutes at 70^{o}C. cDNA was diluted in 600µl of nuclease free water (NFW).

2.2.2. Quantitative real-time PCR

For qPCR, 5µL of cDNA was added to 7.5µL of SensiMix™ SYBR® & Fluorescein Kit (Bioline, #QT525-20). 0.6µL of forward and revers primers each and 1.3µL of NFW. Quantitative PCR was run on 7500 Fast Real-Time PCR system by Applied Biosystems with following settings: 1 cycle 5 minutes at 95^{o}C, 40 cycles of 15 seconds at 95^{o}C, 30 seconds 60^{o}C and 30 seconds at 75^{o}C, melt curve stage 15 seconds at 95^{o}C, 60 seconds at 60^{o}C, 30 seconds 95^{o}C and 15 seconds at 60^{o}C. All samples were run in triplicate. Porphobilinogen deaminase gene was used as an internal reference for all samples. Fold change in gene expression was

calculated using the comparative Ct method. Primers used are listed in Table 2-2.

Gene	Forward primer	Reverse primer
AFP	AAACTATTGGCCTGTGGCGA	TTTTGTCCCTCTTCAGCAAAGC
ALB	CTCGGCTTATTCCAGGGGTG	AAAGGCAATCAACACCAAGGC
CXCR4	CACCGCATCTGGAGAACCA	GCCCATTTCCTCGGTGTAGTT
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
GATA6	GAGCGCTGTTTGTTTAGGGC	GCTGACGTCTAGCTCCTCGG
HHEX	TGCATAAAAGGAAAGGCGGC	TTGCTTTGAGGGTTCTCCTGT
HNF4a	ACTCTCCAAAACCCTCGTCG	CCCTTGGCATCTGGGTCAAA
NANOG	CATGAGTGTGGATCCAGCTTG	CCTGAATAAGCAGATCCATGG
POUF1	AGTGAGAGGCAACCTGGAGA	ACACTCGGACCACATCCTTC
PBGD	GGAGCCATGTCTGGTAACGG	CCACGCGAATCACTCTCATCT
PDX1	GATTGGC GTTGTTTGTGGCT	GCCGGCTTCTCTAAACAGGT
PROX1	ACGTCATCATTCCGAACCCC	TTCCTGCATTGCACTTCCCG
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
SOX2	TGGACAGTTACGCGCACAT	CGAGTAGGACATGCTGTAGGT
TTR	ACCGGTGAATCCAAGTGTCC	GGTTTTCCCAGAGGCAAATGG
BRA(T)	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG
MIXL1	GGTACCCCGACATCCACTTG	TAATCTCCGGCCTAGCCAAA
EOMES	ATCATTACGAAACAGGGCAGGC	CGGGGTTGGTATTTGTGTAAGG
CDX2	GGCAGCCAAGTAAAACCAG	TTCCTCTCCTTTGCTCTGCG
NKX2.1	GCTGCCTAAAACCTGGCGCCG	ATGAAGCGGGAGATGGCGGGGAA
DEANR1	ACATTTGGTAGCCCGTGGAG	TCTTCCCCGGAGAACTAGCA
DIGIT	ACCACTCACGGCAAGCAG	ACGCAGGCAGTCACTGATAA
HULC	ATCTGCAAGCCAGGAAGAGTC	CTTGCTTGATGCTTTGGTCTGT
CARMEN	TAGGTGTTGGCTGAGTGCAG	CCAACCACTCCCCAAACA
CK19	TCCGAACCAAGTTTGAGACG	GCCCCCTCAGCGTACTGATTT
HLXB9	CACCGCGGGCATGAT C	ACT TCCCCAGGAGGT TCG A
HNF1b	GCACCCCTATGAAGACCCAG	GGACTGTCTGGTTGAATTGTCG
SOX9	CTCTGGAGACTTCTGAACGAGAG	CCTTGAAGATGGCGTTGGGG
LGR5	CTCCCAGGTCTGGTGTGTTG	GAGGTCTAGGTAGGAGGTGAAG
GATA2	ACTCCTTCACTCTCAGAGGC	TCGAGGTGATTGAAGAAGAC
GATA5	TCGCCAGCACTGACAGCTCAG	TGGTCTGTTCCAGGCTGTTCC
HEY1	TGGATCACCTGAAAATGCTG	CGAAATCCCAAACTCCGATA
HEY2	AGGCTACTTTGACGCACACG	CAAGTGCTGAGATGAGACACAAG
MSX1	AAACACAAGACGAACCGTAA	GTACATGCTGTAGCCCACAT
MSX2	AGTCGGAAAATTCAGAAGAT	CATGGAGTCTATTGATCTG
TBX2	GGCTTCACCATCCTAAACTCC	AAACGGGTTGTTGTCGATCTT
TBX3	AGTCGGGAAGGCGAATGTTT	AGCGTGATCACTTGGGAAGG
TBX20	AAGGAGGCGACGGAGAACA	TCCTGCCCGACTTGGTGAT
MAF	CTCGTCTTTCCCCAGGACTT	CCTCTTCTGCTTGGCTCTCT
A1AT	ACTTAGCCCCTGTTTGCTCC	CGGCATTGTCGATTCACTGTC
C/EBPa	TATAGGCTGGGCTTCCCCTT	AGCTTTCTGGTGTGACTCGG

Table 2-2 List of qPCR primers.

2.2.3. rRNA depletion and RNA sequencing library preparation

Total RNA extraction from R-PAT M and differentiated samples was performed as per protocol 2.2.1. RNA concentration was measured using Qubit®2.0 Fluorometer (Invitrogen, #Q32857) with the Qubit® RNA BR Assay kits (Thermo-Fisher; #Q10210) and its quality (RIN value) was assessed on 4200 Tapestation System (Agilent technologies; #G2991AA) using RNA ScreenTape Assay kit (#5067-5576). All sequences samples had a RIN value of 10. Ribosomal RNA (rRNA) depletion was performed using NEBNext® rRNA Depletion Kit (NEB; #E6350S) according to manufacturer's instructions. The success of rRNA depletion was evaluated on 4200 Tapestation using High Sensitivity RNA ScreenTape Assay kit (Agilent Technologies; #5067-5579) (RIN ranges 1.0 -3.1). cDNA libraries were prepared using NEBNext® Ultra™ II Directional RNA Library Prep kits for Illumina (NEB; #E7760S) according to manufacturer's instructions. For library multiplexing NEBNext®Multiplex Oligos for Illumina® Index set 1 (NEB; #E7600S) and set 2 (NEB; #7780) were used. Following the library preparation, the concentration and library quality were assessed using the 4200 Tapestation system with High Sensitivity D1000 ScreenTape Assay kit (#5067-5584). Ready libraries were stored at -80°C.

2.2.4. RNA sequencing and bioinformatic data analysis

RNA sequencing and data analysis was outsourced to Babraham Institute, Cambridge, UK. The sequencing was performed on Illumina HiSeq sequencer by Dr Kristina Tabada to yield 30mln paired end reads per sample. Sequencing data quality control and data analysis were performed by Dr Simon Andrews. Reads were mapped using GRCh38_v97 Ensembl human genome.

Differentially expressed genes (DExGs) between BMP and NOG samples were identified using DESeq2 function on SeqMonk Mapped Sequence Analysis tool (Babraham Bioinformatics, Cambridge). Heatmaps were generated using R script written by Dr Simon Andrews on R software. Overrepresentation analysis was performed using WebGestalt (http://www.webgestalt.org/) online tool. GSEA analysis of DExGs was performed using GSEA v4.3.2 software, GSEA graphs were created using the same software (Figure 2-1). Gene sets were downloaded

from Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp).

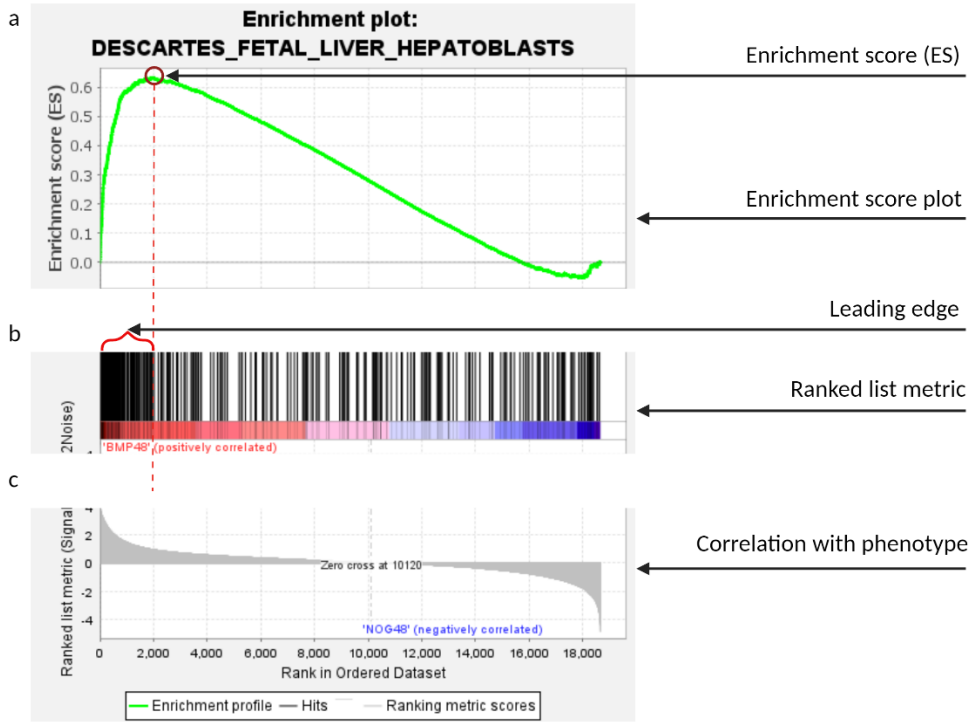

Figure 2-1 Features of the GSEA plot.
DExGs were ranked from the highest to the lowest fold change value. GSEA analysis was done to identify gene sets of interest. a) Enrichment score plot: each gene from a gene set is checked against the ranked list of DExGs from our experiment (BMPvsNOG) and an enrichment plot is created. ES (enrichment score) is the maximum enrichment score reached by a gene set. It represents how much a gene set is overrepresented at the top (upregulated genes) and bottom (downregulated genes) of the ranked gene set; b) Ranked gene metric: each bar represents a gene from the gene set. It visualizes where on the ranked list of DExGs each individual gene from the gene set appears; Leading edge subset: appears before the ES for positive enrichment scores and after the ES for negative enrichment scores. It shows which genes contribute most to the gene set being upregulated/downregulated. c) Correlation with phenotype: the plot is created from the values of the ranking metric of the genes in the tested gene set. The value goes from positive to negative and, in our case, positive value indicates correlation with the first phenotype (BMP48) and negative value indicated correlation with the second phenotype (NOG48).

2.2.5. Chromatin immunoprecipitation (ChIP)

For ChIP experiments, R-PAT M cells were seeded onto 100mm x 15mm round cell culture dishes to yield appropriate amount of cell per condition with easy access for cell scraping (Thermosifher Scientific, #150464). R-PAT M cells were differentiated to D1 and D2 LPCs according to the protocol 2.1.4 in the presence

or absence of BMP4 signalling. For ChIP experiments, BMP4 signalling was blocked with 2 dorsomorphin. ChIP was performed using Pierce™ Magnetic ChIP (Thermofisher Scientific, #26157). Cells were crosslinked using 10ml of 1% formaldehyde (FA; Merck, #F8775) and incubated for 10min at RT. FA was neutralised with 1ml of x10 glycine solution and incubated for 5min at RT. FA/glycine solution was removed from the dish and the cells were washed twice with 10ml ice-cold PBS. 10µl of Halt Cocktail were added to 1ml of ice-cold PBS applied to the dish. Cells were detached from the dish using a scraper (Fisher scientific; # 08-100-241), collected into 1.5ml Eppendorf tube and centrifuged for 5min at 3000g. PBS was removed and cell pellet taken for further procedure. 200µl of Membrane Extraction Buffer was added to the cell pellet and it was mixed by pipetting the pellet up and down and vortexed for 15s. Cells were incubation with the solution on ice for 15min and then spun at 9000g for 3min. Supernatant was removed and the nuclei were resuspended in 200µl of MNase Digestion buffer working solution. 2µl of diluted MNase (dilution factor determined by previous optimisation) were added to the nuclei, vortexed and incubated in the water bath at 37°C for 15 min with mixing by inversion every 5min. To stop MNase digestion, 20µl of MNase Stop Solution were added, vortexed and left on ice for 5min. Nuclei were retrieved by 5min centrifugation at 9000g and removal of supernatant. Nuclei were resuspended in 100µl of 1x IP Dilution buffer containing protease/phosphatase inhibitors. The tubes were sonicated using Diagenode Bioruptor. Three 30s pulses followed by 30s breaks at medium setting were applied. Samples were centrifuged at 9000g for 5min, and supernatants were transferred to a new 1.5ml Eppendorf for immunoprecipitation. 10µl of supernatant for each time point and condition was takes and stored at -20°C to serve as 10% input. The remaining 90µl of the supernatant was diluted in 410µl of 1x dilution buffer. 5µl of ChIP grade anti-SMAD1/5/8 antibody (Cell signalling; #11971) and 5µl of Normal Rabbit IgG (Cell signalling; #2279) was added to test and control samples, respectively. Chromatin and antibodies were incubated overnight at 4°C with constant mixing. 20µl of A/G Magnetic Beads was added to each sample and incubated at 4°C with mixing for 2h. The beads were collected using a magnetic rack and

the supernatant was carefully collected and discarded. 1ml of IP wash buffer was added to the beads and incubated for 5min at RT while mixing. The wash procedure was completed a total of three times. To elute DNA, 150µl of 1x IP dilution buffer was added to the beads and the samples were incubated at 65°C for 30min with vigorous shaking. Proteinase K digestion solution was prepared using 6µl of 5M NaCl, 2µl of 20mg/ml proteinase. Following the 65°C incubation, beads were separated from the eluted DNA using magnetic rack, and the supernatant was placed in the tubes containing Proteinase K digestion solution. Input samples were prepared by thawing on ice and addition of 150µl of 1x IP Elution buffer and proteinase K digestion solution. All samples were vortexed and placed at 65°C for 1.5h. To recover the DNA, 750µl of DNA binding buffer was added to each sample, mixed and placed in the DNA clean up column inserted into a 2ml collection tube. The columns were centrifuged at 10 000g for 1min then washed with 750µl of DNA Column wash buffer. Once the wash buffer was removed by centrifugation at 10 000g for 1 min, the tubes were dried by another centrifugation at 10 000g for 2min. All columns were placed in a fresh collection tube and 50µl of DNA column elution buffer was applied directly to the column membrane. The columns were centrifuged at 10 000g for 1min. The eluted DNA was stored at -80°C until ChIP library preparation.

2.2.6. ChIP library preparation, sequencing and data analysis

ChIP library preparation was outsourced to DeepSeq Department at The University of Nottingham. Libraries preparation and sequencing was performed by Nadine Holmes. Data analysis was outsourced to Babraham Institute, Cambridge and performed by Dr Simon Andrews.

2.2.7. Direct genomic DNA extraction

DNA for PCR screening of targeted R-PAT M cells was extracted using Phire Tissue Direct PCR Master Mix (ThermoFisher, #F170S). Small colony fragment was placed in 0.6ml Eppendorf tube with 10µl of media. 20µl of dilution buffer and 0.5µl of DNA release mix were added to the tube and vortexed for 30s. After short spin, the tube was left at RT for 5min. Next, the tube was placed in a heat block at 98^{o}C for 3min and then spun and stored at -20^{o}C.

2.2.8. PCR Genotyping

Following puromycin selection, surviving colonies were dissected and moved to 24well plate (2.1.10). Small fragment of each colony was saved for direct gDNA extraction (2.2.7). PCR genotyping was done using Phire master mix (MM) provided with Phire Tissue Direct PCR Master Mix. PCR reaction was set up as follows: 2µl of gDNA sample with 0.5µl of each primer, 5µl of MM and 2µl of NFW . Primer sequences and expected band length for each genotyping are in Table 2-3. PCR programme for each genotyping is stated in Table 2-4.

				Expected band length (bp)		
PCR type	Primer name	Primer sequence	Annealing temp.	Wild-type	Correctly targeted	Plasmid integration
Locus	Locus Locus_fw	CTGTTTCCCCTTCCC AGGCAGGTCC	65°C	1692	No band	No band
	Locus Locus_rev	TGCAGGGGAACGG GGCTCAGTCTGA				
5'INT	5'INT Locus_fw	CTGTTTCCCCTTCCC AGGCAGGTCC	65.1°C	No band	991	No band
	5'INT Locus_rev	TCGTCGCGGGTGGC GAGGCGCACCG				
3'INT	3'INT OPTTetR_fw	CCACCGAGAAGCAG TACGAG	69.4°C	No band	1447	No band
	3'INT OPTTetR_fw	TGCAGGGGAACGG GGCTCAGTCTGA				
5'BB	5'BB Backbone_fw	ATGCTTCCGGCTCGT ATGTT	60°C	No band	No band	1227
	5'BB Puro_rev	TGAGGAAGAGTTCT TGCAGCTC				
3'BB	3'BB OPTTetR_fw	CCACCGAGAAGCAG TACGAG	60°C	No band	No band	1802
	3'BB Backbone_rev	ATGCACCACCGGGT AAAGTT				
CTRL	SOX21	AGCCCTTGGGGAST TGAATTGCTG	72.6°C	237		
	SOX21	GCACTCCAGAGGAC AGCRGTGTCAATA-				

Table 2-3 Primer sequences for OPTiKD lines genotyping, annealing temperatures for primers and expected gel electrophoresis band lengths.

PCR programmes for Phire polymerase			
Stage	Temp	Time	
Initial denaturation	98°C	5min	
Denature template	98°C	10s	35 cycles
Anneal primers	*	30s	
Extension	72°C	2min	
Final extension	72°C	30s	
Hold	10°C	∞	

Table 2-4 PCR programme for genotyping reactions with Phire polymerase MM.
*annealing temperature stated in Table 2-3.

2.2.9. Gel electrophoresis

Agarose gel electrophoresis was preformed following genotyping, colony PCR, PCR amplification or plasmids digestions. DNA products were run on 1-3% agarose gels containing 10mg/ml ethidium bromide (Invitrogen; #15585011) at 80V for 45min to 1.5h. DNA samples were either loaded with 6X loading dye (NEB#B57051S) diluted to 1X (digested plasmids and PCR products not amplified with Phire polymerase MM) or loaded directly into a well (Phire MM PCR products). Gels were visualised using LAS-4000 Fujifilm Luminescent Image Analyser.

2.2.10. Immunostaining

For immunostaining, cells were grown and differentiated on 48 well plates. Cells were fixed with 4% PFA for 20min at 4^{o}C followed by two PBS washes. For permeabilization and blocking cells were treated with 0.1% TritonX in 10% donkey serum for 30min. Following that, primary antibody was added for overnight incubation at 4^{o}C. All antibodies were diluted in 1% donkey serum in PBS, this solution was also used as a wash solution. Excess 1^{o} antibody was removed with three 5min washes. 2^{o} antibody was applied for 1 hour incubation at room temperature (RT) and removed with three 5min washes. For 2^{o} antibody incubation, plates were covered to protect from light. For double staining, the procedure was repeated from application of the 1^{o} antibody. DAPI (Sigma, #D9542) staining was applied during the second wash following 2^{o} antibody

incubation. Details of antibodies used are provided in Table 2-5. Cells were imaged using Operetta High Content Imagining System.

2.2.11. Building of the pAAV_puro_MsiKD plasmids

2.2.11.1. Plasmid digestion

pAAV_puro_siKD diagnostic digest:

Diagnostic plasmid digestion was performed using EcoRI and PsiI restriction enzymes. 1µg of the pAAV_puro_siKD plasmid was digested with 1µl of each enzyme in 5µl of CutSmart buffer and 37µl of NFW. The reaction was incubated at 37°C for 1h.

Name	Dilution	Host species
NANOG	1:50	Goat
OCT4	1:50	Goat
SOX2	1:50	Mouse
SOX17	1:100	Goat
FOXA2	1:100	Goat
CDX2	1:50	Mouse
GATA4	1:50	Mouse
BRA(T)CHYURY	1:50	Mouse
Alexa Fluor 488 anti-goat	1:400	Donkey
Alexa Fluor 647 anti-rabbit	1:400	Donkey
Alexa Fluor 488 anti-mouse	1:400	Donkey
Alexa Fluor 647 anti-mouse	1:400	Donkey
DAPI	1:500	N/A

Table 2-5 Details of antibodies used for immunocytochemistry and ChIP.

pAAV puro siKD plasmid preparation for ligation with shRNA inserts: 5µg of the plasmid was digested with 3µl of BglII (NEB; #R0144) and 3µl of SalI (NEB; # R3138S) restriction enzymes in the presence of 3µl of Shrimp Alkaline Phosphatase (rSAP; NEB; #M0371S) for dephosphorylation. 9µl of NEB3.1 buffer and NFW up to 90µl was added. The mixture was incubated at 37°C for 2hrs.

pAAV puro siKD plasmid preparation for Gibson assembly: 5µg of the plasmid was digested with 5µl of BstBI (NEB; #R0519S) and 5µl of HincII (NEB; #R0103S) restriction enzymes in 10µl of Cutsmart buffer (NEB) and NFW up to 100µl. The mixture was incubated for 1h and 30min at 37°C.

2.2.11.2. Design of shRNA oligos

To maximise the chances of high levels of knockdown of each candidate TF, PubMed database was searched for published sequences of shRNA/siRNAs shown to efficiently knockdown our selected genes. Where previously validated shRNA sequence could not be found, Broad Institute's The RNAi Consortium (TRC) shRNA library has been used to select shRNA sequence with high predictive values for induction of knockdown (Table 2-6).

Candidate TF	siRNA/shRNA	Source
TBX2	GGAGCUGUGGGACCAGUUC	(Crawford, McIntyre et al. 2019)
TBX3	AUGCCAAAGAGGAUGUACAUUCA	(Dong, Dong et al. 2018)
TBX20	CTGGATCAACATGGCCATATA	Broad Institute TRC library
HEY1 (A)	UAGAGCCGAACUCAAGUUUCCAUUC	(Brun, Jain et al. 2018)
HEY1 (B)	UUGAGAUGCGAAACCAGUCGAACUC	(Brun, Jain et al. 2018)
HEY2	CUCAGAUUAUGGCAAGAAA	(Wang, Zhu et al. 2019)
MAF (A)	AGAGGGACGCGUACAAGGA	(Bianchi, Bulgarelli et al. 2015)
MAF (B)	CUGGAAGACUACUACUGGA	(Bianchi, Bulgarelli et al. 2015)
GATA2 (B)	GCAUGAAGAUGGAAAGUGG	(Guo, Fu et al. 2016)
GATA2 (E)	UUCUUGGACUUGUUGGACAUCUUCC	(Kanki, Kohro et al. 2011)
GATA5	AAAGUCCUCAGGCUCGAAC	(Feng, Zhu et al. 2019)
MSX1	GCAUUUAGAUCUACACUCU	(Goto, Fujimoto et al. 2016)
MSX2	GCAGGCAGCGUCCAUAUAU	(Nallasamy, Kaya Okur et al. 2019)
Scramble	CCUAAGGUUAAGUCGCCCUCGCUC	https://www.addgene.org/1864/

Table 2-6 Sequence of si/shRNAs for knock down of candidate TF genes.

Once the siRNA/shRNA sequences have been identified for each gene, BLOCK-IT RNAi designer software tool was used to obtain a full shRNA sequence. As per

(Bertero et al., 2016) protocol, additional base pairs were added to each end of the shRNAs sequence to create sticky ends compatible with BglII and SalI restriction enzymes (Figure 2-2). Top and bottom strands were synthetised by Sigma and annealed, phosphorylated and purified creating shRNAs ds oligos ready for cloning into the empty pAAV_puro_siKD vector.

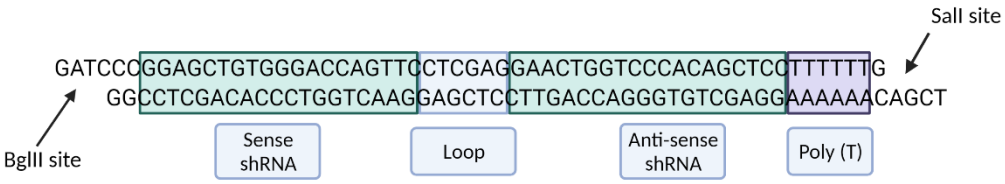

Figure 2-2 Design strategy of the shRNA ds oligos.

2.2.11.3. Annealing of shRNA oligos

5µl of 200µM top oligo and 5µl of 200µM of bottom oligo were added to 1µl of T4 polynucleotide kinase (NEB; #M0201S) and 2µl of T4 DNA ligase reaction buffer (NEB; M0202S), topped up with NFW to 20µl. Phosphorylation, denaturing and annealing was done in a thermocycle according to the following programme (lid at 95°C):

- 37°C for 1h
- 95°C for 5min
- Ramp down to 80°C at 0.1°C/s
- 80°C for 4min
- Ramp down to 75°C at 0.1°C/s
- 75°C for 4min
- Ramp down to 70°C at 0.1°C/s
- 70°C for 4min
- Ramp down to 10°C at 0.1°C/s
- 10°C hold;

Annealed oligos were stored at -20°C for up to two weeks.

2.2.11.4. Ligation of cut vector with annealed shRNA oligos

Annealed oligos were first diluted with NFW at a ratio of 1:500. 4µl of diluted oligos was added to 50ng of cut vector and ligated with the use of 1µl of T4 DNA ligase in 2µl of T4DNA ligase buffer and NFW topped up to 10µl. The mixture was left at RT for 1h. The ligated plasmid was transformed according to protocol 2.2.11.5.

2.2.11.5. Bacterial transformation

For plasmid transformation, NEB® 5-alpha Competent E.coli (High efficiency) kit was used (NEB; #C2987H). Bacteria were removed from -80°C storage and thawed on ice for 10min. 50ng of plasmid was added to the 50µl of bacteria and mixed by gently flicking the tube. The tube was placed on ice for 30min. Next, the tube was placed in water bath at 42°C for 30s and placed straight back on ice for 2min. 250µl of SOC media was added and the tube was placed in shaking incubator at 37°C, 225rpm for 1 hour. Following incubation, 100µl of bacteria in SOC was streaked a previously prepared LB agar plate with 100µg/ml of Ampicillin. The plate was placed at 37°C for 14-16h.

2.2.11.6. Bacterial colony PCR to identify correctly ligated plasmids

Single colonies were picked from the agar plates cultured overnight. Each colony was diluted in 10µl of NFW. 5µl were placed in 200µl LB broth for future expansion and 5µl were taken for PCR. PCR reaction contained 5µl of the diluted colony, 0.5µl of forward and reverse primers each, 0.5µl of dNTP (NEB; #N0447S), 2µl of Taq reaction buffer and 0.125 Taq polymerase (NEB; M0273S), topped up with NFW to 20µl. PCR was run on thermocycler according to the following programme:

Bacterial colony PCR			
Stage	**Temp**	**Time**	
Initial denaturation	95°C	5min	
Denature template	95°C	30s	35 cycles
Anneal primers	60°C	30s	
Extension	72°C	1min	
Final extension	72°C	1min	

Hold	10°C	∞

4µl of loading dye (NEB; #B7025) was added to the samples following the PCR reaction. Samples were run on 1.2% agarose gel for 1h at 80V. Following identification of plasmids with correct band length, 3 candidates for each pAAV_puro_siKD construct were amplified (), purified (2.2.11.10) and sent for Sanger Sequencing with Source Bioscience to verify correct sequence of each insertion.

2.2.11.7. PCR amplification

To produce shRNA cassettes for pAAV_puro_MsiKD plasmid building, pAAV_purp_siKD plasmids with appropriate shRNAs were PCR amplified with the use of primers containing regions of homology for Gibson assembly (supplementary figure). 4ng of plasmid was amplified using 10µl of forward and reverse primers each (5µM), 1µl of dNTPs (NEB; #N0447S), 1µl of Phusion polymerase (NEB; #M0530S) in 10µl of high fidelity Phusion buffer. The reaction was made up with NFW up to 50µl. 4 PCR reaction were made up for each amplified cassette. The thermocycler run on the following settings:

shRNA cassettes amplification			
Stage	**Temp**	**Time**	
Initial denaturation	98°C	1min	
Denature template	98°C	10s	35 cycles
Anneal primers	65°C	30s	
Extension	72°C	30min	
Final extension	72°C	30s	
Hold	10°C	∞	

Following PCR, the 4 reactions were pooled and amplified DNA was purified according to 2.2.11.12. The DNA concentration was measured using NanoDrop-1000 and 2µg of each cassette were run on 1% agarose gel for 1h at 80V. Bands in the region of 400bp were excised and purified following 2.2.11.11 protocol. DNA concentration following gel extraction was measure using NanoDrop-1000. Fragments were stored at -20°C.

2.2.11.8. Gibson assembly

Gibson assembly was performed using Gibson Assembly® Cloning Kit (NEB, #E5510). 150ng of cut pAAV_puro_siKD vector digested with appropriate restriction enzymes, 35ng of each purified DNA insert, 10µl of master mix and nuclease free water up to 20µl were added to 0.6ml Eppendorf tube and mixed. The tube was incubated for 1h at 50°C in a thermocycler. Following the 1h incubation, 2µl of the assembly mixture was transformed in DH5α as described in 2.2.11.5.

2.2.11.9. Plasmid amplification

To amplify a plasmid following bacterial transformation, a fragment of bacterial colony or 5µl of bacterial suspension was placed in 5ml of LB broth with 100µg/ml Ampicillin and incubated at 14-16h in shaking incubator at 37°C.

2.2.11.10. Plasmid purification

All plasmids were purified using PureYield™ Plasmid Miniprep System (Promega, #A1222). 3ml to 5ml of overnight liquid broth culture were spun for 30s at 16000g. The supernatant was collected and discarded. The pellet was resuspended using a pipette in 600µl of TE buffer. 100ml of cell lysis was added and mixed with the cells by inverting 6 times. 350ul of cold Neutralization Solution was added and mixed by inverting the tube then spun for 3min at 16000g. The supernatant was transferred to a PureYield™ column and centrifuged at 16000g for 15s. The column was washed twice: first with 200µl of Endotoxin Removal Wash and then with 400µl of Column Wash Solution. The washed were moved through the column by centrifugation at 16000g. For elution of the plasmid, the column was placed in a clean collection tube and 30µl of nuclease free water (NFW) was applied directly to the column membrane. After 4 min incubation at RT, the column was spun for 1min at 16000g. The eluted plasmid was checked for concentration and purity on NanoDrop-1000 and stored at -20°C.

2.2.11.11. DNA purification from agarose gel

DNA fragments from PCR amplification were purified using QIAquick PCR and Gel Cleanup kit (Qiagen; #28506). PE buffer was prepared on receiving the kit following manufacturer's instructions. DNA band of appropriate length was cut out from the gel using a scalpel, weighted and placed in 1.5ml Eppendorf. PB buffer was added to the tube in a ratio 3:1 of buffer to gel volume. The tube was incubated for 10min at 50°C, vortexing occasionally to dissolve the gel completely. 1 gel volume of isopropanol was added to the tube and mixed. The sample was applied to the QIAquick column and centrifuged for 1min at 16000g. The column was washed with 500µl of QG buffer and centrifuged for 1min at 16000g. Two 750µl washes with Buffer PE followed. The column was dried by centrifuging at 16000g for 1min. 30µl of NFW was applied to the column and incubated for 4min at RT to maximise the yield of purified DNA. The column was spun, and DNA concentration was measured using NanoDrop-1000. PCR fragments were stored at -20°C.

2.2.11.12. PCR product purification

DNA purification directly following PCR amplification was performed using QIAquick PCR and Gel Cleanup kit (Qiagen; #28506). PE buffer was prepared on receiving the kit following manufacturer's instructions. PB buffer was added to the tube in a ratio 5:1 of buffer to PCR reaction volume and mixed. The sample was applied to the QIAquick column and centrifuged for 1min at 16000g. The column membrane was washed twice with 750µl of PE buffer. The column was dried by centrifuging at 16000g for 1min. 30µl of NFW was applied to the column and incubated for 4min at RT to maximise the yield of purified DNA. The column was spun, and DNA concentration was measured using NanoDrop-1000. PCR fragments were stored at -20°C.

2.2.11.13. Sanger sequencing

For sample sequencing samples were sent to Source Bio Science, Nottingham. Sequencing results were visualised and compared using SnapGene Viewer 6.0.5.

2.3. Statistical analysis

All statistical analysis, excluding RNA sequencing, was performed using GraphPad Prism. Data is presented as mean ± standard deviation (SD), determined based on three technical replicates for qPCR. Flow cytometry data was based on biological replicates. Statistical significance was determined using the most relevant statistical test, and indicated as follows: ns ≥ 0.05; *p ≤ 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. Applied statistical tests are stated in the figure legends.

Chapter 3. Characterization of hiPSCs differentiation protocols for the use in modelling of early human endodermal development

3.1. Introduction

3.1.1. Pluripotent stem cells as a model of human development

The exact molecular mechanisms governing each stage of hepatogenesis are still not fully understood. Animal models, such as Xenopus, mice, rats or chicken helped to identify the major genes and signalling pathways governing liver development, however not all off the findings of animal research translate to humans (Lal et al., 2016, Odom et al., 2007). Since their derivation, PSC have been applied for modelling of human development. They provide a species relevant model and circumvent the ethical issues connected to the maintenance of human embryos in in *vitro* culture. PSC can be differentiated into virtually any cell type of the body by addition of specific developmental signals at carefully controlled time intervals. Hannan group has established protocols for PSC differentiation into various lineages of endodermal origin (Hannan et al., 2013a, Hannan et al., 2013b, Hannan et al., 2015, Cho et al., 2012, Sampaziotis et al., 2015). Each protocol attempts to simulate events occurring during development, as described in the introduction (1.2) and induces PSC to go through progenitor stages to terminally differentiated cells by addition of small molecules or growth factors at specified times (Figure 3-1). The first stage of each protocol is directing PSC towards definitive endoderm, the earliest progenitor of endodermal organs. In our protocol, Activin A and Wnt3a are added to cell media to induce PSC to differentiate towards DE. Activin A is a member of TGFβ family that mimics the action of Nodal. It is sufficient to induce the expression of endodermal genes, such as SOX17 or GSC, but not to suppress pluripotency factors, such as NANOG or OCT4 (Touboul et al., 2010). Addition of Wnt3a during the DE specification has been shown to improve the level of expression of endodermal genes and suppression of pluripotency genes (Hay et

al., 2008). Once DE identity is established, as assessed by the expression of endoderm specific genes, DE can be further patterned along the anterior-posterior axis to form foregut, midgut or hindgut depending on which signalling pathways are activated. ActA induces expression of foregut gene HHEX, while supressing CDX2. CHIR99021, a GSK3β inhibitor and hence Wnt signalling activator, induces hindgut fate in DE cells, as shown by expression of CDX2 (Hannan et al., 2013a). After 4 days of signalling with either ActA or CHIR99021, DE cells will form foregut or hindgut monolayer, respectively. HG cells can further be differentiated to gut organoids, while foregut cells are precursors of thyroid, lung, liver or pancreatic cell types.

To induce FG cells towards pancreatic fate, we add Retinoic Acid (RA) which drives the specification towards pancreatic fate. BMP inhibitor Noggin is added, as BMP signalling blocks pancreatic specification and drives FG cells towards hepatic lineage. Additionally, ActA signalling is blocked using SB431542 as it also directs FG cells to alternative fates and blocks the expression of pancreas specific genes. FGF10 signalling blocks the expression of gut marker, CDX2 and maintains proliferation of the differentiating cells (Cho et al., 2012).

Induction of liver progenitor fate from FG cells is achieved by signalling with BMP4 and FGF10, two factors necessary for the formation of the liver bud (Rossi et al., 2001, Shin et al., 2007). Additionally, in our protocol we inhibit ActA signalling with SB431542, as it has been shown to improve the expression of hepatic markers in liver progenitor cells (Touboul et al., 2010). Once liver progenitor cells (LPCs) are established, they can be matured to either hepatocyte like cells or biliary epithelial like cells by the use of appropriate growth factors and small molecules (Hannan et al., 2013b, Sampaziotis et al., 2015).

There are numerous protocols for differentiation of iPSCs into cells of endodermal origin. They can differ in numerous aspects, such as the use of different growth factors or small molecules, basal media or the matrix on which the cells are cultured. The timeframes for each cell type can also differ between the protocols. As currently there are no internationally agreed standards on

what the best route of differentiation towards each lineage is, it is important to show that each protocol used renders cells with correct gene and protein expression profile, as determined with the help of previous research. Therefore, our project starts with the characterization of our differentiation platform showing that the protocols we use produce cells of the desired type.

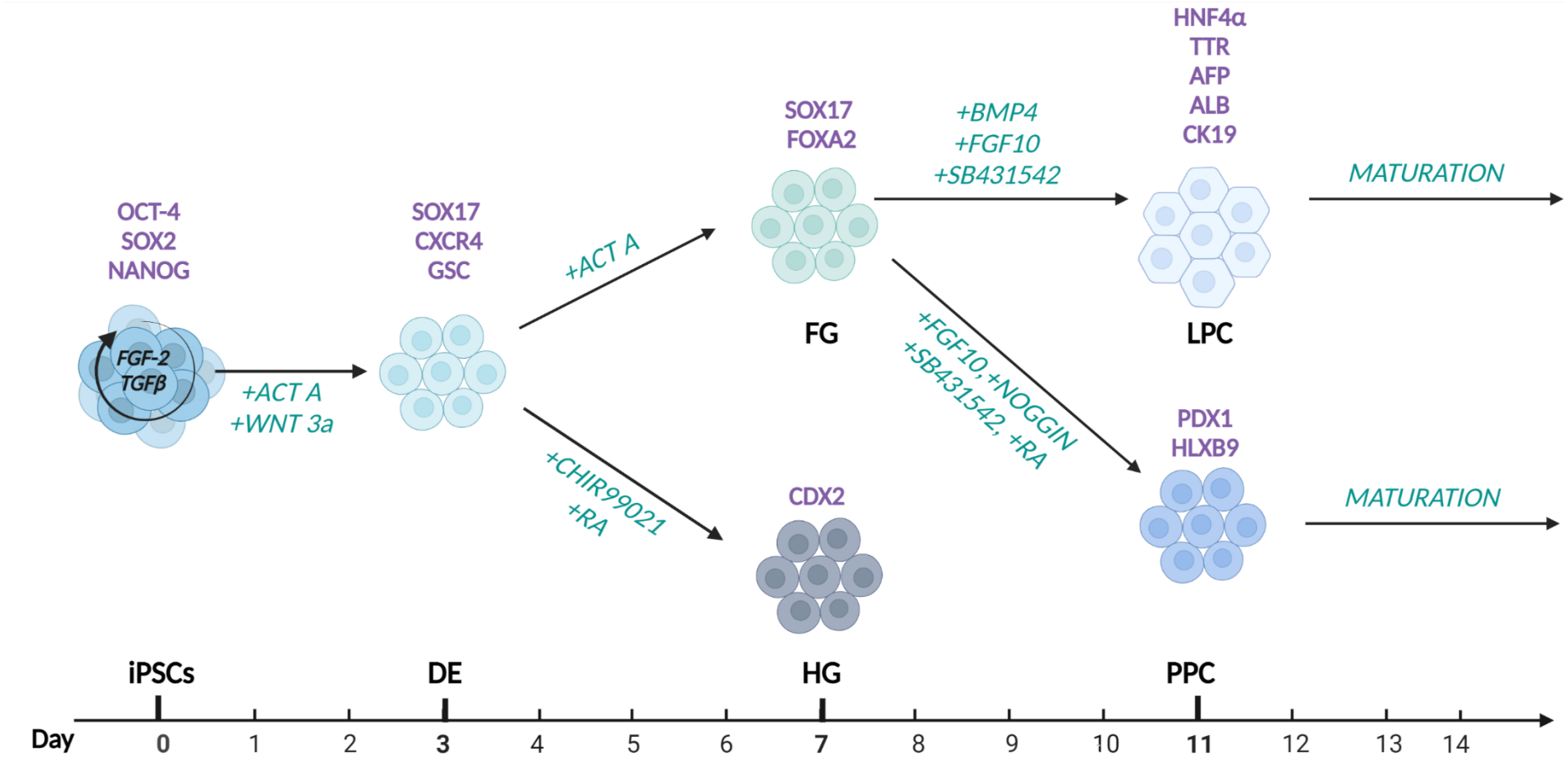

Figure 3-1 The overview of protocols for human iPSCs differentiation towards endodermal lineages used in this thesis.

3.1.2. Transfection methods for nucleic acid delivery into foregut cells

The objective of this thesis is to study the genetic regulators of early endodermal fate choices, in particular the transcription factors (TFs) responsible for mediation of BMP4 signalling during LPC specification. The most common way to study a gene's function is its knockdown (e.g.: siRNA), knockout (e.g.: using CRISPR/Cas9) or overexpression followed by analysis of the consequences of the genetic modification. Those studies require the introduction of nucleic acids or proteins into the cells, a process called transfection. First transfection method was described in the 1960s and since then numerous new approaches have been developed (Chong et al., 2021). The most suitable transfection method is determined by the purpose of the experiment. Stable transfection, when the genetic material integrates into the host genome, is used for the creation of stable cell lines, large scale protein production or gene therapy. Transient transfection, when the delivered nucleic acids are temporarily inducing changes in the cells, can be used to study gene's function in small scale experiments e.g.: gene knockdown with siRNA (Dong et al., 2018).

Whether the transfection is to be transient or stable, the nucleic acid material has to be delivered into the cells. There are several methods available for DNA/RNA delivery into the cells, briefly summarized in Table 3-1. Ideally, the selected method for transfection should be highly efficient with minimal toxicity to the cells. However, other factors must also be considered when choosing the method, such as cost, available equipment, operator's skill and safety. Once the method is selected, it requires a degree of optimization to achieve optimal transfection efficiency.

Method		Description	Advantages	Disadvantages
Physical	Electroporation	Application of electrical current to briefly increase the permeability of the cell membrane and allow the entry of nucleic acids	High efficiency, even in difficult to transfect cells; An easy procedure;	High toxicity; Requires specialist equipment;
	Sonoporation	Modification of cell's membrane permeability by the use of ultrasound waves;	An easy procedure;	High toxicity; requires specialist equipment;
	Gene microinjection	Manual delivery of the genetic material by the use of a fine needle;	High delivery accuracy and efficiency; Control of the amount of the delivered material; Low toxicity;	Highly skilled operator required; Laborious; Small scale experiments;
	Optical transfection	Creation of small, transient holes in the cell membrane using light (e.g.: laser beam);	Can be applied to single cell transfection; high efficiency	Toxicity; requirement of specialist equipment;
Chemical	Cationic lipids	Positively charged lipid aggregates that encapsulate the negatively charged nucleic acids and mediate the entry into the cells by merging with the phospholipid bilayer of the cell.	High efficiency; Inexpensive; an easy procedure;	Efficiency dependent of cell type and culture conditions; Toxicity;
	Cationic polymers (e.g.: DEAE-dextran, polylysine)	Cationic polymers form nucleic acid-polymer complexes which can be taken up by the cell by endocytosis;	Inexpensive; An easy procedure;	Variable transfection efficiency and toxicity;
	Calcium phosphate	Positively charged calcium cations bind with negatively charged nucleic acids forming a precipitate that can be taken up by the cell;	Low coast; good efficiency; applicable to w wide variety of cells; An easy procedure;	Efficiency can be easily affected by e.g.: change in the pH;
	Magnetic beads;	Complexes of iron oxide and nucleic acids are introduced into the cells using magnetic field.	High efficiency; An easy procedure;	For adherent cells only;
Viral	Adenoviruses	A DNA virus that enters the cells by binding to CAR receptor. It is then internalized via integrin-mediated endocytosis and transported to the nucleus.	Highly efficient; non-integrating;	Can induce strong immune response in host cells; higher handling risk;
	Adeno-associated viruses	A DNA virus that enters the cells via receptor-mediated endocytosis and is transported into the nucleus.	Lower risk of immune response in host cells; lower risk of insertional mutagenesis;	Limited packing capacity up to 4.9kb;
	Retroviruses	RNA virus that integrates its RNA into the host DNA using reverse transcriptase.	Stable transfection; lower probability to trigger inflammation;	Possibility for insertional mutagenesis and gene disruption;

Table 3-1 A brief summary of available transfection methods (Chong et al., 2021, Fus-Kujawa et al., 2021)

3.2. Chapter aims and objectives

In this chapter, we aim to characterize our differentiation platform and demonstrate that it is suitable for modelling of early stages of human liver development while optimizing a transfection method for nucleic acid delivery into cells with the following objectives:

- Human iPSCs can be applied to modelling of human development
- Foregut monolayer and hiPSCs can be efficiently transfected using lipid and mechanical methods

3.3. Results

3.3.1. Characterization of iPSC differentiation to endodermal lineages for study of early human development

R-PAT M hiPSC cell line was differentiated to definitive endoderm (DE) and foregut cells following the protocol specified in Figure 3-2a. The morphological changes happening during the differentiation to FG are presented in Figure 3-2b. Upon induction of differentiation to DE cells begin to proliferate and undergo morphological changes consistent with epithelial to mesenchymal transition (EMT). R-PAT M cells form loose colonies and have high nucleus to cytoplasm ratio. By DE D3 the cells form a densely packed monolayer and increase in size. Cells that fail to differentiate die off. Establishment of a homogenous layer of DE cells allows for anteroposterior patterning depended on the delivery of specific growth factors. ActA induces FG fate which results in further proliferation and modification of cell morphology to a rhomboidal shape. Morphological changes were accompanied by changes in gene expression as analysed by qPCR. Pluripotency genes gradually decrease as the differentiation progresses, while mesendoderm genes: BRA(T) and MIXL1, transiently increase, peaking at DE D2. (Figure 3-3a). This indicates that the cells progress through the primitive streak stage. As the differentiation continues, endodermal genes are upregulated with peak SOX17 and FOXA2 levels at DE D3. There is also upregulation of CXCR4, GATA4 and EOMES (Figure 3-3b). Once DE monolayer is established at DE D3, cells are further induced towards anterior foregut fate by signalling with Act A only. High levels of SOX17, FOXA2 and CXCR4

are maintained though the FG stage of differentiation. The specificity of FG differentiation is confirmed by HHEX expression and lack of midgut- and hindgut-specific markers: PDX1 and CDX2, respectively (Figure 3-4a). At FG D4 there is also no expression of liver or lung specific genes indicating that the FG cells are not primed towards any specific cell lineage (Figure 3-4b). We were also able to detect increase in the levels of DE specific lncRNAs: DEANAR and DIGIT, while the levels of mesoderm specific lncRNA CARMEN or liver/pancreatic cancer associated lncRNA HULC remained virtually undetectable (Figure 3-5).

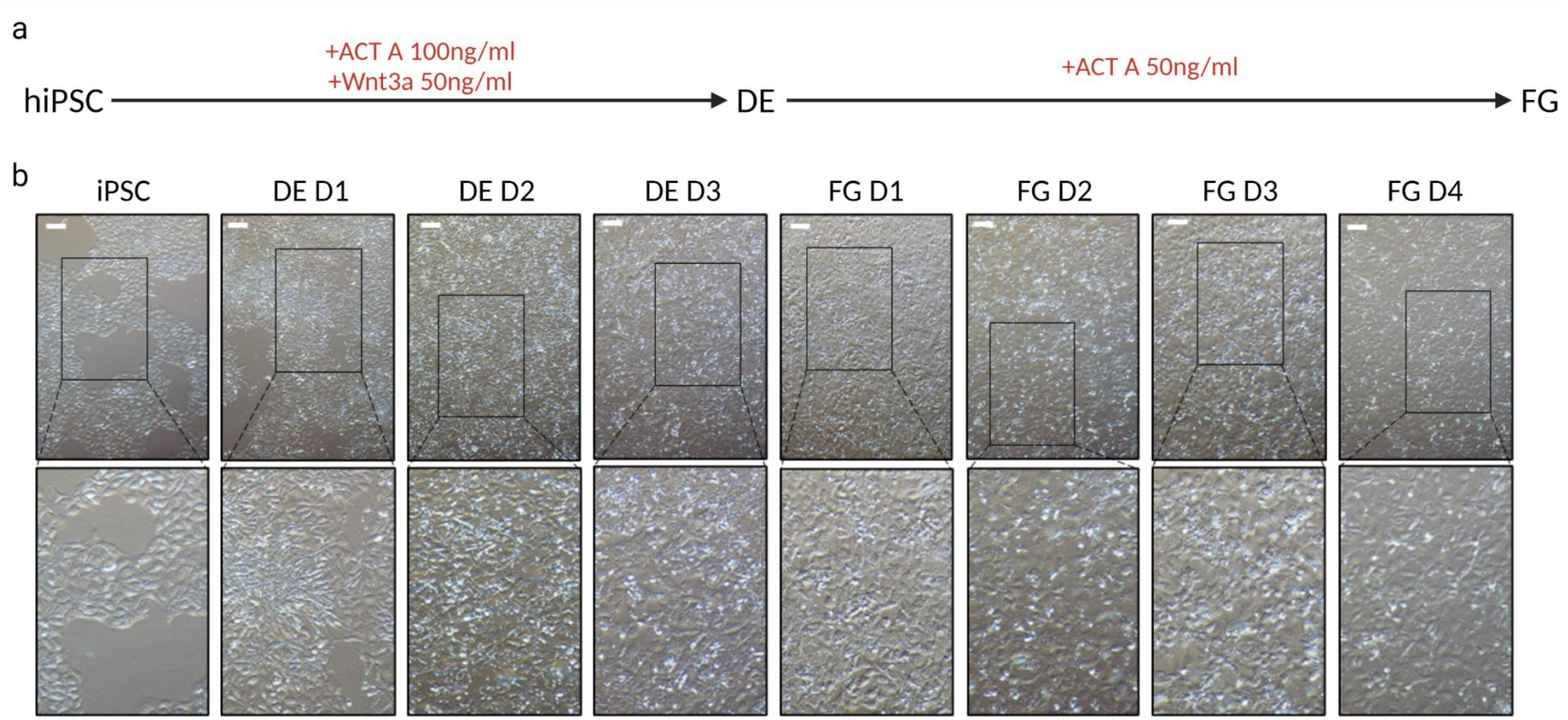

Figure 3-2 R-PAT M differentiation to DE and FG.

a) Schematic presentation of differentiation of hiPSC to DE and FG cell types; b) Light microscopy images of R-PAT M (p21) hiPSC differentiating to DE and FG. Scale bar = 200μm.

Immunostaining analysis showed that by DE D3 less than 15% of cells expressed NANOG, less than 3% of cells expressed OCT4 and less than 1% SOX2 (Figure 3-6). SOX17 was expressed by over 90% of cells and FOXA2 by over 95% of cells by DE D3. BRA(T) levels peaked between DE D1 and DE D2, with 55% and 64% of cells staining positively for this marker, respectively (Figure 3-7).

Immunostaining of foregut cells showed continued high expression (over 95%) of SOX17 and FOXA2 and no expression of CDX2 (Figure 3-8).

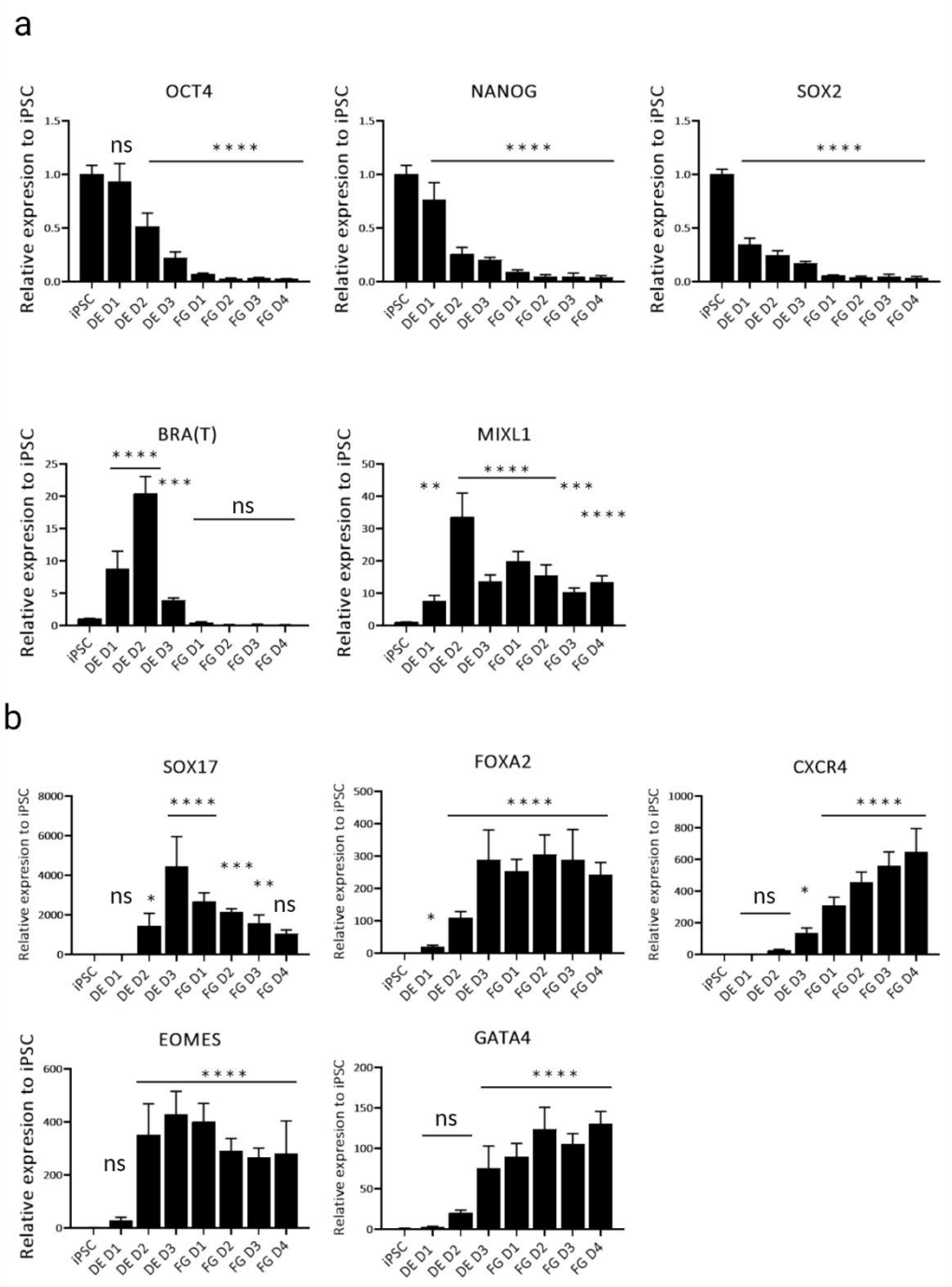

Figure 3-3 Gene expression changes during R-PAT M differentiation towards DE and FG. qPCR analysis of changes in a) pluripotency genes; b) mesendoderm genes; c) DE specific genes; Mean expression value at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=2)

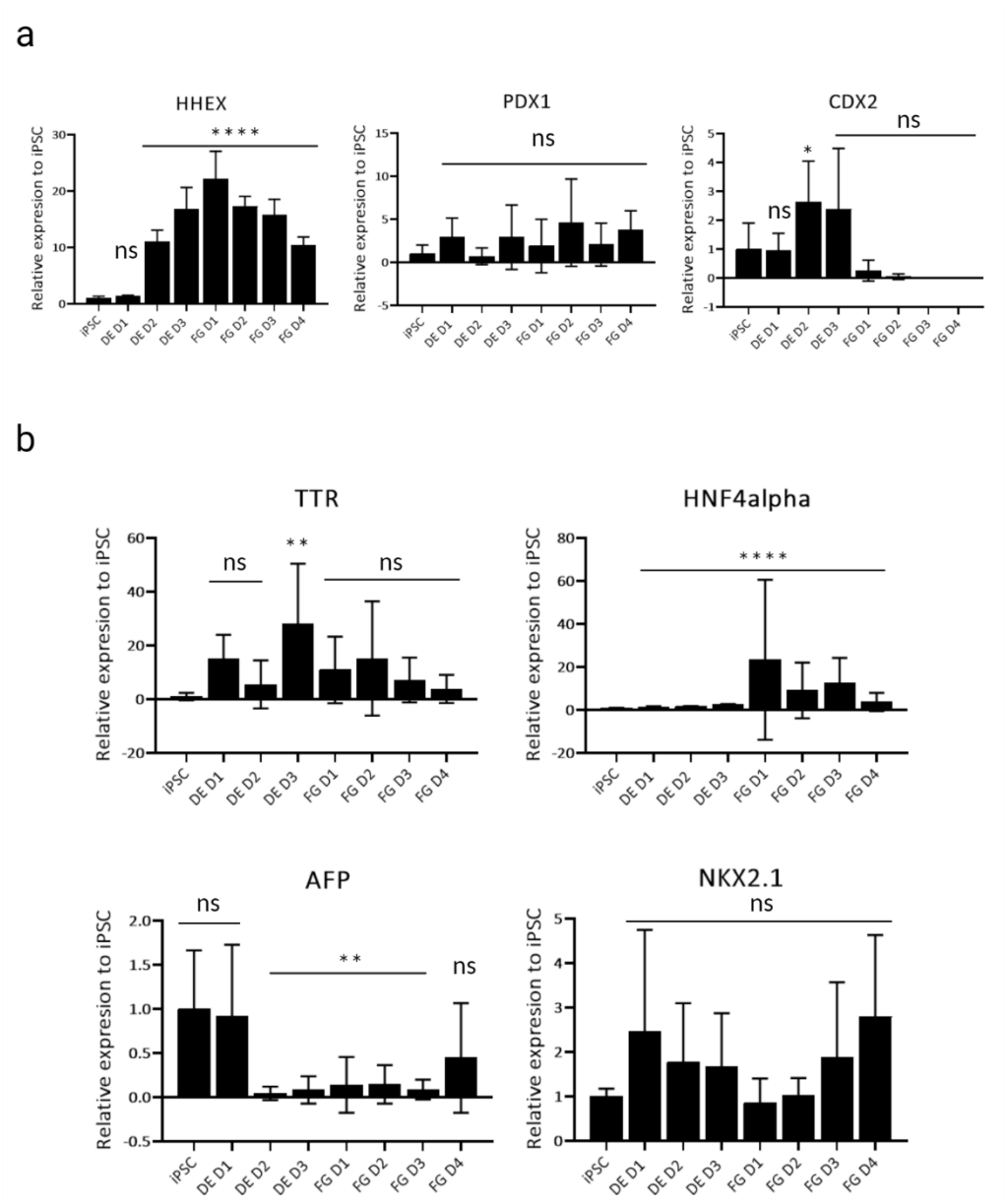

Figure 3-4 R-PAT M differentiation renders cells expressing FG specific markers.
a) QPCR analysis shows expression of FG specific marker HHEX with no expression of mid- and hindgut specific markers. b) FG cells are not primed towards liver or lung progenitors. Mean expression value at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=2)

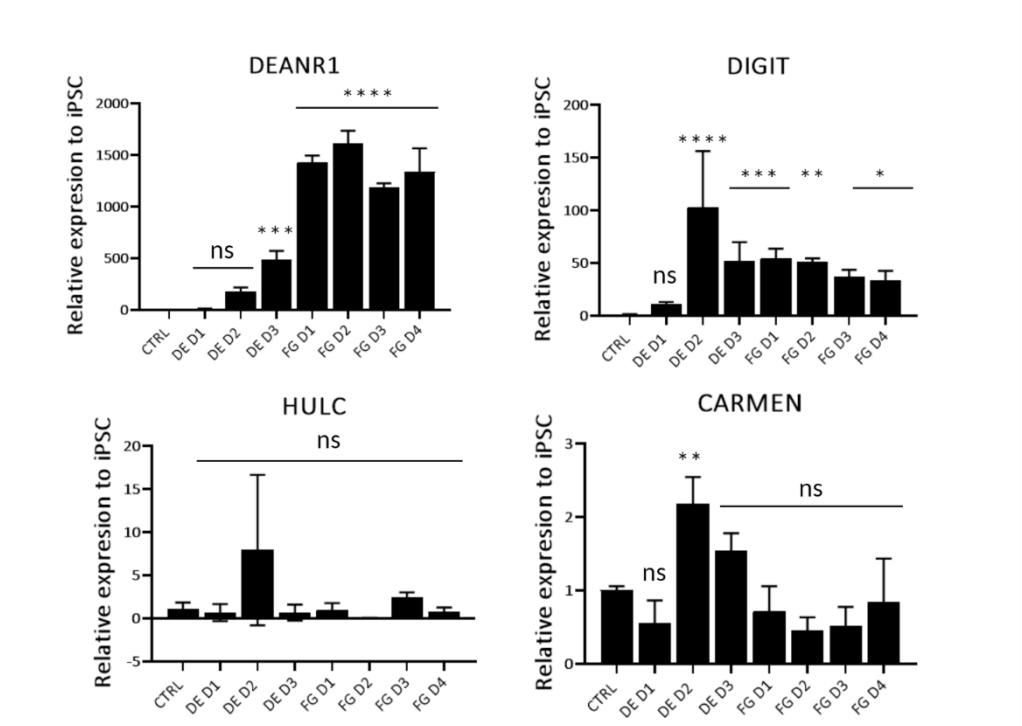

Figure 3-5 LncRNA expression during R-PAT M differentiation towards DE and FG.
QPCR analysis shows upregulation of endoderm specific lncRNAs DEANR1 and DIGIT, while cancer specific lncRNAs are not expressed. Mean expression value at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=2).

Next, R-PAT M hiPSC cell line was differentiated to definitive endoderm (DE) and hindgut cells following the protocol specified in Figure 3-9a. Appropriate marker expression profile was checked by qPCR and immunostaining. Once signals inducing the formation of HG are delivered after DE D3, the cells elongate and form a more densely packed monolayer with spheroids appearing usually around HG D4 (Figure 3-9b). Gene expression profile by qPCR shows upregulation of hindgut specific factor CDX2 from HG D1 and rapid suppression of FG marker CXCR4. PDX1, a gene specific for midgut development is virtually undetectable. HNF4α, TF important for liver as well as colon development, is upregulated on HG D2 and there is a continued expression of broader endodermal factors such as FOXA2, SOX17 and GATA4 from DE stage throughout HG stage (Figure 3-9c). Immunostaining analysis confirmed high expression of endodermal markers SOX17 and FOXA2, and upregulation of hindgut specific marker, CDX2 (<95% of cells positive by HG D4) (Figure 3-10

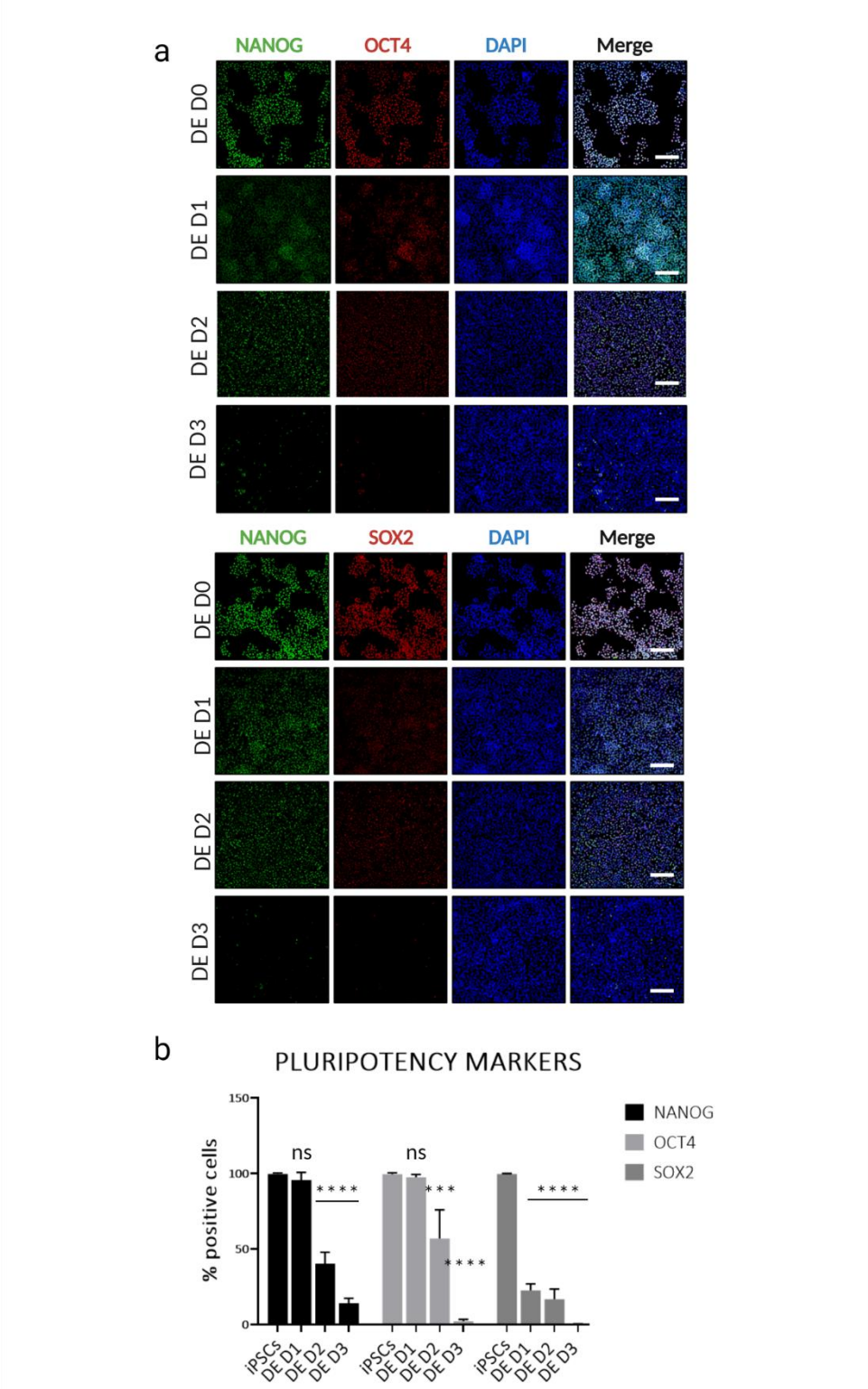

Figure 3-6 Pluripotency marker expression profile during R-PAT M differentiation towards DE.

a) Immunostaining images of differentiating R-PAT M cells. Scale bar = 200μm; b) Quantification of immunofluorescent images shows gradual decrease in pluripotency markers; Mean fluorescence intensity at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=3).

Following successful differentiation towards FG, R-PAT M cells were further induced towards hepatic and pancreatic fates following protocols in Figure 3-11a. During the specification of LPC from the FG, the cells undergo subtle morphological changes. The cells enlarge and become more rounded. During PPC specification from FG, the morphological changes are more pronounced and result in the formation of tight monolayer (Figure 3-11b). On the molecular level, LPC start expressing hepatic transcription factors such as TTR and HNF4α from LPC D1 and by LPC D4 there is a good expression of hepatic (AFP, ALB) and cholangiocyte (CK19) genes, demonstrating the bipotential nature of LPCs. At the same time, there is no expression of pancreas specific genes such as HLXB9 or PDX1, indicating that the pancreatic fate has been efficiently suppressed. PPCs show good expression of pancreas specific markers such as HLXB9, PDX1, SOX9, HNF1β and GATA6, with no expression of hepatic markers such as AFP or TTR, indicating that the protocol specifically produces PPCs (Figure 3-12).

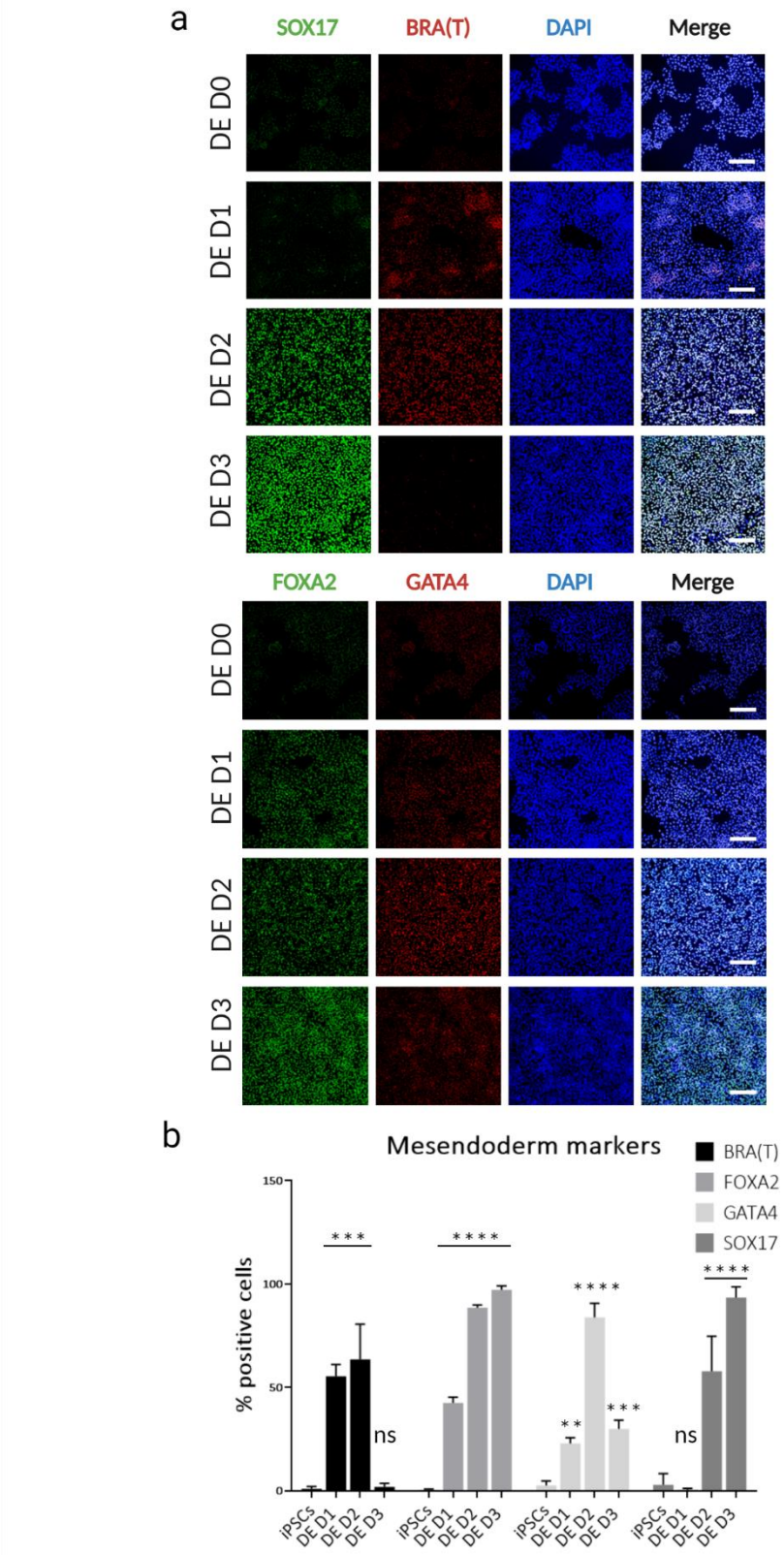

Figure 3-7 Mesendoderm marker expression during R-PAT M differentiation towards DE.

a) Immunostaining images of differentiating R-PAT M. Scale bar = 200μm; b) Quantification of fluorescence intensity shows transient increase in BRA(T) gene and gradual increase of SOX17 and FOXA2; Mean fluorescence intensity at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=3).

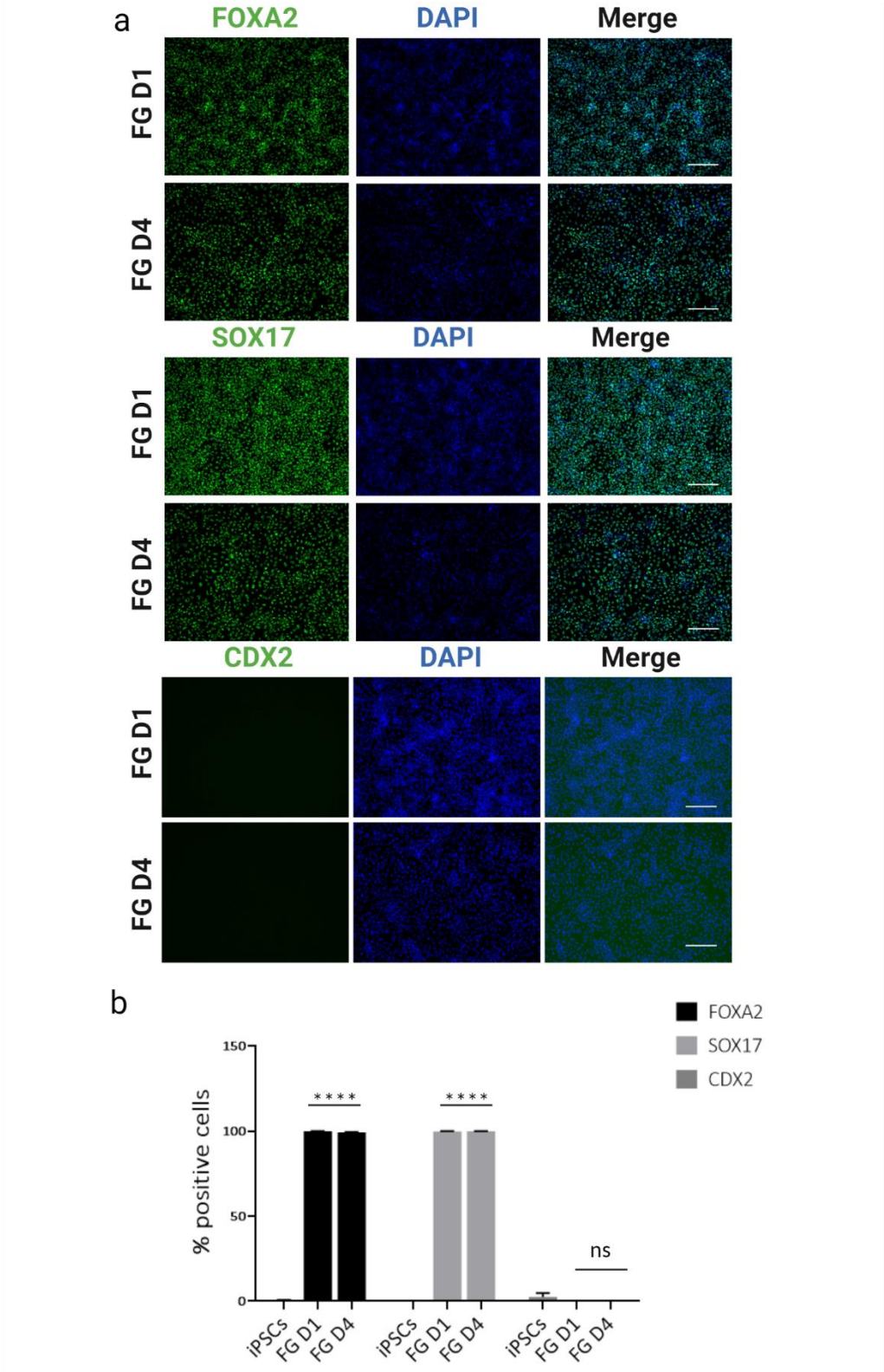

Figure 3-8 Endodermal marker expression in R-PAT M cells differentiated to FG.
a) Immunostaining images of differentiating R-PAT M cells. Scale bar = 200μm; b) High percentage (>95%) of FG cells express endoderm markers, SOX17 and FOXA2. Mean fluorescence intensity at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=3).

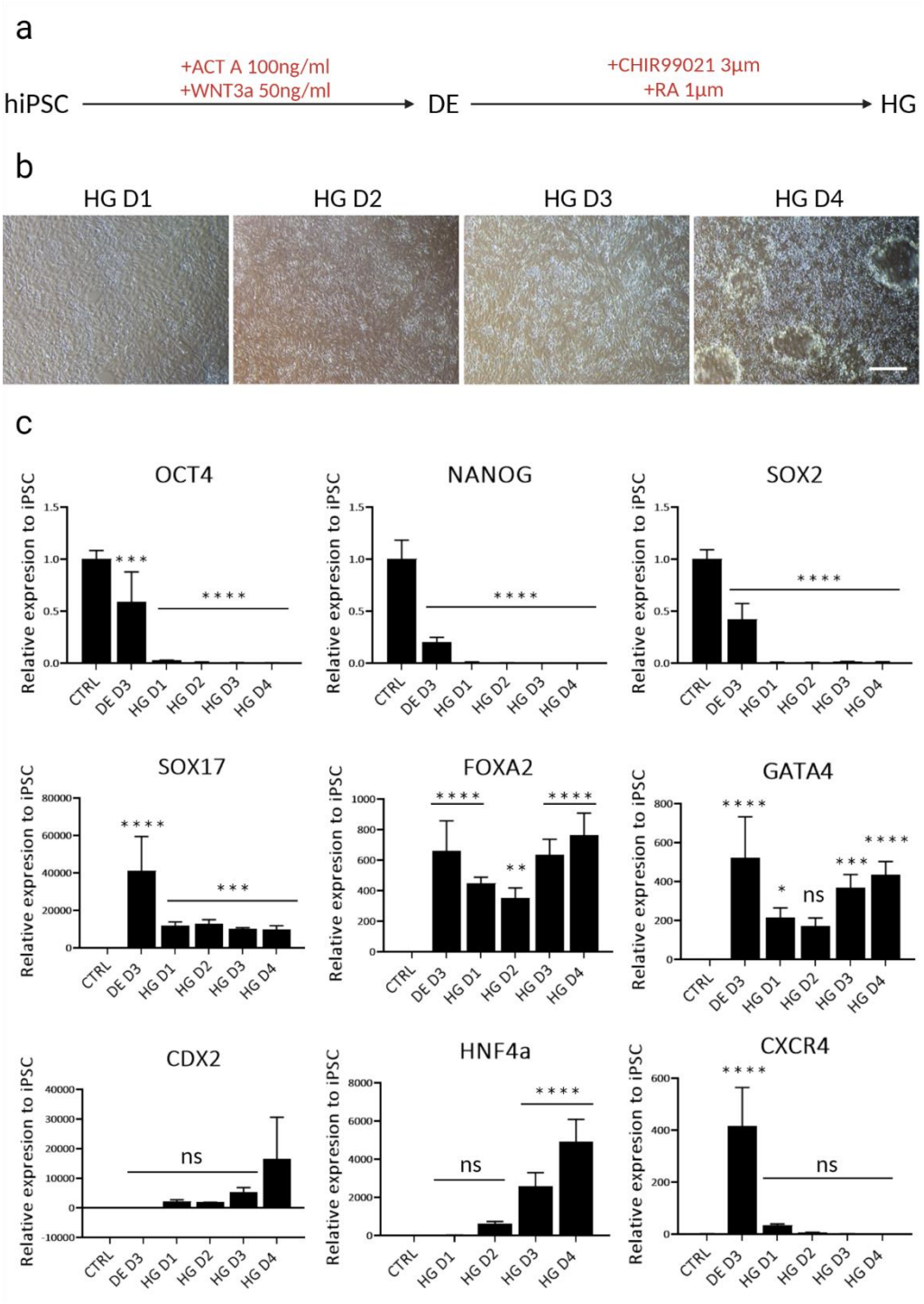

Figure 3-9 R-PAT M differentiation to HG.

a) Schematic of differentiation of hiPSC towards hindgut fate; b) Microscopy images of morphological changes during DE specification towards HG. Scale bar=200µm; c) QPCR analysis of gene expression changes. Mean fluorescence intensity at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=3).

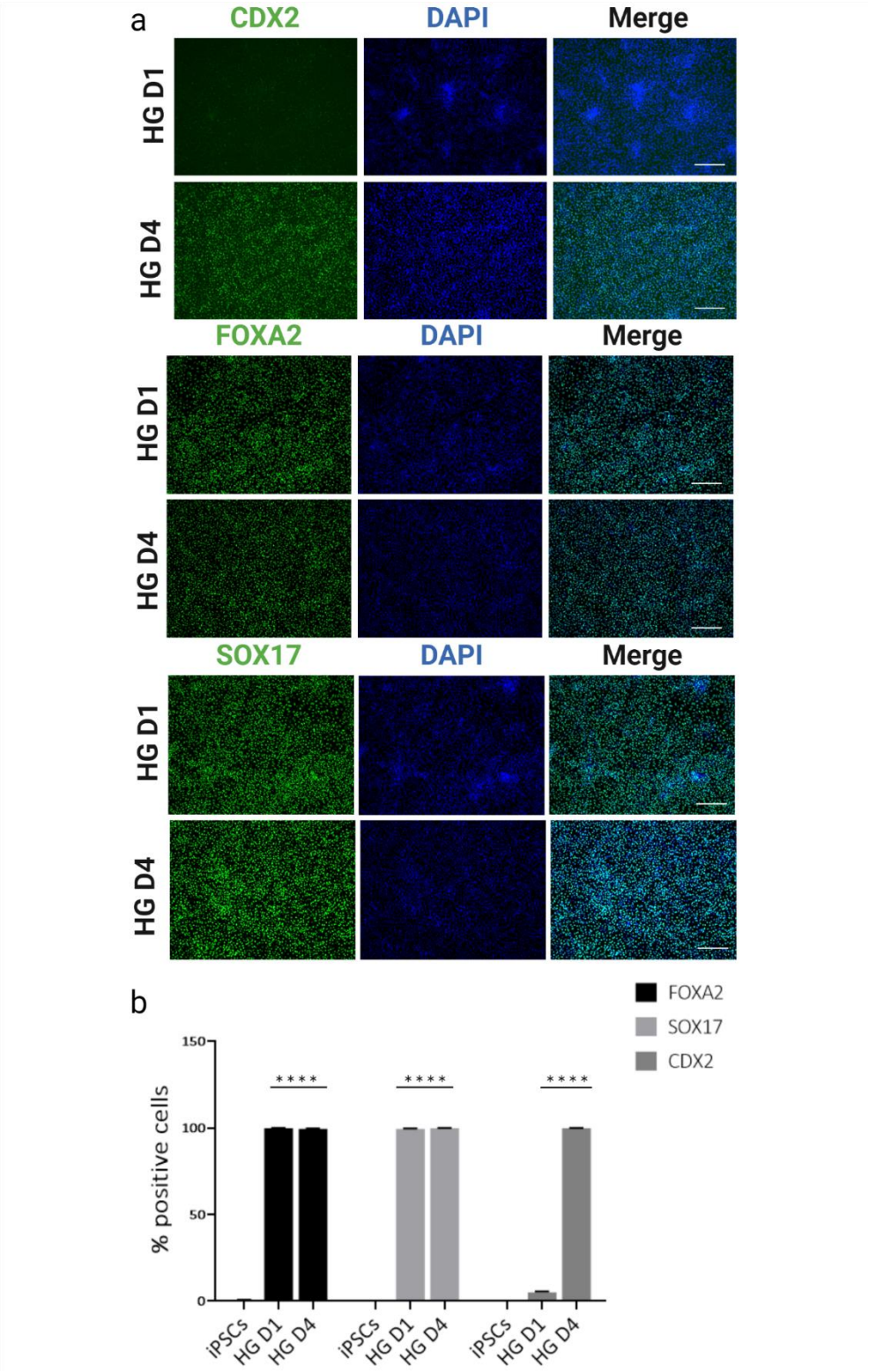

Figure 3-10 Endodermal marker expression in R-PAT M cells differentiated to HG.
a) Immunostaining images of differentiating R-PAT M cells. Scale bar = 200μm; b) At HG D4 high percentage of cells (>95%) express CDX2. Mean fluorescence intensity at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=3).

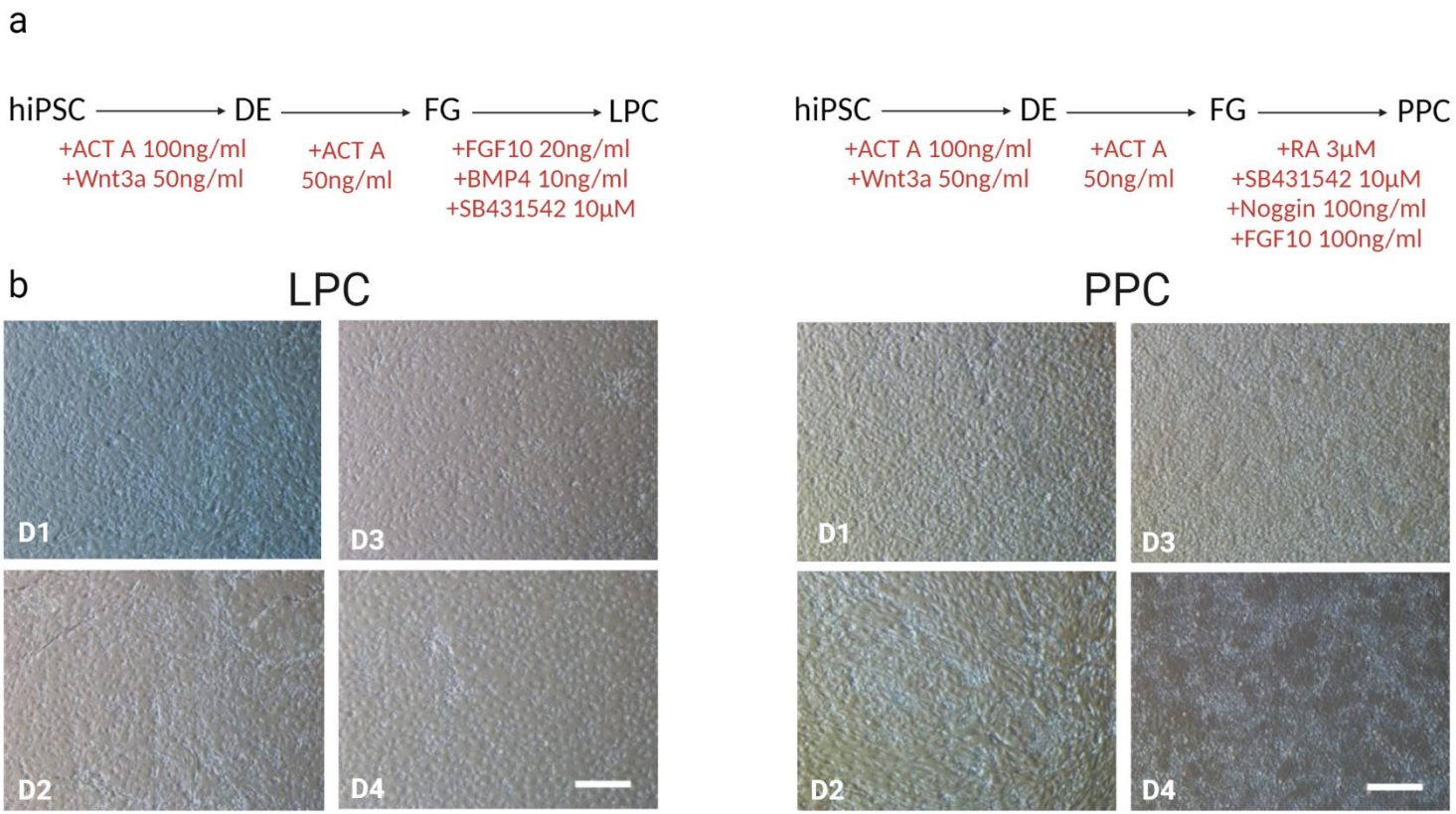

Figure 3-11 R-PAT M differentiation towards LPC and PPC.

a) Schematic of differentiation protocols; b) Microscopy images of morphological changes during differentiation. Scale bar = 100 µm;

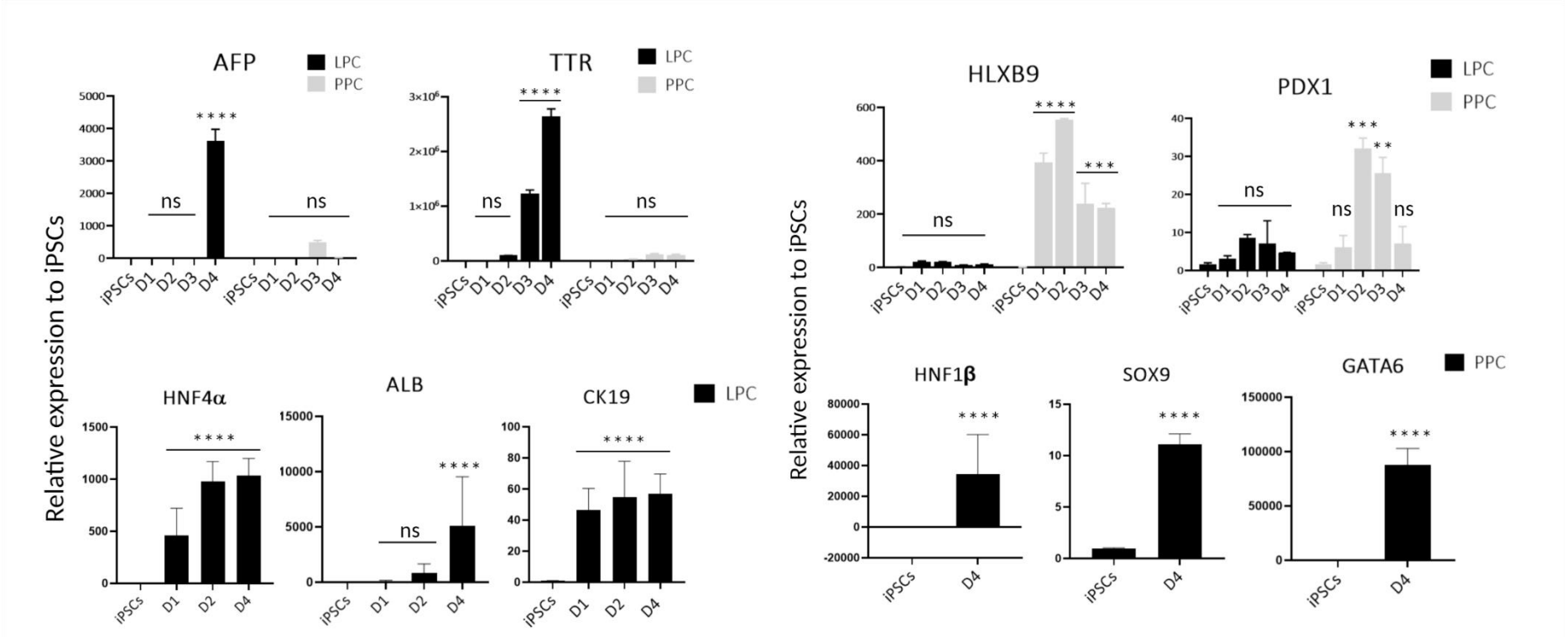

Figure 3-12-Gene expression profile of R-PAT M differentiation to LPC and PPC.

QPCR analysis of gene expression in LPCs and PPCs. Mean fluorescence intensity at each day compared to iPSCs using one way ANOVA with Dunnett's multiple comparison test (n=3).

3.3.2. Manipulation of signalling pathways during endodermal linage development

To further demonstrate that our platform is a reliable model for human endoderm development we have manipulated specific signalling pathways to show their importance for directed differentiation in line with events occurring during the in *vivo* development.

CDX2 TFs has been shown to be essential for the establishment of intestinal identity during patterning of the primitive gut tube in mice (Gao et al., 2009) and that its expression is activated by Wnt signalling (Sherwood et al., 2011). Therefore, inhibition of Wnt signalling during HG specification from DE should abolish CDX2 expression and prevent the acquisition of intestinal identity by DE cells. R-PAT M cells were differentiated towards HG fate with either the addition of CHIR99021 or DKK-1. CHIR99021 is a small molecule activator of Wnt signalling. It activates Wnt signalling by selective inhibition of glycogen synthase kinase 3 (GSK-3), an enzyme in the Wnt signalling pathway that is part of a complex degrading β-catenin, a transcriptional co-activator. Dickkopf related protein 1 (DKK-1) is a Wnt antagonist. It inhibits Wnt signalling by forming a complex with LRP5/6 receptors, making them unavailable for the Wnt ligand. Cells differentiated to HG in the absence of Wnt signalling show morphology resembling that of FG rather than HG, with cells remaining loosely packed and rhomboidal in shape (Figure 3-13a). The cells fail to express HG specific TF CDX2. There is a slower and weaker suppression of FG specific CXCR4 in cells treated with DKK-1 and weaker activation of HNF4α. There are differences in the level of expression of other endodermal markers such as SOX17 and GATA4, while inhibition of Wnt signalling seems to have no effect on FOXA2 expression. DKK-1 treated cells also started expressing TTR gene, which is a marker of liver specification and fail to upregulate a marker of intestinal progenitor cells, LGR5 (Figure 3-13b). This shows that our development model mirrors the results of in *vivo* experiments on mouse embryos. As similar experiments are not possible on

human embryos, hiPSC differentiation can be a reliable alternative for a more detailed study of the human liver development.

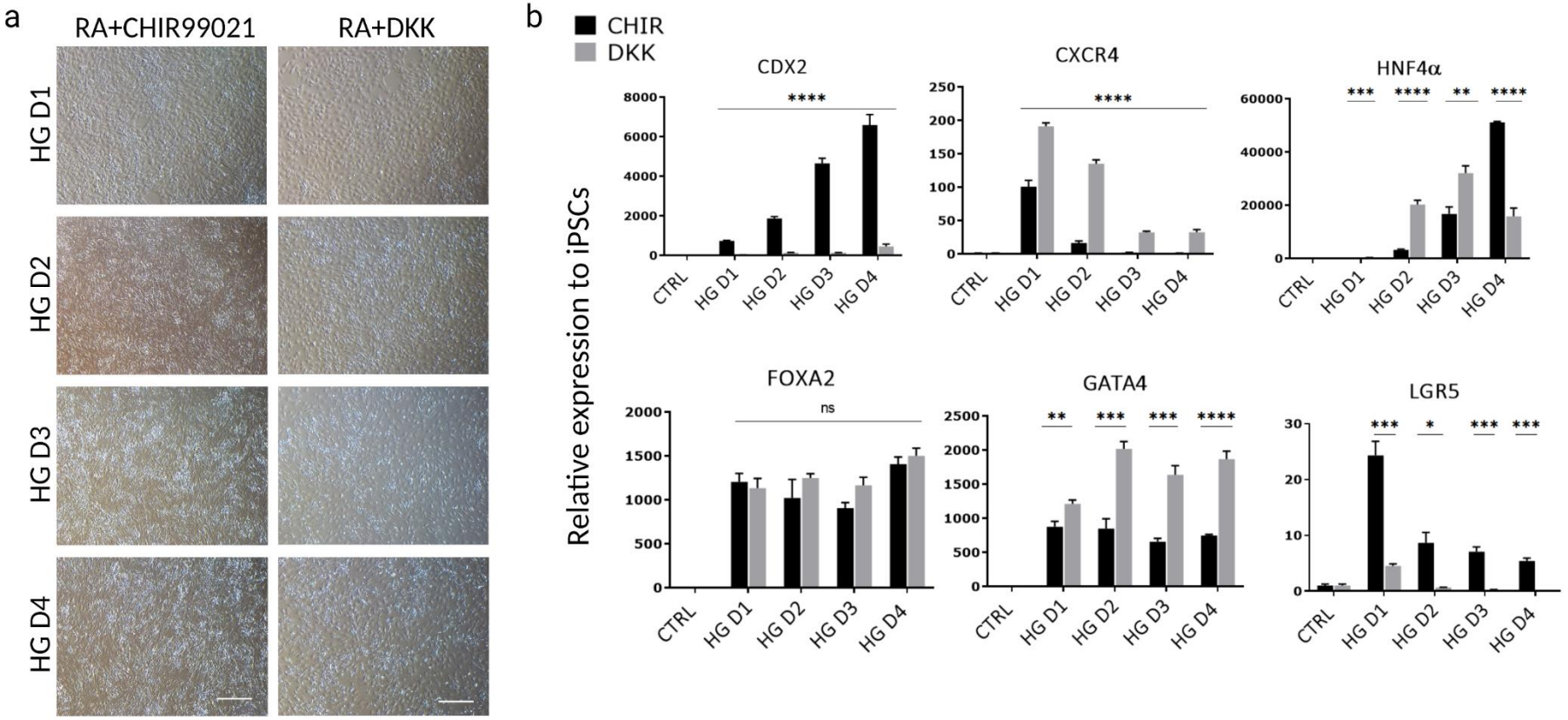

Figure 3-13 Wnt signalling pathway manipulation during specification of HG cells.

a) Light microscopy image of morphological differences. Scale bar = 100μm; b) QPCR analysis of gene expression profiles of HG cells differentiated in the presence (CHIR) or absence (DKK-1) of Wnt signalling. Data presented as mean ± SD. Multiple unpaired t-tests with Welch correction (n=2).

3.3.3. Transfection optimization of foregut

The interest of this study are TFs responsible for the specification of LPC from foregut cells under the direction of BMP4 signalling. To study the gene function, cell transfection of the foregut monolayer right before induction towards hepatic fate will be necessary.

Our choice of method of transfection was chemical transfection with Lipofectamine, due to its cost, simplicity of use and previous experience. GFP plasmid was used to monitor transfection efficiency by fluorescent microscopy or flow cytometry. R-PAT M cells differentiated to FG were transfected with varying ratios of Lipofectamine to DNA to identify the amount of reagent giving the best transfection efficiency. The cells were transfected at FG D2, FG D3 and FG D2 and FG D4 to identify the best timing for transfection. GFP fluorescence was checked 24 hours post transfection. Fluorescent microscopy images revealed very few cells positive for GFP (Figure 3-14a). Flow cytometry analysis confirmed low transfection efficiency with less than 10% of cells expressing GFP when cells were transfected at FG D2 and FG D4 (Figure 3-14b). The 10% rate of transfection success is insufficient for our experimental needs, therefore we tested several other lipid transfection reagents available. Four reagents were selected for testing: X2 (Mirus), FuGene (Promega), K2 (Biontex) and K4 (Biontex). The selected reagents were appropriate for the transfected material and cell type. FG D3 cells were transfected with two different ratios of reagent to DNA for each tested reagent. Transfection efficiency and cell toxicity was qualitatively assessed using light and fluorescent microscopy. X2, FuGene and K4 at 2:1 ratio seemed to have the highest transfection efficiency although overall, the transfection efficiency appeared very low for all tested reagents. Additionally, 4:1 ratio of all reagents seemed to cause substantial cell toxicity as evident by the thinning of the FG monolayer 24hrs post transfection (Figure 3-15).

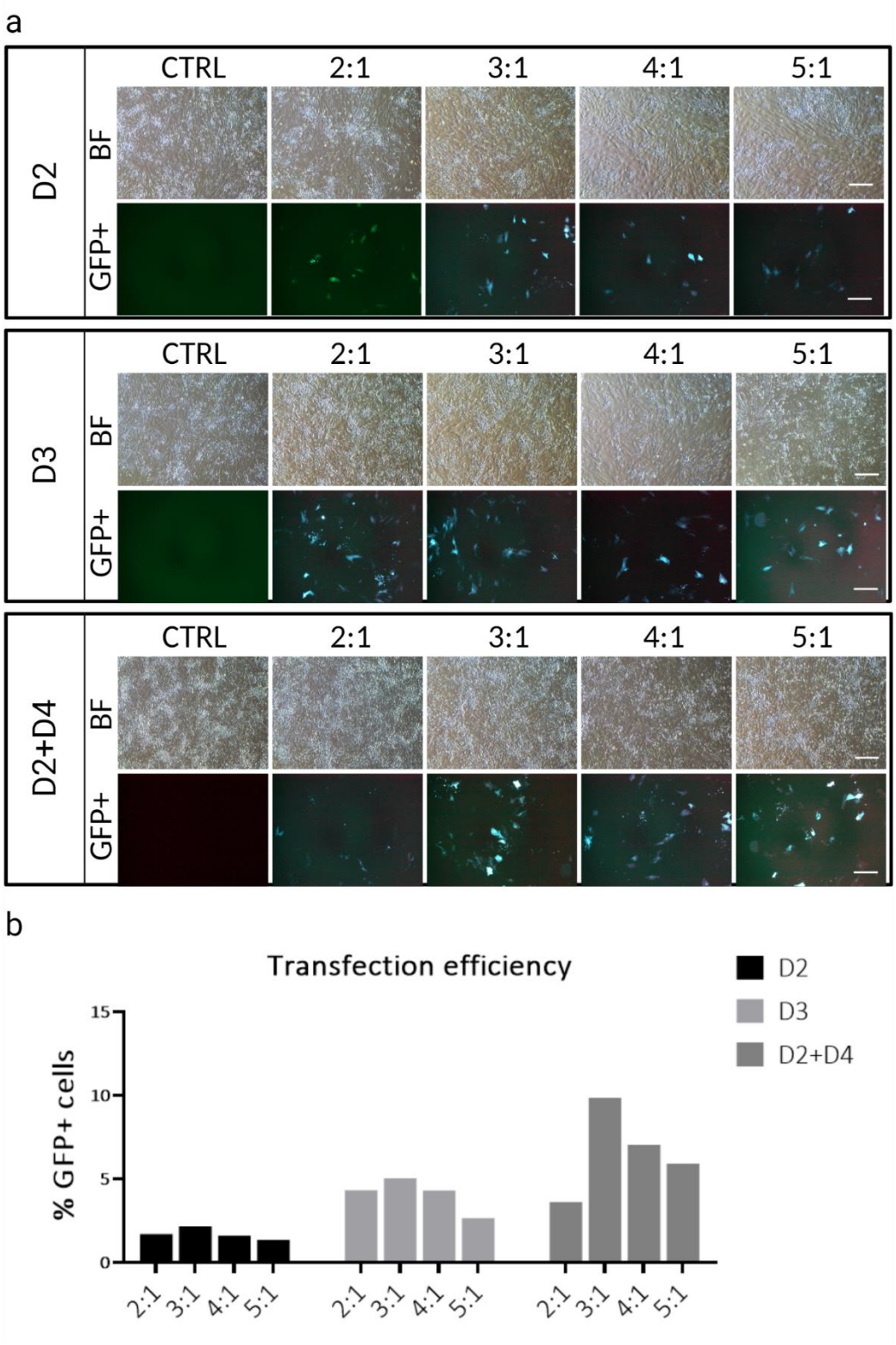

Figure 3-14 Determination of transfection efficiency using Lipofectamine.
a) Light and fluorescent microscopy images of FG cells transfected at various points during the specification and with varying Lipofectamine:DNA ratio. Scale bar = 100μm; b) Quantification of GFP+ cells using flow cytometry; n=1.

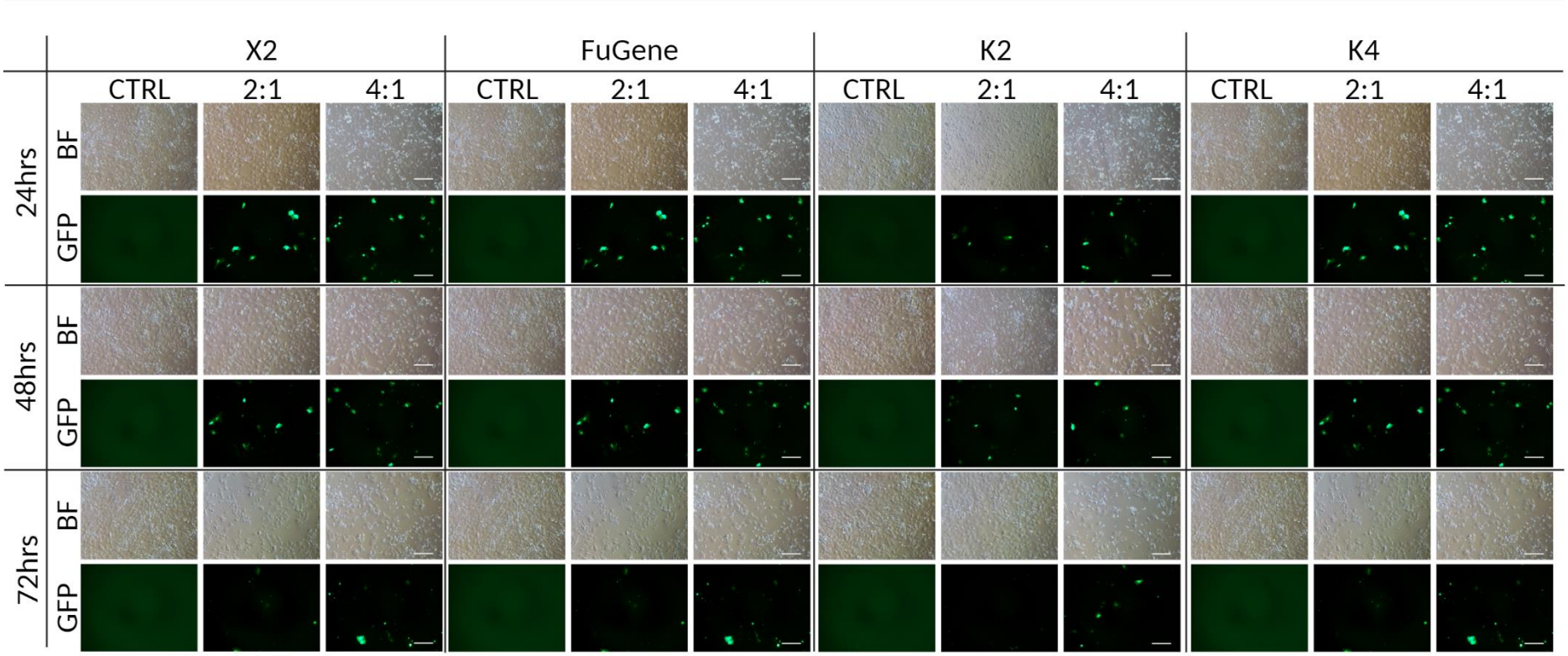

Figure 3-15 FG transfection efficiency tests.

R-PAT M cells differentiated to FG were transfected at FG D3 with varying ratios of reagent to DNA of 4 different reagents. Cell viability and GFP expression were checked using light and fluorescent microscopy 24, 48 and 72hrs post transfection. Scale bar: 100μm.

3.3.4. Transfection optimization of iPSCs

As the transfection of FG monolayer was inefficient in our hands, we decided to change the approach to gene studies and employ a tetracycline inducible system that can be introduced into the cells at the undifferentiated stage and its effects studied at a chosen time of cell differentiation. Therefore, it was now necessary to identify an efficient method for hiPSC transfection.

First, we tested four different lipid reagents (Figure 3-16). K2 and K4 showed very few GFP positive cells and appeared cytotoxic. FuGene and LT1 reagents appeared to achieve the best transfection efficiency (either ratio) from the reagents tested with minimal cell toxicity. However, on close examination both of those reagents seemed to transfect cells mainly at the edge of the colonies. R-PAT M cells form colonies that are denser in the middle and more loosely packed on the edges. The lipid reagents seemed to mainly access the cells on the edge of the colonies while a substantial amount of the tightly packed cells in the middle of the colony remained untransfected (Figure 3-17).

Next, due to the difficulty of accessing the tightly packed cells in the middle of the colony, we tested a physical method of transfection: nucleofection. For nucleofection cells are dissociated which would eliminate the problem of tightly packed colonies being inaccessible to the transfection reagents.

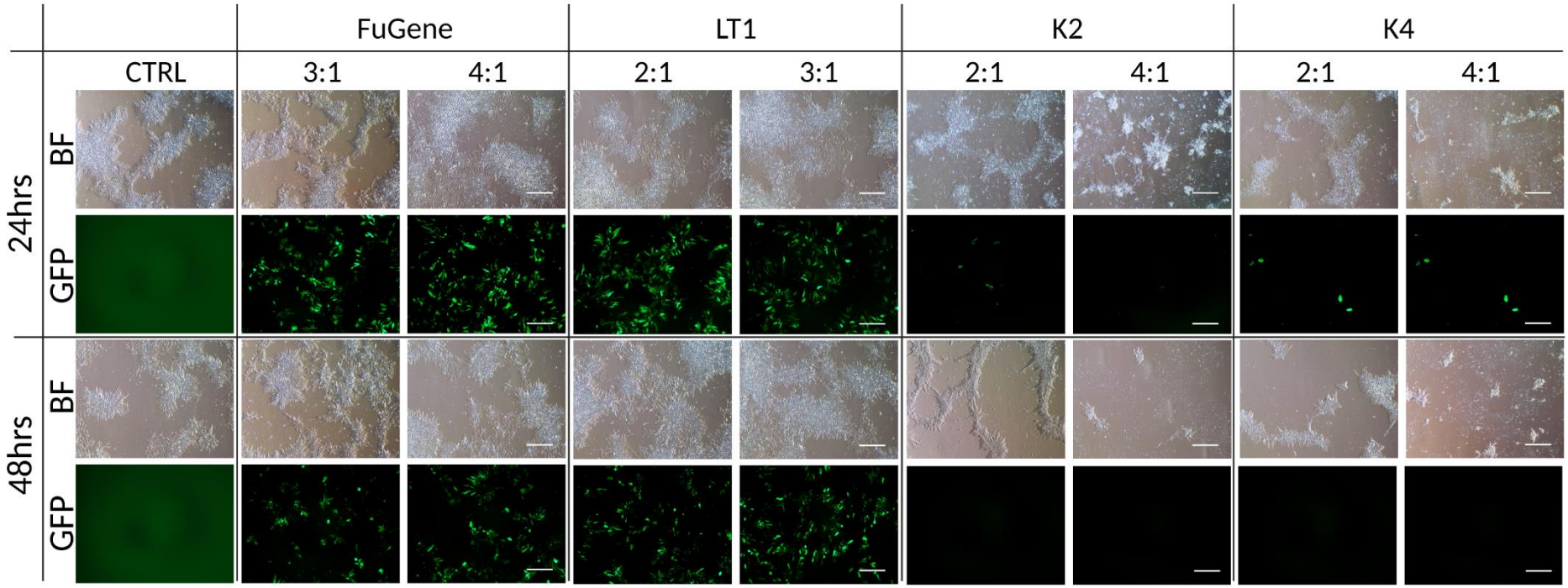

Figure 3-16 R-PAT M transfection efficiency testing.

R-PAT M cells were transfected with varying ratios of reagent to DNA. GFP expression was tested 24 and 48hrs post transfection. Scale bar: 100μm.

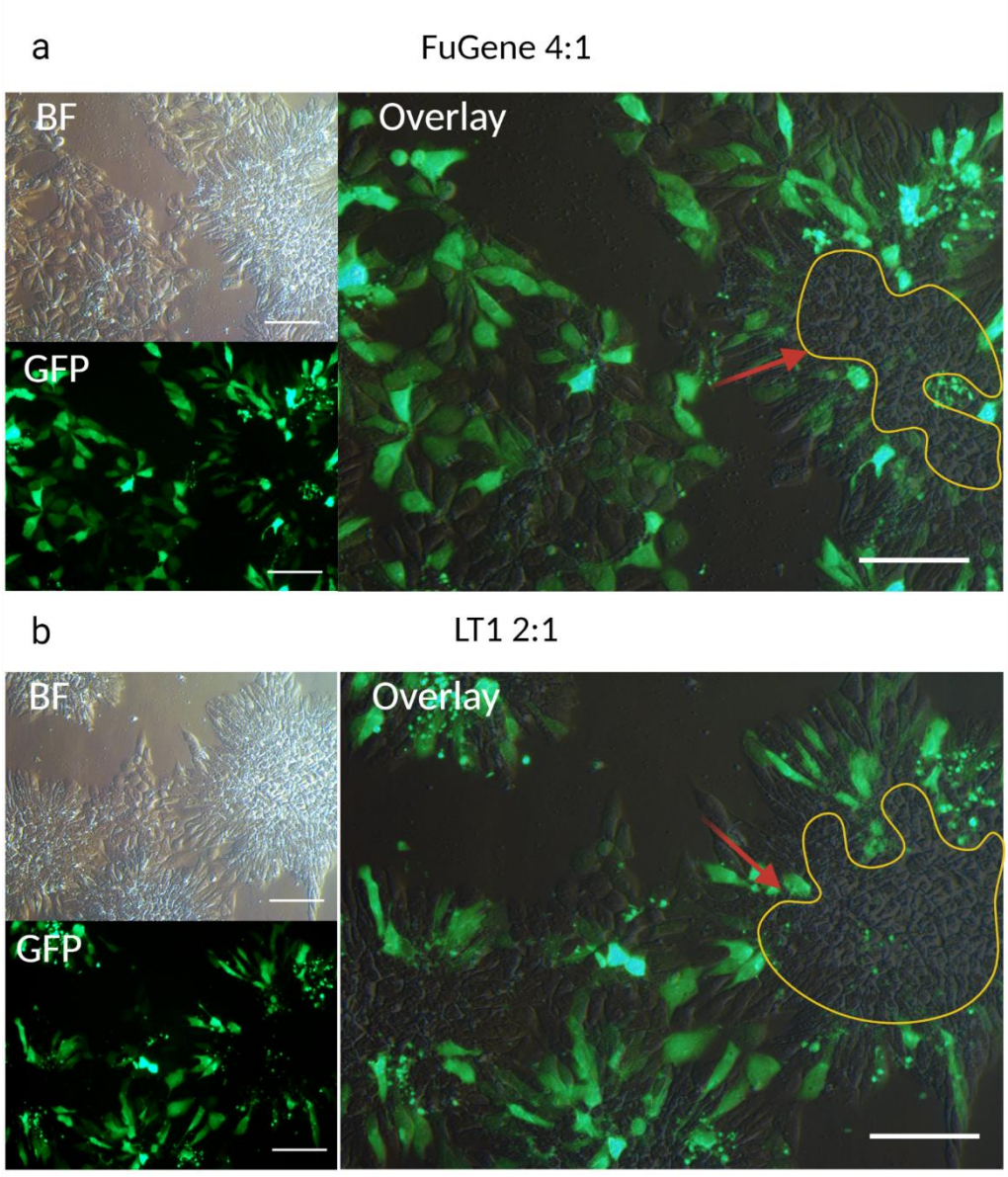

Figure 3-17 R-PAT M transfection with lipid reagents.
a), b) Light and fluorescent microscopy images of R-PAT M cells transfected with FuGene and LT1 regents, respectively. Red arrows point to denser areas of the R-PAT M colonies that remain untransfected. Scale bar = 100μm.

To achieve maximum transfection efficiency, we screened seven nucleofection programmes recommended for hiPSC on Amaxa-4D nucleofector to identify the most optimal one. R-PAT M cells were transiently transfected with pmaxGPF vector provided as positive control with the Lonza P3 primary cell nucleofection kit. GFP expression was checked microscopically at 24 and 48hrs post nucleofection (Figure 3-18a) and cells were collected for flow cytometry analysis 48hrs post nucleofection. For flow cytometry analysis cells were additionally

stained with PI dye to assess cell viability. All seven programmes showed good survival of the cells (62-74%). The programme with highest proportion of successfully transfected cells was DN-100 (66.4%) and it was used for all subsequent transfection experiments (Figure 3-18b). Next, we proceeded with optimization of the cell number per nucleofection to see if it can further improve nucleofection efficiency. Between 200k and 320k of cells per nucleofection was tested with the DN-100 nucleofection programme. Each cell number resulted in high transfection efficiency of at least 70.22%. Highest average transfection efficiency was achieved when 100k cells were nucleofected per nucleofection cuvette (85.32% ± 4.8%). This transfection efficiency differed significantly only form the 320k cells used per nucleofection (70.22% ± 2.76%). Therefore, the number of cells per nucleofection of between 100k and 300k could be used to achieve high proportion of transfected cells (Figure 3-19).

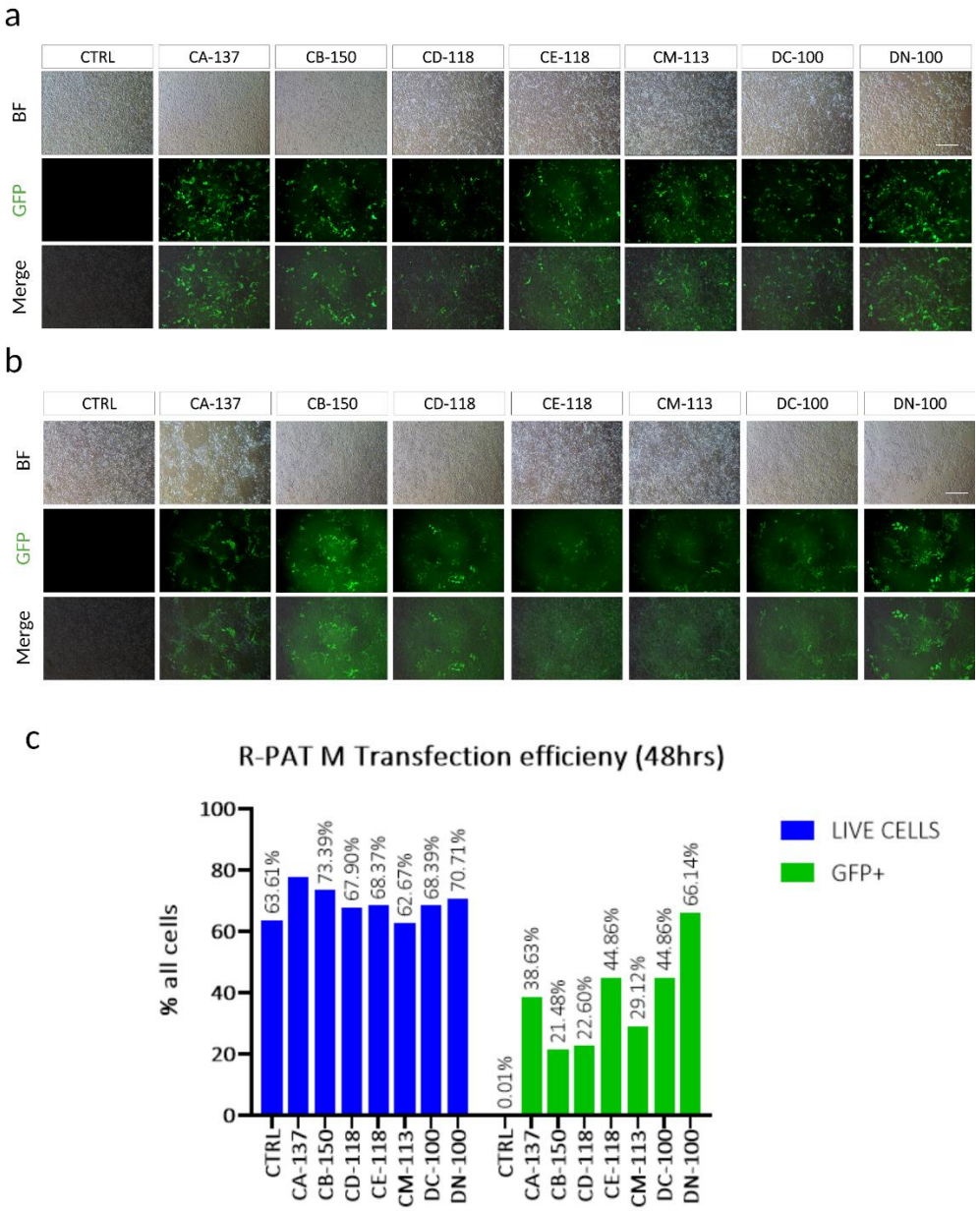

Figure 3-18 Optimization of transfection efficiency of R-PAT M using Amaxa-4D nucleofector.

a),b) Live fluorescent imaging of transfected R-PAT M cells using different nucleofection programmes on Amaxa-4D. Scale bars = 200μm; c) Comparison of cell viability and transfection efficiency across seven nucleofection programmes. Transfected cells were analysed using flow cytometry at 48hrs post nucleofection (n=1).

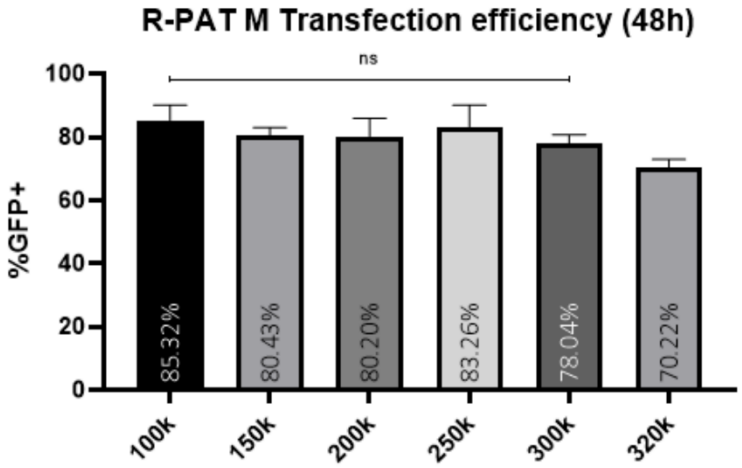

Figure 3-19 Optimization of seeding density for transfection of R-PAT M hiPSC line.
Varying numbers of cells were transfected with pmaxGFP plasmid. 48hrs post transfection cells were collected for flow cytometry analysis of GFP expression. Data presented as mean ± SD (n=3). One way ANOVA with Tukey's multiple comparison test.

3.4. Discussion

3.4.1. HiPSC as models of early human endodermal development

The differentiation protocols applied in this project mimic the pathway of natural early development of human endoderm. The cells are first directed towards a definitive endoderm identity with transient expression of mesendoderm genes as happens during the primitive streak formation and subsequent emergence of the three germ layers. From DE, the cells can further be specified into one of the regions of primitive gut tube, depending on which cells of endodermal origins are to be established. Foregut specification allows for further differentiation towards liver, lungs, pancreas or thyroid while hindgut enables differentiation of intestinal fates (Zorn, 2008). LPCs are specified by exposing foregut cells to FGF and BMP4 signalling to mimic signals from surrounding cardiac mesoderm and septum transversum mesenchyme, respectively (Rossi et al., 2001). In vivo, this signalling results in the specification and proliferation of bipotential hepatic progenitors that form liver bud and subsequently expand into the surrounding mesenchyme, eventually leading to the formation of foetal liver. At each stage of development, a different set of TFs is activated orchestrating the expression of appropriate genetic programme. This allows the identification of each cell type in the in vitro cell culture. The protocols used for differentiation in this project render cells that specifically express appropriate genes at each stage, supporting our assertion that differentiated hiPSCs can be applied as models of early endodermal development. Furthermore, the disruption of differentiation programme by e.g.: removing one of the signalling molecules leads to similar consequences as seen in in *vivo* studies. This is shown in the experiment where Wnt signalling was blocked during HG specification from DE. As reported for in *vivo* setting, this results in the lack of CDX2 expression, weaker HNF4α expression and inappropriate expression of FG-related markers such as CXCR4 (Sherwood et al., 2009).

Although animal models have contributed hugely towards our understanding of development, the need for more relevant models becomes apparent as detailed studies reveal how even highly conserved elements can function differently between two species. Odom and colleagues showed that four hepatic TFs have vastly different binding sites in humans and mice, even though those TFs are highly conserved between the two species (Odom et al., 2007). In another study, single cell RNA-seq. of human and mouse preimplantation embryos showed that, although some gene expression patterns are shared between the species, there are differences in TF expression and signalling pathway employment between mouse and human pluripotent epiblast (Blakeley et al., 2015). It is, therefore, necessary to find more accurate models of early human development. As the research on human embryos is highly controversial and limited, the use of pluripotent stem cells, especially induced pluripotent stem cells, offers a valuable alternative. It can also contribute to a decrease in the use of foetal tissue for biomedical research which also comes with ethical issues.

In this project the cells are cultured and differentiated as a 2D monolayer. This allows for well controlled, reproducible and focused experiments being conducted on a specific cell type. This type of cell culture can successfully contribute to expanding knowledge of molecular mechanisms governing endodermal development. For example, Fisher and colleagues were able to show that GATA 6 is essential for DE formation (Fisher et al., 2017) and Twaroski and colleagues identified mediators of FGF2 signalling during the hepatic progenitor cell formation (Twaroski et al., 2015).The advantage of 2D monolayer culture is the relative ease and cost effectiveness, consistent access of all cells to nutrients delivered by the culture media, application to high throughput phenotyping and ease of specific molecular manipulation with e.g.: siRNAs or overexpression vectors. However, there are also drawbacks to this type of cell culture, chief among them the fact that cells rarely exist in a 2D environment in the body. In *vivo*, cells are surrounded by other cells, have contact with extracellular membrane (ECM) proteins and are exposed to mechanical forces. The dynamic of how cells are exposed to soluble molecules is also quite

different, consequences of which can be especially observed in development during spatial patterning by e.g.: components of BMP signalling pathway (reviewed in (Bier and De Robertis, 2015). The lack of 3D environment and complete ECM can have an impact on cell migration, proliferation or morphology and affect overall cell behaviour (reviewed in (Baker and Chen, 2012).

There are improvements to 2D cell culture systems that can help mimic a more natural environment. Micropatterned substrates and nanotoporgaphic materials have been used to imitate adhesion and mechanics that influence cell behaviour in 3D environment. Micropatterned substrates are made by coating the cell culture dish with cell-adhesive or protein-absorption-resistant regions that can be arranged in various designs to control cell shape, cytoskeletal architecture, spreading, multicellular organization and stem cell differentiation (Vega et al., 2018, Zhang et al., 2022). Nanotopographic materials are substrates with nano-scale topological features that can alter not only how the cells attach to the substrate but also the mechanical properties of the cell culture environment. The stiffness of the substrate can affect cell adhesion and proliferation (Zhou et al., 2022) or direct adult stem cell differentiation (Xie et al., 2018). Microfluidic platforms have been developed to address the issue of nutrient distribution that more accurately reflects in *vivo* conditions. Static or dynamic platforms can recreate morphogen gradient present during development improving the models of tissue patterning and specification during development (Cui et al., 2020).

Alternatively, cells could be grown in 3D as either self-assembled organoids or seeded onto scaffolds. 3D cell culture allows for more complex cell-cell or cell-ECM interactions, improved cell morphology, better spatial organization and more intricate model of in *vivo* environment. Addition of scaffolds can also help in provision of physical or chemical cues to the cells further improving the accuracy of the model.

Another layer of complexity can be added by 3D models utilizing multiple cell types within the organoid. Takabe and collegues have created self-assembling

liver bud organoids composed of three iPSC-derived progenitor populations: hepatic endoderm, STM cells and endothelial progenitors (Takebe et al., 2017). Such models can be applied to the study of factors important for cell-cell interactions or cell assembly during liver bud formation and are an interesting avenue to explore once candidate TFs crucial for LPC specification have been identified.

However, 3D culture also comes with a variety of disadvantages. Larger 3D structures can present a challenge for nutrient delivery resulting in heterogeneity or necrotic areas. Likewise, uniform targeting of the cells with transfection reagents for molecular manipulation would also not be as straightforward as in a 2D system. It is also substantially more difficult to image 3D cell structures. It is more expensive and more difficult to handle, sometimes requiring expensive equipment (e.g.: bioreactors), whereas scaffold-based 3D cell culture can present a challenge when retrieving cells for further analysis.

In this project we wanted to identify BMP4-activated TFs that are crucial for hepatic specification using whole transcriptome analysis. For such experiment, high numbers of cells are necessary ideally collected from pure cultures of the specific cell type of interest. As 2D cell culture allows for synchronized, reproducible and sequential differentiation of iPSCs towards the desired cell type, it was appropriate and sufficient for the purpose of this project. Even distribution of nutrients and signalling molecules, ability to target cells for genetic studies, as well as ease of cell collection were of most importance for collection of data and further study of molecular mechanisms governing the specification of LPCs.

3.4.2. Transfection methods

The aim of transfection was to deliver siRNAs or overexpression plasmids (OEP) at the late stage of FG specification to ensure that once the LPC specification begins there are appropriate levels of siRNAs/OEP to prevent the expression/start the production of the gene of interest. The initial method selected for FG transfection was chemical using lipid reagents. This method is straightforward, does not require specialist equipment and has been reported

to have good efficiency (Dalby et al., 2004). Therefore, after our first choice of regent, Lipofectamine, failed to efficiently transfect the FG monolayer we have decided to test several other available reagents. However, all the tested reagents showed very poor transfection efficiency. The most likely reason for poor transfection efficiency was the confluency of the FG monolayer. For most lipid reagents the recommended confluency levels during transfections, as stated in the kit inserts, is 60-90%. Differentiation towards LPC yields a monolayer at DE D3 and during FG specification proliferation is more rapid during the first two days of the process, slowing down in FG D3 and FG D4. This could create unfavourable conditions for transfection as cells are more likely to take up nucleic acids when actively dividing. Additionally, the FG monolayer is tightly packed creating additional hindrance for the delivery of transfection reagents. Observations made when undifferentiated hiPSC colonies were transfected with lipid reagents support this explanation. On inspection with fluorescent microscope, the tightly packed centres of the colonies remain virtually GFP free and most transfected cells are on the edges of the colonies, in areas where they are not as packed as in the middle. There is, however, a report of successful foregut transfection with siRNA, using Lipofectamine reagent, therefore it is possible that the issue causing low transfection efficiency in our study was due to another factor (Banerjee et al., 2018).

Transfection using viral vectors is frequently applied in situation where lipid regents have failed. Adenoviruses, adeno-associated viruses or retroviruses can all infect non-dividing cells and facilitate gene knockdown or overexpression. However, the method is quite challenging and laborious, requiring strict biosafety controls. Therefore, we did not decide to attempt this approach for this project.

Physical methods could also possibly be applied for the transfection of FG monolayer. However, most physical methods require expensive specialist equipment. Amaxa nucleofector (Lonza) available in our lab can be used for transfection of adherent cells. However, it requires an additional specialist module that would have to be purchased at a substantial cost.

Another way to attempt FG transfection would be to try the dissociation of the FG layer and seeding FG cells at a confluency more appropriate for subsequent lipid transfection or even nucleofecting the FG cells. One of our protocols describes generation and long-term maintenance of FG stem cells in culture, that includes FG dissociation. However, it has not yet been optimized for culture on Matrigel, a cell matrix adapted by our lab for all our differentiation protocols. We have made attempts to optimize the protocol, but they were unsuccessful (data not shown). Due to time limitations, we have looked for a solution that eliminates the need to transfect FG monolayer and instead target hiPSCs that are easily dissociated. The inducible system we have selected will enable us to study gene knockdown at the desired stage of differentiation. The inducible nature of the engineered construct ensures that the knockdown studies are specific to the differentiation stage of interest. Furthermore, this system has also been published for knockout studies (Bertero et al., 2016). For overexpression studies inducible systems can also be applied (Lange et al., 2020, Kim et al., 2015).

Although transfection of iPSCs using lipid reagents showed better results than FG transfection, the proportion of targeted cells was still not optimal. Additionally, it was observed that the lipid reagents were only accessing iPSC at the edges of the colonies and a substantial proportion of tightly packed cells in the middle of the colonies were not targeted. To overcome this issue and improve transfection efficiency we employed a physical method of transfection. Nucleofection of dissociated iPSCs has given excellent transfection rate, without excess cell death. To mitigate for possible toxic effects of exposure to high voltage, we optimized transfection of high number of iPSC to maximize the number of surviving cells.

Chapter 4. BMP4 signalling mediates the specification of liver progenitor cells from anterior foregut endoderm during hiPSCs differentiation

4.1. Introduction

RNA sequencing is one of techniques that utilizes next generation high throughput sequencing for analysis of the transcriptome. It can provide detailed insight into the expression level or relative changes of each transcript during defined developmental stage or in response to various stimuli, such as activation of a signalling pathway. Over the last decade, it has largely replaced previous transcriptome analysis techniques, such as microarray, due to its higher resolution, sensitivity and capacity for new transcript discovery. (Whitley et al., 2016). It enabled the discovery of non-coding transcripts such as microRNAs or long non-coding RNAs (lncRNAs) that are involved in the regulation of gene expression, protein translation and modulation of chromatin states (Han et al., 2015). This technique can be used to examine transcriptome changes during PSCs differentiation. Using RNA sequencing, Twaroski and colleagues have identified a number of genes activated or supressed by FGF2 signalling during the early stages of LPC specification. The study identified a WNT signalling suppressor, NKD1, as one of the direct targets of FGF2 signalling (Twaroski et al., 2015). It is established that Wnt suppression is necessary for hepatic specification and this study has contributed to the understanding of molecular mechanism of how this is achieved. To strengthen the information form RNA sequencing it can be combined with another high throughput technique: chromatin immunoprecipitation followed by sequencing (ChIP). Protein-DNA interactions are vital for BMP4 induced SMAD-mediated regulation of gene expression. Identifying the genomic locations to which SMAD1/5/8 bind can contribute to our understanding of how BMP4 signalling mediated the acquisition of hepatic fate in FE cells. Chromatin immunoprecipitation is a popular technique that allows the detection of protein-DNA interaction in the

context of living cells. It has been applied to identifying the locations of binding sites of TFs, histones and other proteins (Furey, 2012). The technique involves reversibly crosslinking DNA to its interacting proteins in living cells. The cross-linking was initially achieved using the UV light, now largely replaced with formaldehyde. Once the crosslinking is done, the cells are lysed, and the chromatin (DNA with associated histones and proteins) is fragmented by sonication or nuclease digestion. An antibody against a protein of interest is used to capture the DNA fragments the protein interacts with. Antibody-protein complexes are isolated using magnetic beads and DNA is freed from the associated protein by reversing the process of crosslinking. The retrieved DNA can be analysed by qPCR, chip microarray or next generation sequencing (Figure 4-1) (Mundade et al., 2014). As the cost of next generation sequencing decreased, ChIP-seq. has become the most popular method for DNA fragments analysis. Sequencing offers higher resolution (single base pair level), expanded coverage and better sensitivity compared to qPCR or microarray. ChIP seq. can help identify binding sites of TFs, motif sequences to which they bind and help understand mechanisms involved in differential gene expression (Lu et al., 2013).

Currently, no direct targets of BMP4 signalling in the context of hepatic specification have been identified. The molecular mechanism by which this pathway induces hepatic identity in the FE is unknown. To understand the genes activated upon this pathway we employ RNA sequencing in the early stages of LPC specification. Additionally, to be able to identify direct targets of BMP4 signalling we also perform chromatin immunoprecipitation (ChIP) with an antibody against SMAD1/5/8 followed by sequencing.

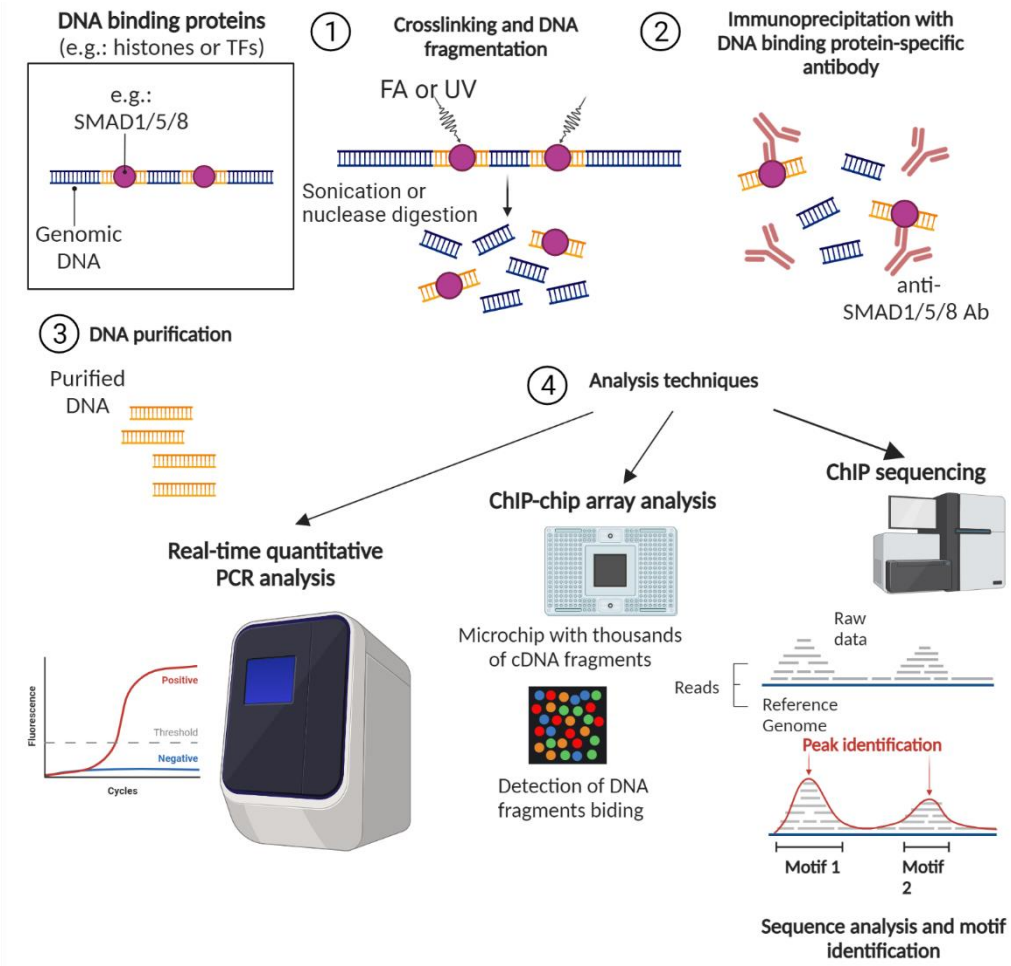

Figure 4-1 Overview of chromatin immunoprecipitation technique.
Image created using Biorender templates with modifications.

4.2. Chapter aims and objectives

In this chapter, we apply our differentiation platform to study the effects of BMP4 signalling during FE specification towards the hepatic lineage. We employ RNA sequencing to examine wider transcriptional changes induced by this pathway followed by ChIP sequencing that will allow more specific identification of direct BMP4 signalling targets. The objectives of this chapter are:

- BMP4 signalling is necessary for LPC specification from FE during hiPSCs differentiation

- BMP4 signalling induced transcriptomic changes in FE cells during LPC specification
- SMAD1/5/8 bind to promoters of BMP4 upregulated genes during LPC specification

4.3. Results

4.3.1. BMP4 signalling is necessary for LPC specification from foregut endoderm during hiPSC differentiation.

BMPs secreted by STM have been shown to be essential for liver specification in mice (Rossi et al., 2001), zebrafish (Shin et al., 2007) and chicks (Zhang et al., 2004). Therefore, many protocols for generating hepatocyte-like cells from hiPSC involve addition of BMP signalling, along with FGF, during the specification of LPCs from FE (Cai et al., 2007, Si-Tayeb et al., 2010b, Hannan et al., 2013b). However, the essential role of BMP signalling for hepatic specification during hiPSC differentiation has not been shown. We identified two different inhibitors of BMP4 signalling: Noggin (NOG) and Dorsomorphin (DOR). Inhibition of BMP4 signalling rather than simple omission from the media was in our experience necessary as different hiPSCs respond to different concentrations of BMP4 during LPC specification (unpublished data). This may be caused by different sensitivity of the cell lines to the signalling or by endogenous secretion of BMP molecules by FE cells. The right BMP4 concentration for differentiation R-PAT cell line has already been determined by our lab. To prevent the endogenous secretion from confounding our analysis, BMP4 signalling had to be blocked. As per our protocol, R-PAT M cells were first differentiated towards DE using Act A and WNT3a signalling, followed by specification towards FE using Activin A only. The FE cells were then induced towards LPC fate by addition of FGF10, SB431542 with either BMP4 or one of the inhibitors. The inhibitors were added for an extended period of time, from first day of specification only (+1) to all four days (+4). Light microscopy and real-time qPCR were used to assess the consequences of BMP4 inhibition on LPC specification. Figure 4-2a shows light microscopy images of LPC day 4 differentiated in the presence or inhibition of BMP4. When NOG or DOR are added during first day of the differentiation only, the morphological differences are difficult to notice. However, addition of inhibitors during all four days of specification impacts the morphology of the LPC monolayer. When BMP4 is inhibited, the monolayer appears less compacted and the cells resemble the cells of FE. The expression of major markers of

hepatic specification is drastically affected (Figure 4-2). The mRNA levels of one of the major TFs involved in hepatic specification, HNF4α, are significantly reduced once the BMP4 inhibition lasted more than one day. Other markers of hepatic specification, AFP, ALB or TTR, are significantly reduced even when BMP4 inhibition lasted for one day only. When the BMP4 inhibition lasted for 4 days, there is virtually no mRNA expression of HNF4α, AFP or ALB, and TTR levels are also severely reduced compared to LPC specification in the presence of BMP4. These results show that BMP4 signalling acts upstream of HNF4α and other LPC markers and indicated that BMP4 signalling is essential for the differentiation of LPC from hiPSC-derived endoderm.

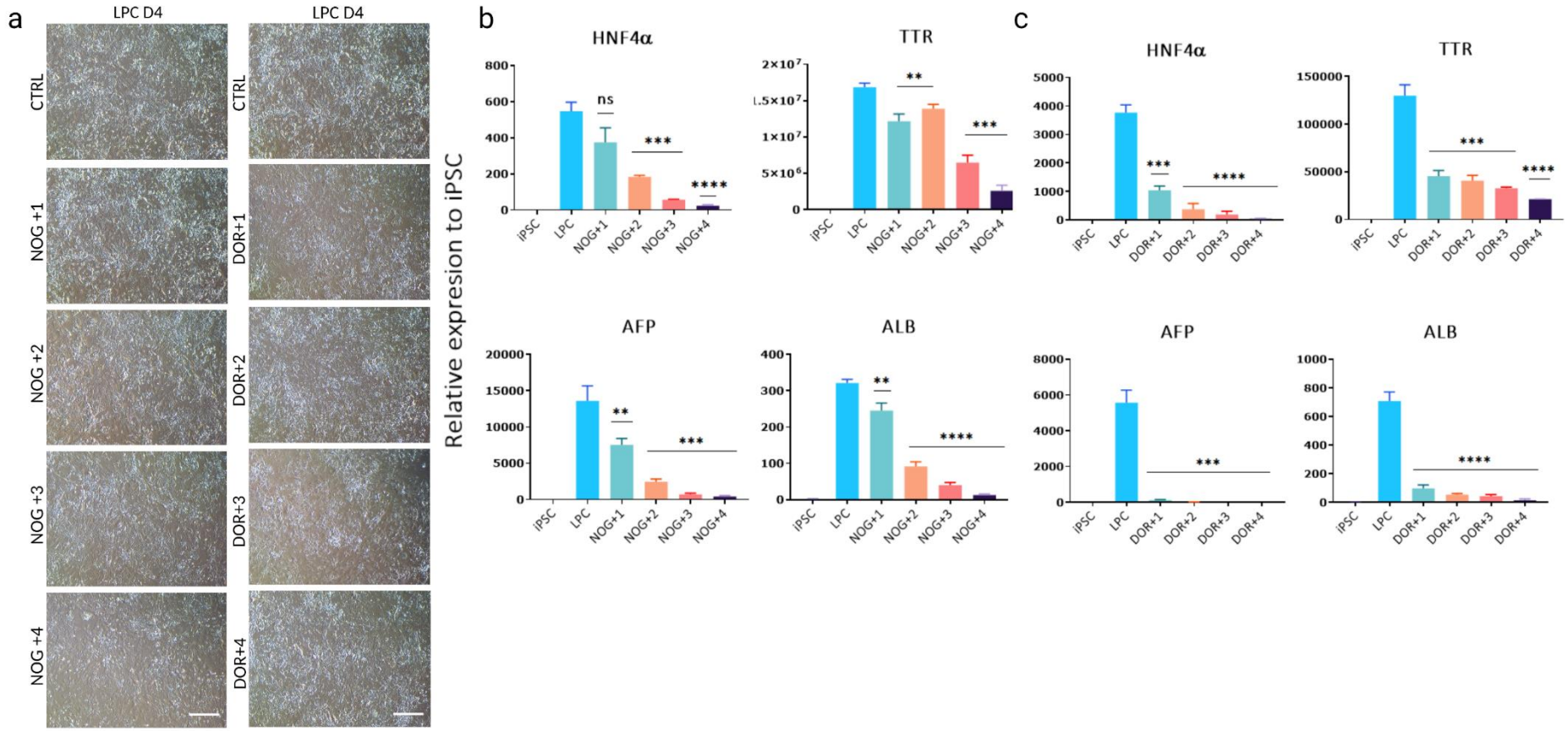

Figure 4-2 Effects of BMP4 inhibition during LPC specification of hiPSCs.

a) Light microscopy images of LPC day 4. Scale bar = 200μm; b) and c) qPCR analysis of mRNA levels of early liver specific genes: Values presented as mean ± SD; Unpaired t-test, n=1.

4.3.2. RNA sequencing and ChIP sequencing of LPC samples differentiated in the presence or absence of BMP4 signalling.

To gain a better understanding of how BMP4 signalling induced the hepatic fate in FE cells we performed transcriptome analysis of LPC cells differentiated in the presence or absence of BMP4 signalling. Additionally, to identify direct targets of SMAD1/5/8 effectors of this signalling, ChIP-sequencing was performed. R-PAT M cells were differentiated to LPC in the presence (BMP) or absence (NOG) of BMP4 signalling. Samples for RNA were collected from undifferentiated samples and at two time points of LPC specification: 24h and 48h to identify the early genes controlled by BMP4. For RNA sequencing, BMP4 was blocked with NOG. To capture changes in both protein coding genes as well as non-coding ones, ribosomal RNA depletion was performed instead of polyA+ selection. Strand specific RNA sequencing libraries were prepared to enable identification of sense/antisense and overlapping transcripts. Libraries were sent off for sequencing to Babraham Institute, Cambridge. Sample were read at 30 x 10^6/sample sequencing depth. Internal quality control confirmed good quality of sequencing experiment. On average, 93.57% ± 0.43% of the reads aligned uniquely to the GRCh38_v97 Ensembl human genome.

For ChIP-sequencing experiment, due to technical problems with Noggin protein, BMP4 signalling was blocked with small molecule dorsomorphin (DOR). Although the difference in experimental set up between RNA- and ChIP-sequencing was not ideal, the inhibition of hepatic specification is equally efficient using both inhibitors as presented in 4.3.1. ChIP crosslinking was done using formaldehyde to capture the direct interaction of DNA and proteins. Due to time limitations, we opted for a ready-made kit for ChIP preparation with MNase digestion for DNA fragmentation instead of more common sonication. MNase digestion condition required some optimization. Typically, 1.5-2.5µl of diluted MNase enzyme are used for 4 x 10^6 cells, as per protocol. We tested three different amounts of the MNase enzyme on our samples and checked the digestion of the chromatin by agarose gel electrophoresis. 3µl of the enzyme produced most clear bands of the expected lengths: 160bp, 320bp and 480bp

which corresponds to the 1, 2 and 3 nucleosome units (Figure 4-3a). Input produced from the 3µl sample was further tested following DNA purification to ensure good fragment distribution before library preparation (Figure 4-3b). The 4200 Tapestation identified bands that roughly correspond to the nucleosome units. Library preparations were outsourced, and post library preparation check was also performed. Figure 4-3c shows an example DNA library of one of the samples with majority of fragments forming a peak around the 300bp ladder mark. This corresponds to single nucleosome unit fragment with added 120bp adapter. Following library preparation samples were sequenced and output was quality checked by bioinformaticians at DeepSeq, Nottingham. The sequencing passed the quality assessment with high mean quality scores (>30), normal distribution of GC content (39%±0.00455%) and very low sequence duplication levels (3.31% ±0.002728%). Initial analysis identified some peaks but could not find any differentially expressed ones. The data was passed on to bioinformatics department at Babraham for further analysis by Dr Simon Andrews. Overall look at the distribution of peaks failed to detect enrichment between ChIP and Input samples. A QQ plot comparing all the samples also revealed minimal differences between all sequenced samples (Figure 4-4a). Pairwise comparison revealed some differentially expressed (DE) hits (Figure 4-4b) but a closer look using a genome browser showed that the enriched peaks grouped near holes in the assembly in the centromeric regions of chromosomes (Figure 4-4c). These holes are usually caused by highly repetitive sequences. Further analysis showed that almost all (83%) of the enriched regions lie within 50kb of the assembly hole as opposed to a random sample where only 9% were found within this distance (Figure 4-4d). This result, together with the overall lack of enrichment, strongly suggest that the ChIP experiment has failed.

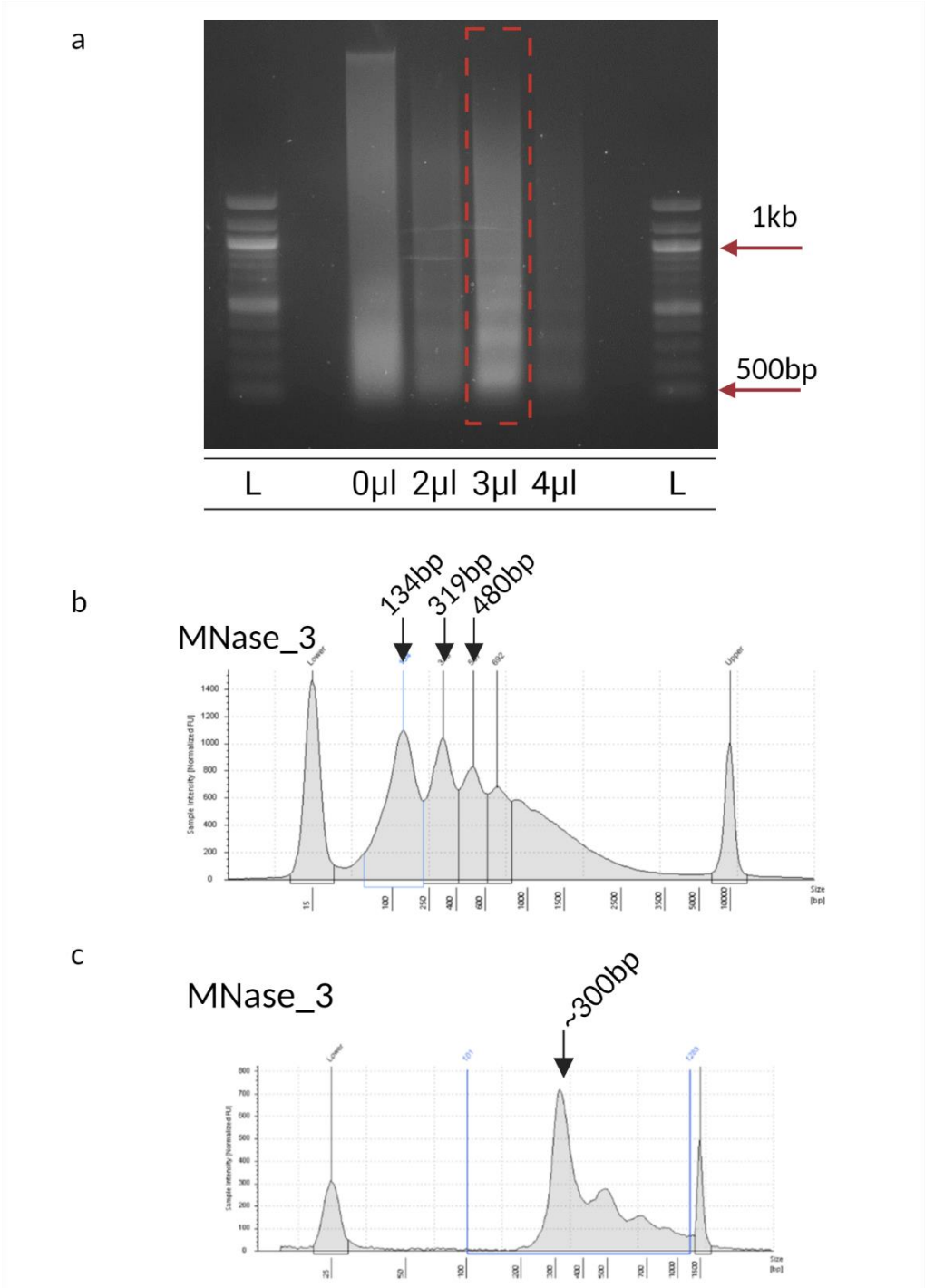

Figure 4-3 ChIP optimization.

a) Chromatin electrophoresis following MNase digestion with varying number of µls of the diluted enzyme. Samples were run on 1% agarose gel. L: 100bp ladder; b) Input sample run on 4200 Tapestation System showing fragmentation of Input DNA when digested with 3µl of diluted MNase; c) DNA fragments distribution following library preparation.

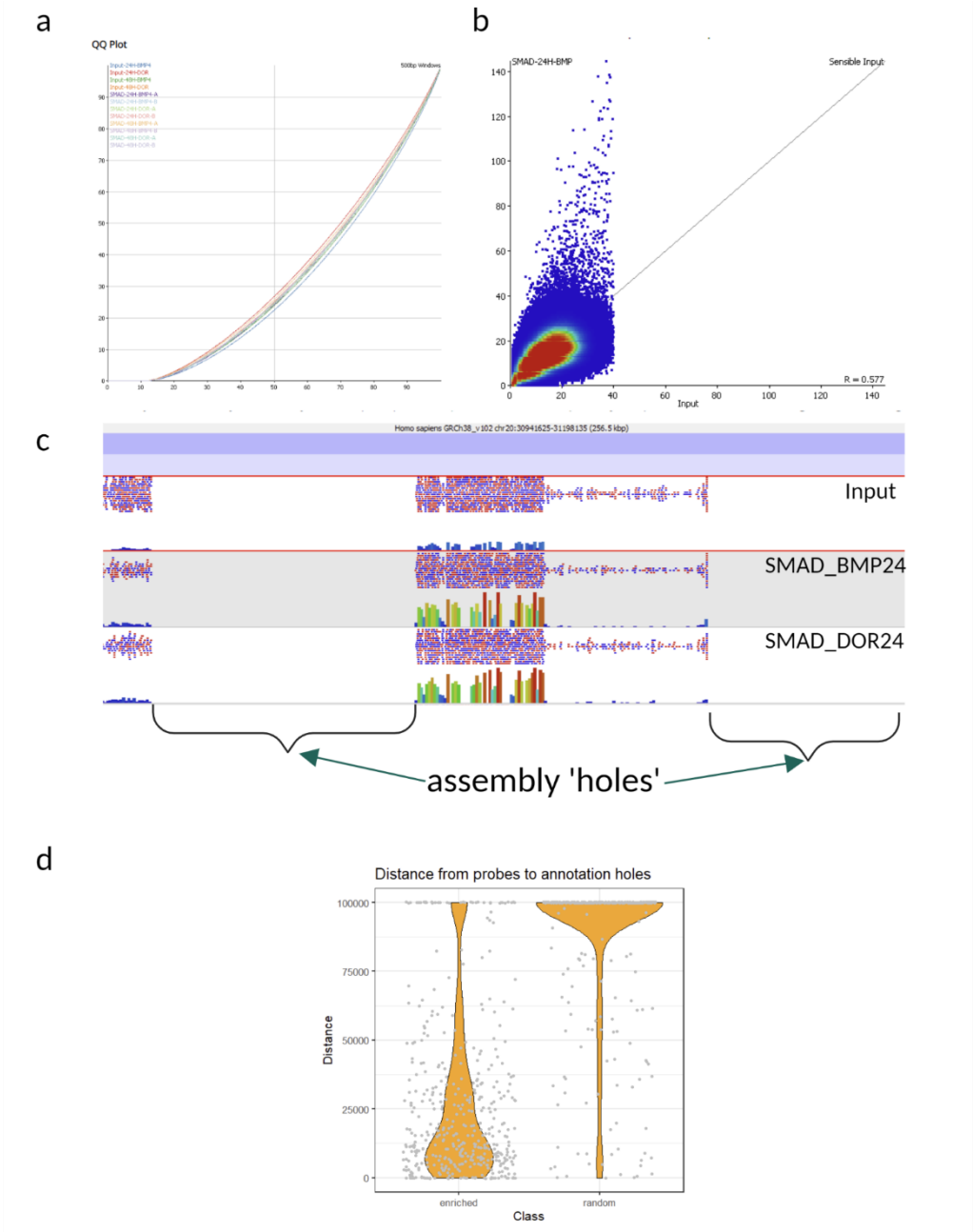

Figure 4-4 ChIP-sequencing data check.
a) QQ plot comparing the distribution of peaks in all the samples; b) Scatter plots visualising some differentially expressed peaks in the 24h samples; c) SeqMonk software visualisation of some of the top differentially expressed hit showing their grouping and proximity to 'assembly holes'; d) Violin plot showing the distance of SMAD enriched regions in our samples from assembly holes compared to a random selection of regions.

4.3.3. Transcriptome changes induced by BMP4 signalling

The main aim of this thesis is to understand the molecular mechanism by which BMP4 induces gene expression changes in the FE to specify the cells towards LPC fate during hiPSCs differentiation. To identify the early gene expression changes induced in FE cells by BMP4 signalling, we differentiated R-PAT M cells

toward LPCs in the presence (BMP) or absence (NOG) of BMP4 signalling. RNA was isolated from undifferentiated R-PAT M cells as a reference, and from two early timepoints of LPC specification: 24h and 48h, and transcriptome analysis was executed. Principle component analysis (PCA) was performed to examine the variation between the different sample groups (Figure 4-5a). The most variation was due to the transcriptome differences between undifferentiated samples and differentiated ones. R-PAT M samples clustered together and away from the LPC samples. This clear distinction in the transcriptome can be expected from two different cell types. The next feature that introduced most variation in the data was the treatment of FE samples with BMP vs NOG. BMP samples clustered together and away from the NOG samples, showing that BMP4 signalling induces changes in the transcriptome of the differentiating FE cells. Although triplicates of both NOG samples show a greater variance between each other, they still cluster together and are easily distinguished on the correlation tree (Figure 4-5b).

Differentially expressed genes (DExGs) between all samples were identified using DESeq2 function on SeqMonk Mapped Sequence Analysis Tool. DExGs were filtered based on $\log_2$FoldChange > 1 and FDR < 0.05 (Figure 4-6a). The heatmap of DExGs shows that undifferentiated R-PAT samples show clearly different gene expression pattern compared to both BMP and NOG samples at both timepoints. In BMP samples a cluster of genes whose expression is upregulated compared to NOG samples and, additionally, becomes more upregulated in the 48h sample can be identified. Similar observation cannot be made about NOG samples, the gene expression pattern at both timepoints appears quite similar. Closer look at specific gene groups associated with various stages of hiPSCs differentiation show, as expected, expression of pluripotency markers in R-PAT group, and down regulation of those markers in the BMP and NOG samples. One unexpected exemption is expression of NANAOG in the NOG48 sample. Endodermal genes are down regulated in R-PAT cells and upregulated in both BMP and NOG samples. GATA4 and GATA6 are increased in the BMP samples indicating the continued role of these TFs in hepatic

specification. SOX17 and CXCR4 definitive endoderm markers are downregulated in the BMP samples.

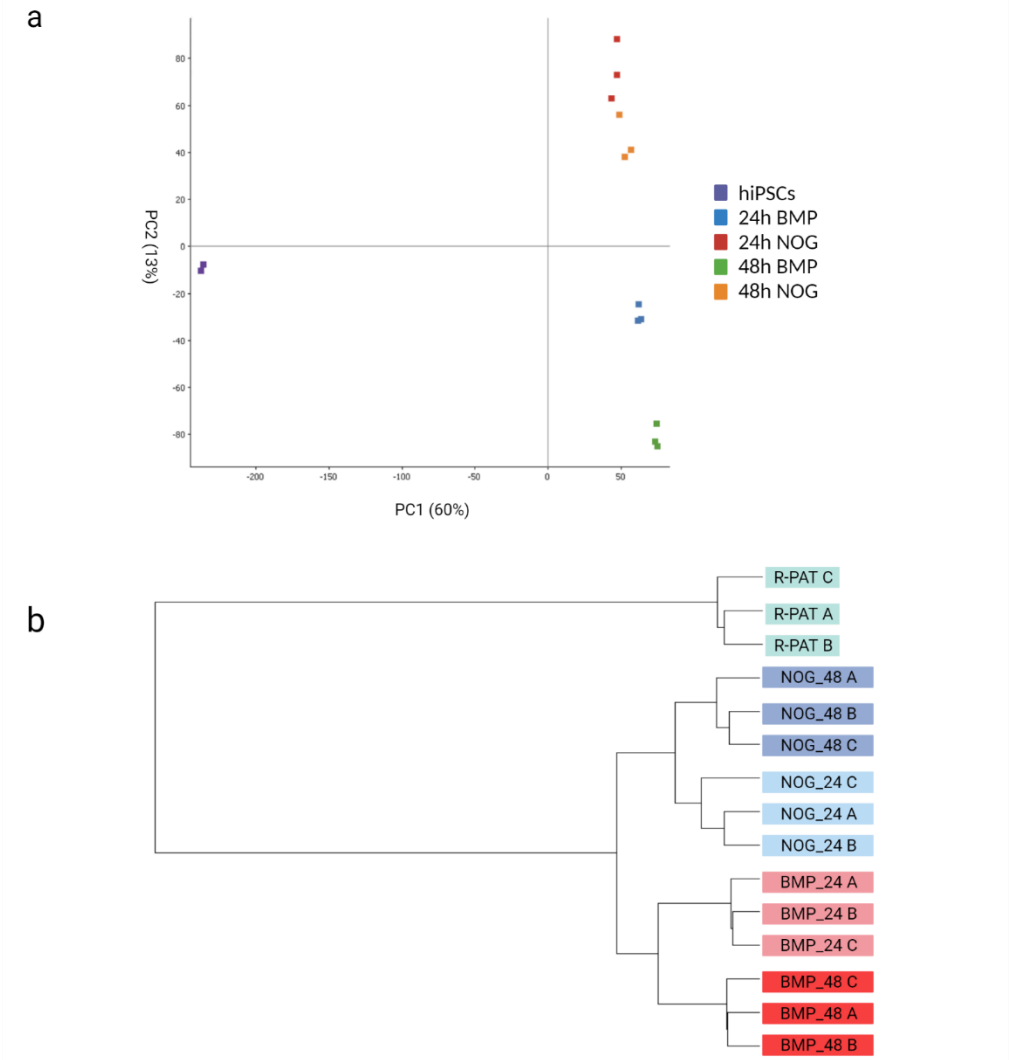

Figure 4-5 Global overview of the RNA sequencing data.
a) PCA plot demonstrated the relationship between the undifferentiated R-PAT M samples and samples differentiated to LPC in the presence (BMP) or absence (NOG) of BMP4 signalling; b) Correlation tree presents the relationship between samples and the triplicates. No outliers were identified;

Similar suppression can be seen in the NOG samples although delayed. HHEX shows clear downregulation in the BMP samples and upregulation in NOG ones. Liver specific genes such as HNF4α, HNF1α, ALB, TTR, AFP and SERPINA1 are strongly upregulated in BMP48 sample, day 2 of LCP specification. There are also

some biliary genes upregulated in the BMP48 sample, such as KRT19 and KRT7, HNF1β or CFTR consistent with the bipotential nature of the LPCs.

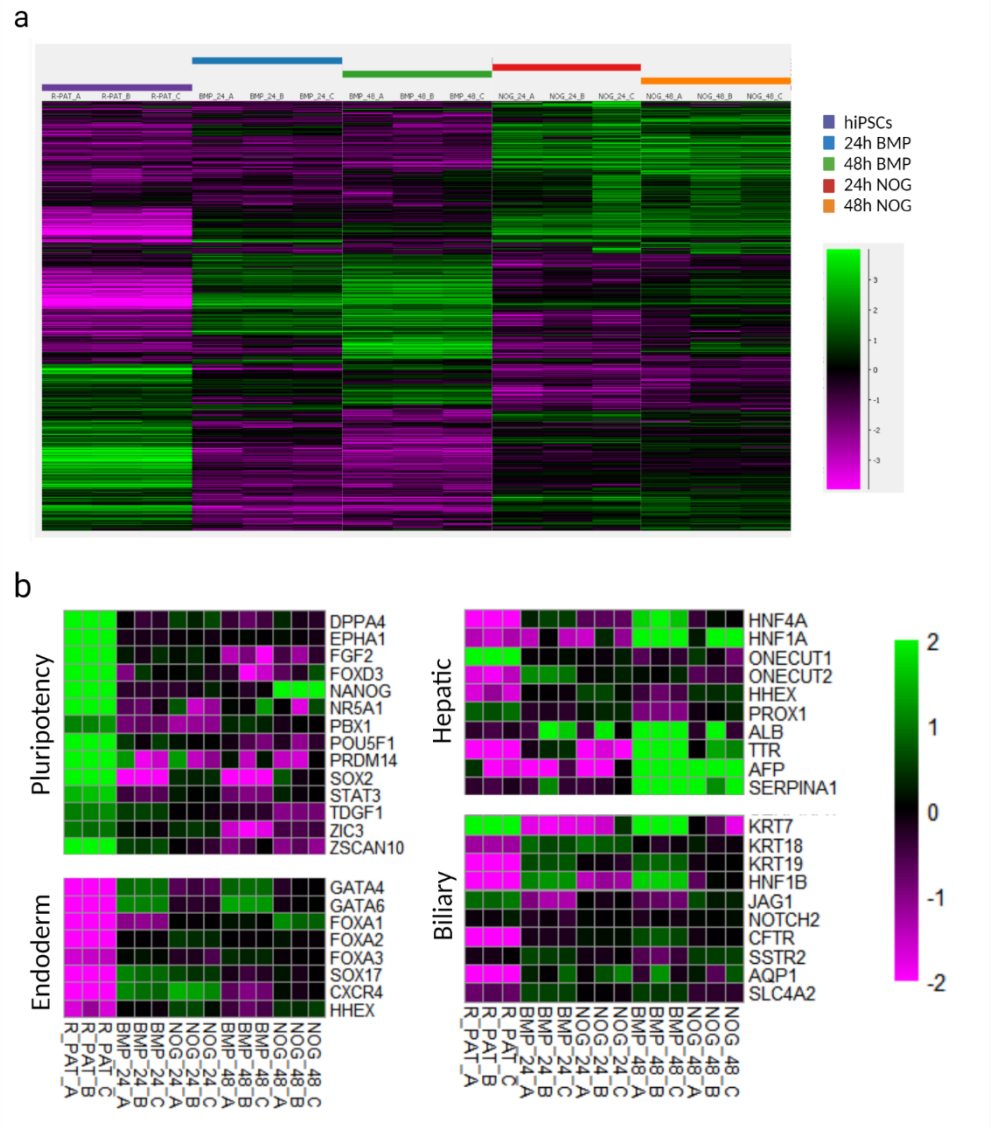

Figure 4-6 DExGs between all sample groups.
a) RNA sequencing gene expression heatmap showing all DEx transcripts between the three sample groups; b) RNA sequencing gene expression heatmaps showing changes in pluripotency, endodermal and hepato-biliary genes between the three sample groups. Log_2RPM values were median centred for normalization. Up-regulation is represented by green shading; down-regulation is represented by magenta shading.

To understand how BMP4 signalling affects the transcriptome of differentiating LPCs, we looked at differentially expressed genes between BMP and NOGs at both timepoints. Altogether, there were 26,293 and 26,513 DEx transcripts

between BMP and NOG at 24h and 48h, respectively. After applying filtering criteria (log_2FoldChange > 1, and FDR<0.05) 1721 and 2336 DEx transcripts were identified at 24h and 48h, respectively. We have identified transcript classes of interest within the DExGs: protein coding genes (PCGs) and long non-coding RNAs We also annotated DEx TFs within the PCGs. Figure 4-7 details the number of upregulated and downregulated genes in BMP samples.

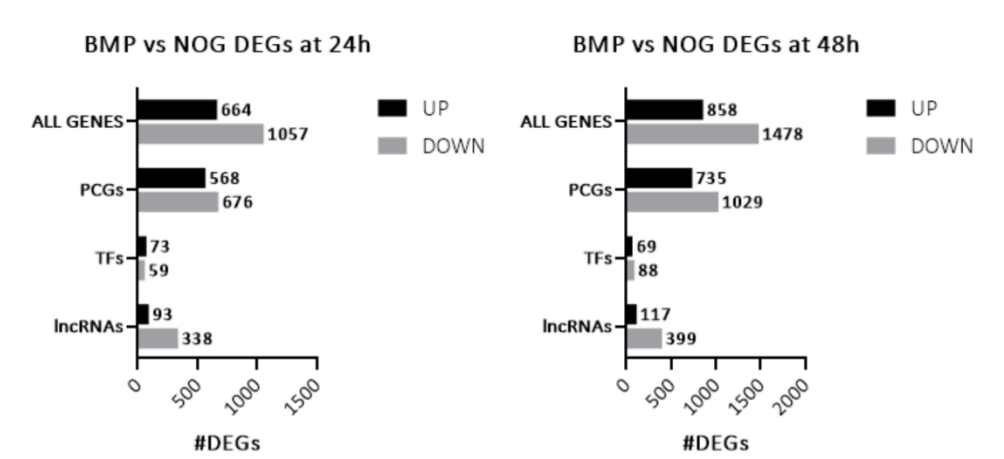

Figure 4-7 Analysis of DExGs between BMP and NOG samples.
Number of DEx genes between BMP and NOG samples. DExGs were identified using DESeq2 and filtered on the basis of log_2foldchange >1 and FDR<0.05;

Further, to gain a general idea of how BMP4 exerts its effects on FE cells, overrepresentation analysis (ORA) was performed on differentially expressed PCGs focusing on pathway analysis and possible biological functions. ORA analysis was performed using WebGestalt (http://www.webgestalt.org/) online tool. Stricter criteria for inclusion in ORA analysis were applied (FDR <0.05 and log_2FC >2) due to limitations of the software. This reduced the number of BMP24 and BMP48 upregulated genes to 208 and 283, respectively, and the number of NOG24 and NOG48 upregulated genes to 275 and 431, respectively.

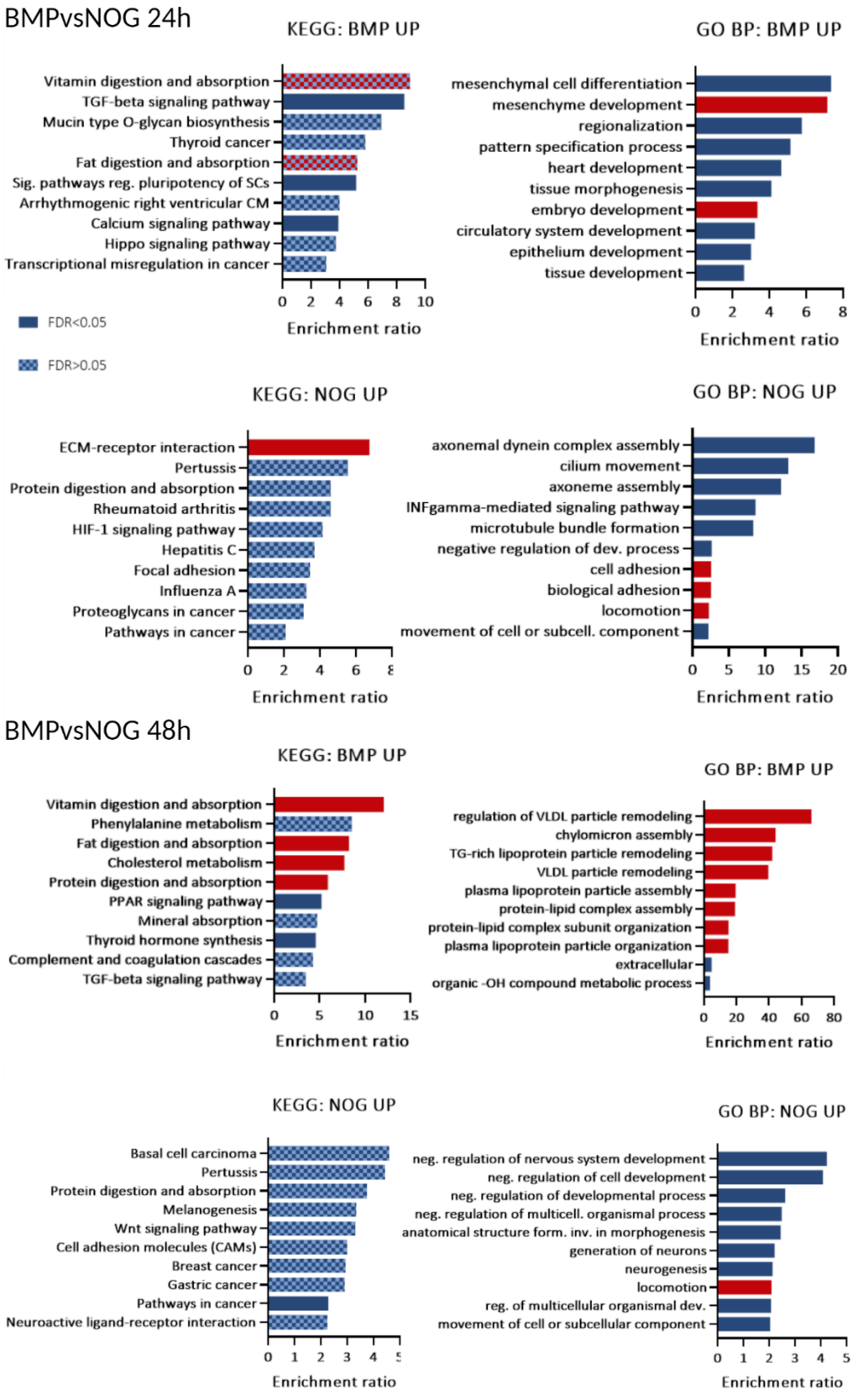

Figure 4-8 Overrepresentation analysis (ORA) of differentially expressed PCGs in BMPvsNOG samples at24h and 48h of LPC specification.

Top 10 enriched KEGG and BP pathways upregulated in BMP and NOG samples are presented. DEx PCGs with $\log_2$FC>2 and FDR<0.05 were included in the analysis. Pathway analysis performed using WebGestalt. Pathways of interest are highlighted in red.

Gene ontology of biological processes (GO BP) analysis of genes upregulated in BMP samples at 24h identified mainly terms associated with generic development such as 'regionalization', 'pattern specification', 'tissue morphogenesis' and 'tissue development' (Figure 4-8). Surprisingly, it also identified gene groups associated with cardiac development such as 'heart development' and 'circulatory system development'. The situation is much more focused in the 48h samples where almost all the top ten most upregulated terms are associated with the processing and assembly of lipid particles ('regulation of VLDL particle remodelling', 'chylomicron assembly' or 'TG-rich lipoprotein particle remodelling'). In terms of pathway enrichment, KEGG analysis of BMP24 upregulated genes identified only three significant pathways among the top then enriched ones ('TGFβ signalling pathway', 'signalling pathways regulating pluripotency of stem cells' and 'calcium signalling pathways'). The analysis also detected pathways associated with 'vitamin digestion and absorption' and 'fat digestion and absorption' however, the enrichment is not significant at 24h. This changes in the 48h BMP samples where the analysis of upregulated genes identified pathways involved in metabolism of vitamins, proteins and fats ('vitamin digestion and absorption', 'fat digestion and absorption', 'protein digestion and absorption'). The KEGG analysis also identified PPAR signalling pathway as significantly enriched at 48h in the BMP samples.

The genes upregulated in NOG samples have also been analysed. GO BP at 24h identified terms such as 'cell adhesion', 'biological adhesion' or 'locomotion' as overrepresented in the NOG upregulated genes. Some of those terms are also present in the result of GO BP analysis of the NOG48 samples. To explore this further we looked at heatmaps of 'cell adhesion' and 'locomotion' gene sets that were upregulated in NOG samples (Figure 4-9). We can see overexpression of E-cadherin (CDH1), a gene that is upregulated in Prox1 null mice. This phenotype is characterized by inability of LPCs to migrate out of the liver bud (Sosa-Pineda et al., 2000).Other genes, such as claudins or ICAM1, although not reported to be involved in hepatocyte migration, are known to have such a role in other

tissues and cancer. The GO BP also identified terms associated with the cytoskeleton such as ‘axoneme assembly’ or ‘microtubule bundle formation’. GO terms enriched at 48h NOG samples are mainly linked to a negative regulation of aspects of development, and unexpectedly, nervous system development (‘generation of neurons’, ‘neurogenesis’). KEGG pathway analysis identified only one significantly enriched pathway at each timepoint in the NOG upregulated genes: ‘ECM-receptor interaction’ at 24h and ‘pathways in cancer’ at 48h.

Gene Set Enrichment Analysis (GSEA) was also performed using GSEA v4.3.2 software (Figure 4-10). First, transcriptome checks against C8 (‘cell type signature gene sets’) category revealed enrichment of genes associated with foetal liver hepatoblasts, at both BMP sample timepoints, confirming the correct progression of our differentiation. Check against C1 (‘hallmark gene sets’) showed an enrichment of many gene associated with hepatocyte function in BMP24 and BMP48 samples. Apart from already identified categories in ORA linked to fat and protein metabolism, GSEA also showed upregulation of genes important for coagulation, haem metabolism and glycolysis. Their enrichment can already be detected at 24 hours of LPC specification. although it is not statistically significant. However, at 48h the enrichment increases and it becomes statistically significant. KEGG pathway analysis confirmed upregulated expression of cell adhesion molecules and ECM-receptor interactions in NOG24 and NOG 48 samples.

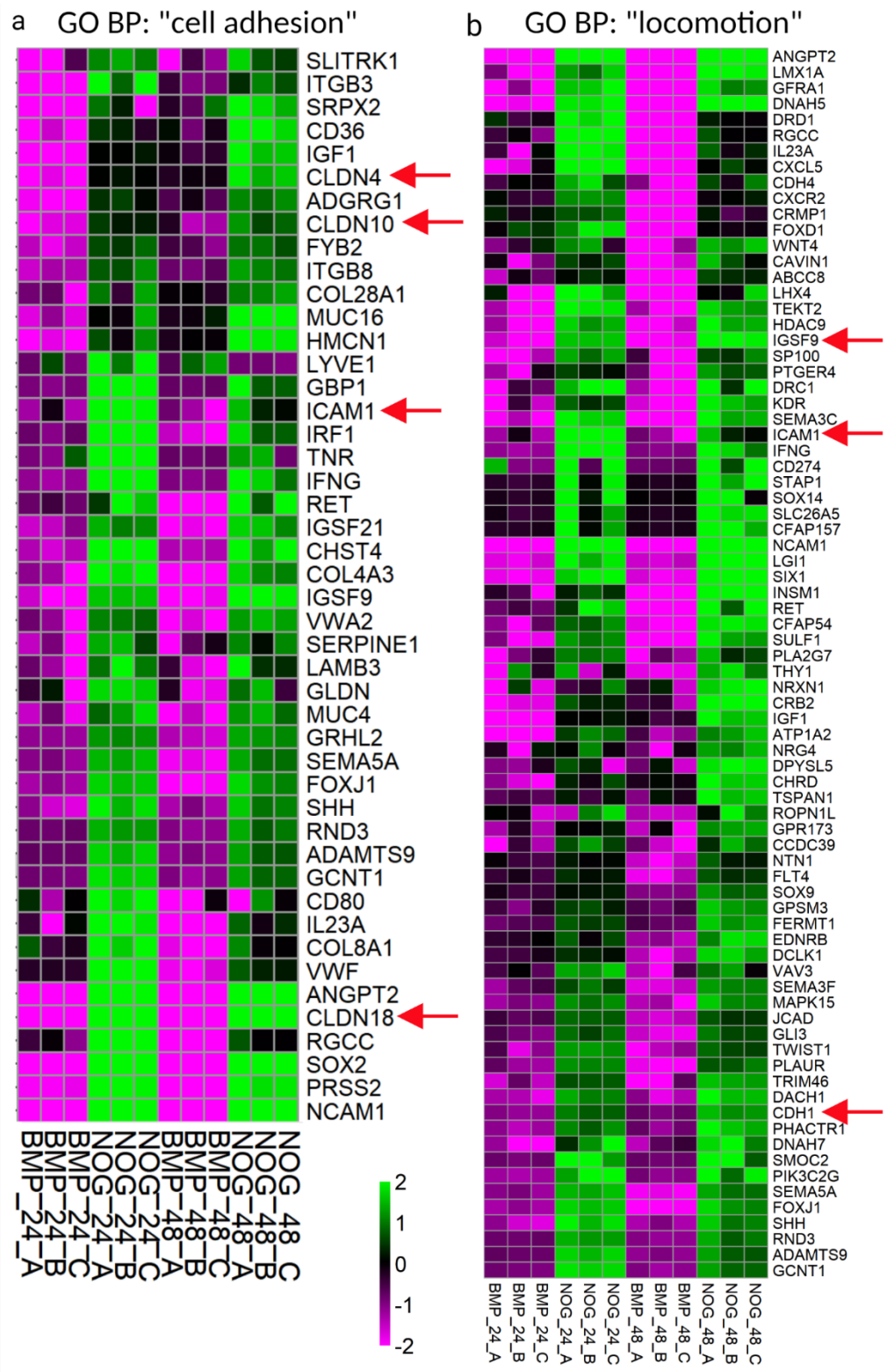

Figure 4-9 Heatmaps representing selected enriched GO terms of genes upregulated in NOG samples.

a) Heatmap of NOG24 upregulated 'cell adhesion' associated genes; b) Heatmap of NOG48 upregulated 'locomotion' genes. Red arrows point to genes that will be discussed.

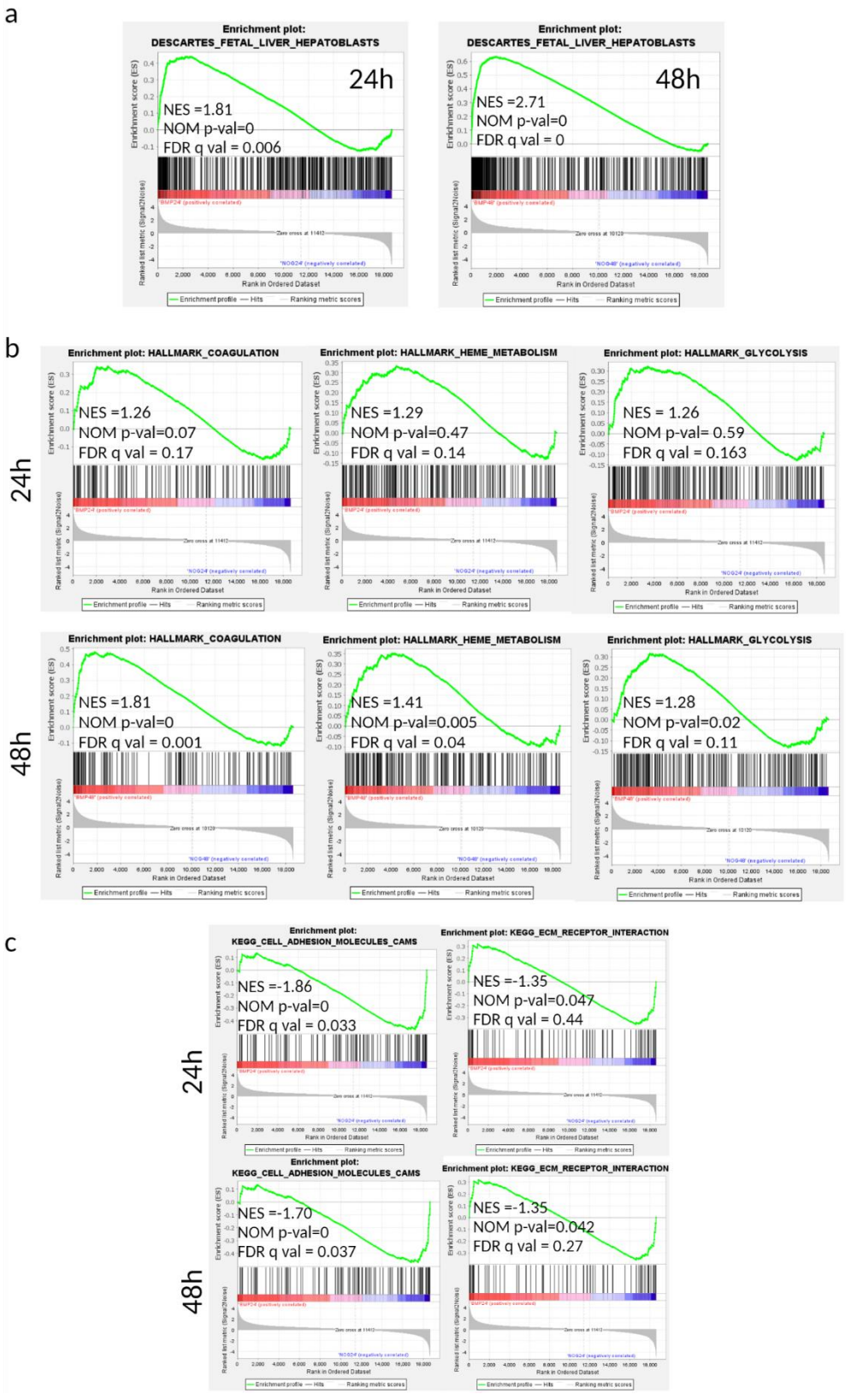

Figure 4-10 GSEA graphs showing enrichment of genes from different categories.
a) C8 'cell type signature genes sets' at 24h and 48h; b) Hallmark category enriched at 24 and 48h; c) C2 'curated gene sets' category at 24h and 48h.

Taken together, the ORA and GSEA revealed BMP4 regulated early expression of genes associated with hepatocyte function. It also suggest that BMP4 signalling regulates the expression of cell adhesion molecules and ECM-receptor interactions, downregulating their expression possibly in preparation for LPC migration out of the hepatic bud.

We were also interested if BMP4 may regulate the other pathways involved in LPC specification: FGF or Wnt. Both BMP and FGF signalling is necessary for liver gene induction in mice, chicks and zebrafish and the removal of either of the signals prevents hepatic specification (Rossi et al., 2001, Zhang et al., 2004, Shin et al., 2007). At the same time, Wnt signalling needs to be suppressed in the FE (McLin et al., 2007). BMP and FGF signalling cross talk in many other aspects of development such as cardiogenesis, bone formation or nervous system development and they can regulate each other's activity (Schliermann and Nickel, 2018). We have identified genes involved in FGF and WNT pathways using The Molecular Signatures Database on GSEA website (https://www.gsea-msigdb.org/). From the identified genes, we looked at the ones that show differential expression between BMP and NOG samples (Figure 4-11). We noticed that all four of FGF receptors are upregulated in BMP samples at both timepoints. This may indicate that BMP4 positively regulates FGF signalling during LPC specification. In the case of WNT signalling associated genes, there is upregulation of WNT inhibitors, such as NKD1 and FRZB in BMP samples, which would be in line with the need for suppression of this signalling. However, at the same time Wnt inhibitors such as WIF-1 and DKK-1 are downregulated in the BMP samples. Similarly, WNT receptors FZD2 and FZD5 are downregulated in BMP samples but FZD4 is upregulated.

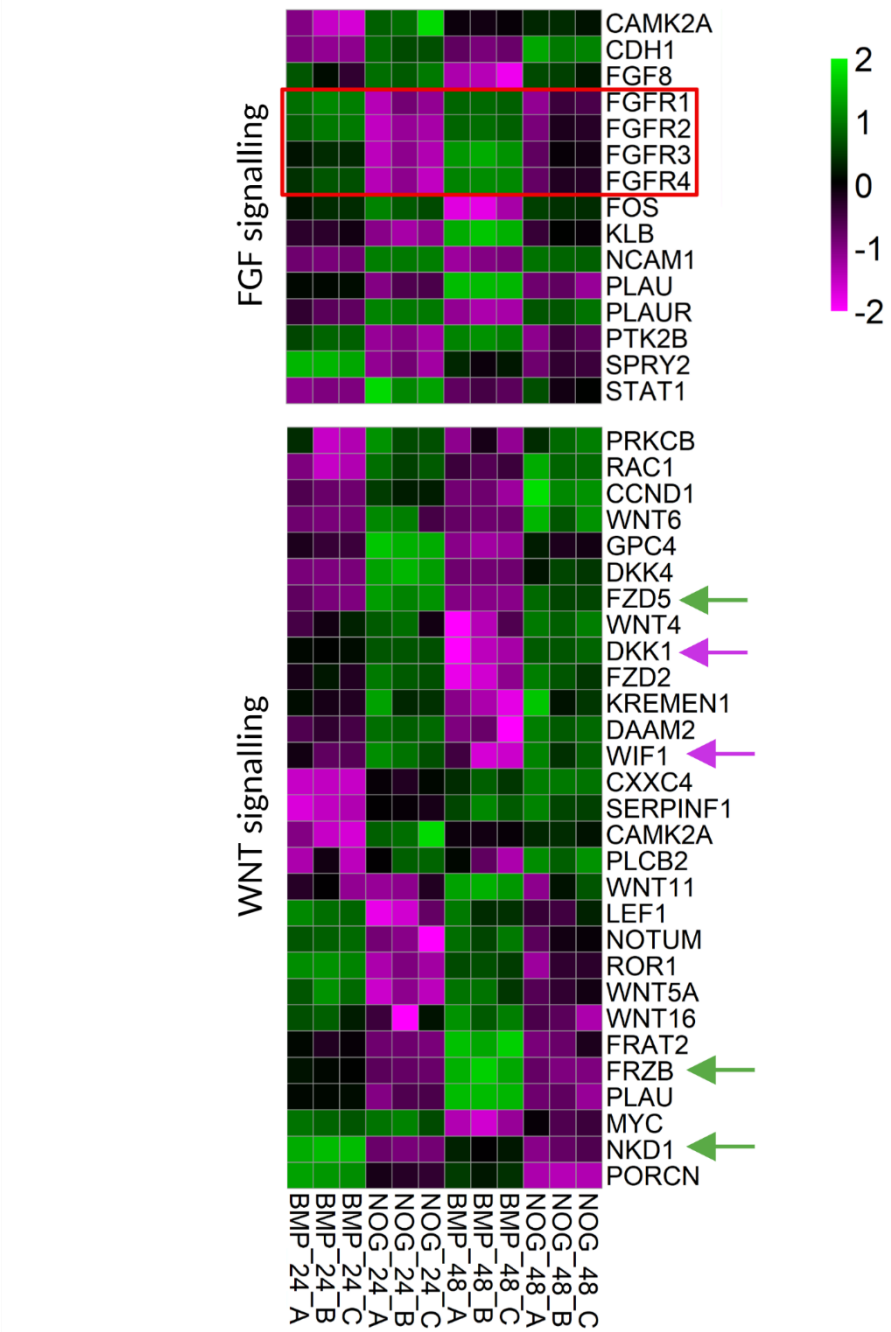

Figure 4-11 Heatmap of differentially expressed genes associated with FGF and WNT signalling.

FGF and Wnt signalling associated genes were identified using publicly available databases (www.gsea-msigdb.org). Only the genes that showed differential expression in our samples are presented in the heatmaps (FDR <0.05). Upregulation is indicated by the green shading and downregulation by red. Arrows and red box highlight genes included in the discussion.

4.3.4. Transcription factor changes induced by BMP4 signalling during LPC specification

Transcription factors (TFs) are important regulators of gene expression during cell type specification, developmental patterning and regulation of signalling. Many TFs important for liver development and function have been identified, however only very few examples of TFs controlling the earliest stage of liver specification are known to date and very little is known on how BMP4 signalling induces hepatic identity in FE cells. Our RNA sequencing identified a list of DExGs that were upregulated or downregulated at the onset of LPC specification. To identify TFs among our DExGs, we have screened them against a list of human TFs published by Lambert and colleagues (Lambert et al., 2018). Using the criteria of FDR < 0.05 and $\log_2$Fold Change > 1, we have identified 132 (73 upregulated and 59 downregulated) DEx TFs and 157 (69 upregulated and 88 downregulated) DEx TFs at 24h and 48h, respectively, in BMP vs NOG samples. This suggests that BMP4 signalling promotes cell type-specific expression pattern by regulating a number of TFs. Volcano plots show the differentially expressed TFs separately at 24h and 48h (Figure 4-12 a, b). Table 4-1 lists the DEx TFs at both time points. In the upregulated list, we can see some TFs that have already been reported to play a role in the early stages of liver development (see Table 1-1). HNF4α, HNF1β, GATA4, GATA6 and TBX3 are upregulated at 24h and 48h, while HNF1α and CEBPα show an increase at 48h. The fact that the expression of all these hepatic factors is downregulated when BMP4 signalling is blocked with Noggin suggests that BMP4 acts upstream of all those TFs to regulate their expression during LPC specification.

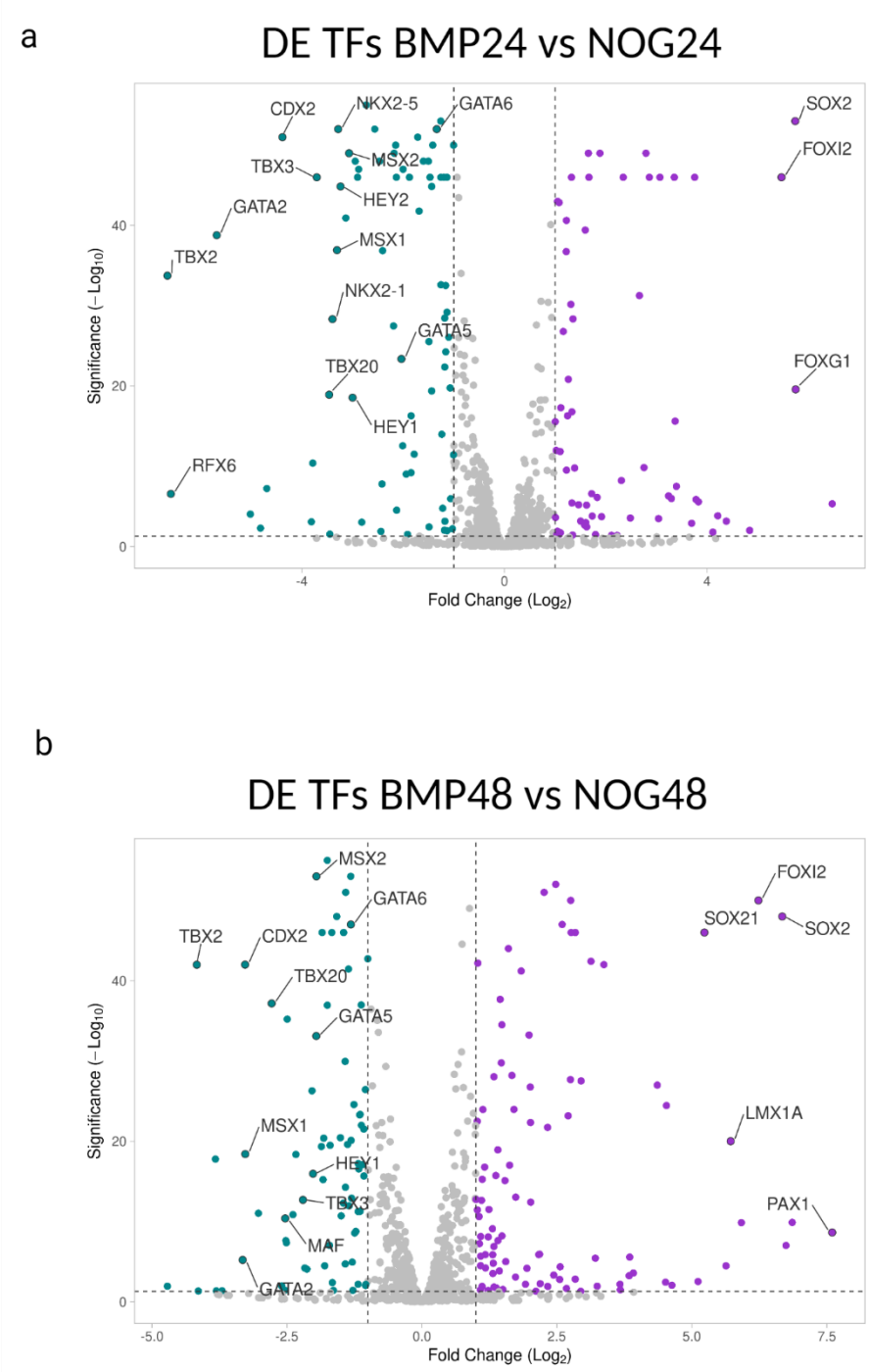

Figure 4-12 BMP4 signalling regulates the expression of TFs during LPC specification.
a) and b) Volcano plots generated using VolcaNoseR showing DEx TFs between BMP24 vs NOG24 and BMP48 vs NOG48 samples, respectively. Upregulated (cyan) and downregulated (magenta) genes are highlighted based on threshold of log2FoldChange >1 and p-value < 0.05.

24h	**UP**	TBX2, RFX6, GATA2, RXRG, IRX6, ZEB2, CDX2, HMX3, GLIS1, TBX3, TBX20, NKX2-1, MSX1, NKX2-5, HEY2, SMAD9, MSX2, HEY1, VENTX, ESRRG, NKX1-2, NKX6-2, EVX1, MXI1**, HNF1B,** TFEC, HESX1, GATA3, MAF, PITX1, MYCN, SP8, LEF1, GATA5, SNAI2, FOXQ1, FOXF2, GLIS3, MYB, HNF4G, TFCP2L1, EPAS1, ISL1, SP5, GATA4, KLF13, DMBX1, KLF8, KLF4, GRHL3, TEAD4, GATA6, ZIC2, RORA, ZIC5, SOX5, PRDM16, SKIL, HIC1, ZBED3, MITF, MKX, PITX2, ZNF792, **HNF4A,** ZNF764, ZNF560, ZFPM1, ZNF44, TSHZ1, HOXB13, ZNF799, ZNF761
	DOWN	FOXP1, HOXB3, SIX4, PAX7, SOX4, FOXA1, BHLHE22, IRF9, SOX15, CXXC4, LMX1B, SOX6, HIVEP2, ZNF521, TCF4, STAT1, OSR1, TIGD4, RUNX1, SPDEF, SNAI3, ZNF860, NFIA, GLI3, POU5F2, ZNF385D, PRDM6, XBP1, ETS2, TRERF1, BHLHA15, CREB5, IRX3, ZBED2, TWIST1, GRHL2, BATF2, DACH1, SOX3, RARB, FOXJ1, LHX4, IKZF2, ZNF157, ZBTB7C, SOX21, PAX9, IRF8, LMX1A, IRF1, ETV7, TFAP2C, SP100, SIX1, SOX14, FOXI2, SOX2, FOXG1
48h	**UP**	IRX6, TBX2, CREB3L3, GATA2, CDX2, MSX1, HESX1, TBX20, DPF3, MAF, GLIS1, **HNF1A**, EPAS1, VENTX, ESRRG, TBX3, RXRG, RFX6, SP8, EY1, GATA5, MSX2, **HNF4A**, MITF, **HNF1B**, MYCN, FOXF2, MXI1, NKX1-2, NKX2-1, **CEBPA**, ZIC2, ZEB2, GLIS3, NKX2-5, HNF4G, BCL11A, PITX2, KLF4, ZNF280A, KLF9, PITX1, OVOL1, ZIC5, ZNF114, SKIL, GATA6, KLF13, ISX, SMAD9, ZBED3, EBF4, SNAI2, ALX1, TSHZ1, GRHL3, TFCP2L1, SOX5, ZFPM2, OVOL2, FOSL2, MYCL, ZFP42, MTERF4, SNAI1, TEAD4, DBP, HLX, CREB3L2
	DOWN	HHEX, ETS1, RUNX1, SALL2, ZNF519, PROX1, ZNF233, SIX3, CARF, ZNF66, POU5F2, PLAGL1, SCMH1, HIVEP2, ZEB1, HOXB3, GSC, E2F1, SPDEF, CASZ1, MYC, MYBL1, TRERF1, ZIC3, ZMAT4, ZNF367, NFIA, FOSB, ZMAT1, TCF4, FOXC1, SOX11, IRX5, ETS2, SNAI3, SOX15, XBP1, ZNF521, EGR1, CREB5, SOX6, LMX1B, IKZF1, SIM1, SIX4, SOX9, E2F2, NKX2-2, TWIST1, OTX1, FOXO6, RARB, FOS, SP100, ZBED2, GLI3, PAX7, ZBTB7C, IRX3, OSR1, NR0B1, IKZF2, GRHL2, POU3F1, FOXI3, IRF1, DACH1, EHF, ZNF648, FOXJ1, NANOG, IRF8, ZNF157, DMRTA2, PAX9, LHX4, SOX3, FOXD1, SOX14, SOX21, ETV7, LMX1A, FOXG1, FOXI2, SOX2, SIX1, INSM1, PAX1

Table 4-1 Differentially expressed TFs at 24h and 48h of LPC specification between BMP and NOG samples.
TFs of interest are highlighted in red. TFs in bold are mentioned in the text.

In the top 20 most upregulated genes at 24h and 48h there are multiple members of the same families of TFs known to be involved in various aspects of human development. Three members of the TBX family of TFs are upregulated upon BMP4 signalling: TBX2, TBX3 and TBX20. TBX2 is the first and second most upregulated TFs at 24h and 48h, respectively. Highly on the upregulated list of

both time points is another member of the GATA family: GATA2. GATA5 is also upregulated at both timepoints, while GATA3 is increased at 24h only. GATA2 and GATA3 are mainly associated with blood development, as well as kidney, prostate, nervous system and adipose tissue development (Tremblay et al., 2018). Two members of the HEY family of TFs are in the top 20 most upregulated genes at 24h: HEY1 and HEY2. HEY1 is also upregulated at 48h. HEY proteins mainly function as transcriptional repressors, either solo or bound to co-factors or other TFs. They have been reported to be involved in the mammalian development of the heart, muscle, bone and vascular and nervous systems (Weber et al., 2014). MSX1 and MSX2 are upregulated at both timepoints. These TFs are known targets of BMP4 signalling during early tooth development in mice (Vainio et al., 1993) and have been shown to act in a redundant manner in murine cardiac development (Chen et al., 2008). They are also involved in nervous system development (Duval et al., 2014) and early limb development (Becic et al., 2018). Another TF that shows a significant upregulation upon BMP4 signalling is MAF. MAF knockout mice present with smaller cytoplasmic volume of liver cells suggesting it may be involved in the early liver development (Zhang et al., 2004). However, this TFs have mainly been reported as having a role in development of the lens, kidneys, T-cells and nervous system (Zhang and Guo, 2015).

To validate these findings, we performed quantitative gene expression analysis of a subset of DExGs, both upregulated and downregulated, during LPC differentiation from hiPSCs and confirmed differential expression of those genes (Figure 4-13). MSX2 was the only tested TFs that failed to show a significant change in expression between presence and absence of BMP4 signalling during LPC specification at both timepoints in contradiction to the RNA sequencing data.

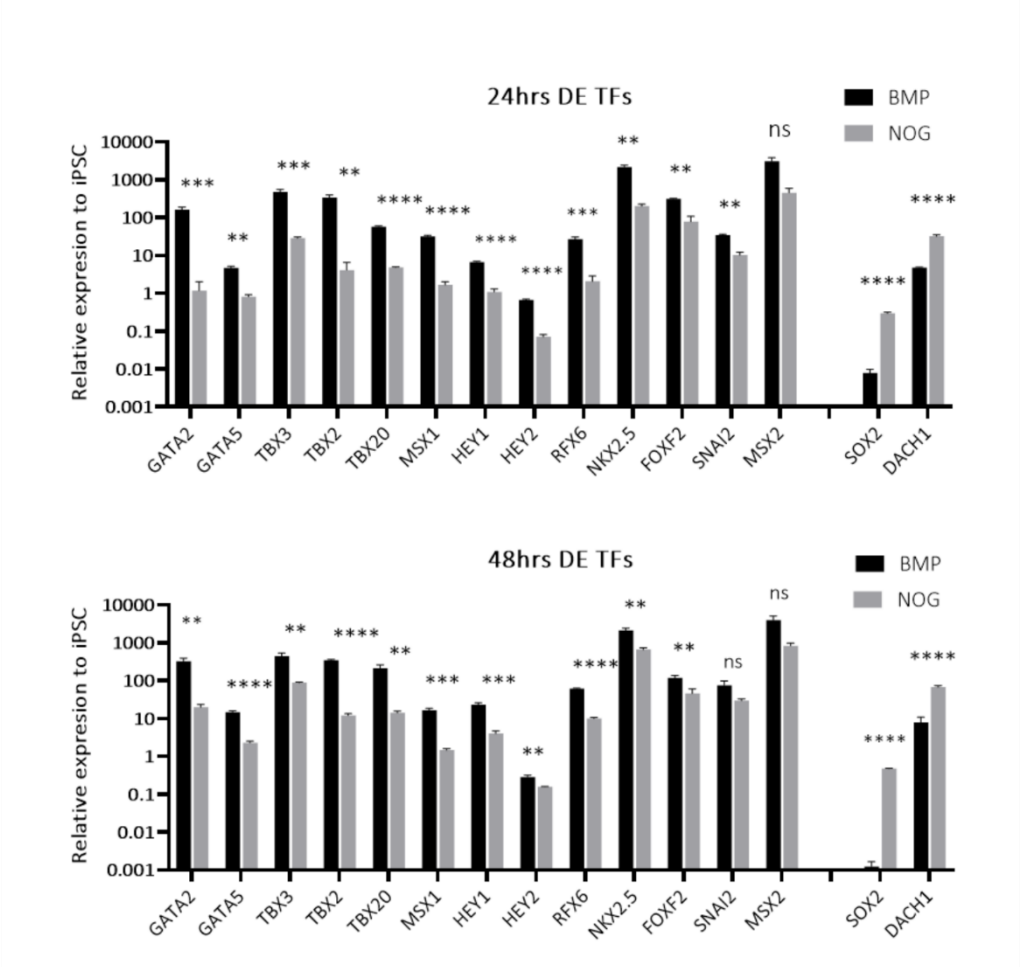

Figure 4-13 Validation of RNAseq. results by quantitative PCR.
Selected genes were chosen for validation of the RNA-seq results. Mean±SD from three technical replicates (n=1). Significance was determined with multiple unpaired t-tests in GraphPad Prism

Our RNA sequencing experiment identified a number of BMP4 upregulated TFs that have not been previously reported with a role in hepatic specification. Due to the failure of our ChIP experiment, we are not able to determine if these TFs are directly regulated by SMAD1/5/8 proteins or if their expression is induced indirectly. However, their upregulation upon BMP4 signalling and known roles in other aspects of early human development makes them interesting candidates for further study of molecular mechanisms governing LPC specification.

4.3.5. Long non-coding RNA (lncRNA) changes upon BMP4 signalling during LPC specification

Transcriptome analyses using high throughput sequencing have revealed that only around 2% of our genome codes for proteins. The majority of human transcriptome in made up of non-coding RNA of which 98% is ribosomal RNA, while the rest is divided between short non-coding RNAs (<200nt; e.g.: micro RNAs) and long non-coding RNAs (>200nt). While short non-coding RNA has been studied and characterized for over 20 years, the lncRNAs have only recently been look at in greater detail and it has become clear that this class of transcripts has important functions in gene imprinting, development, cancer and other diseases. To be able to identify lncRNAs that may be controlled by BMP4, strand specific libraries were prepared. This is particularly important for lncRNAs identification as these transcripts can overlap other genes.

The Esamble genome against which the RNA sequencing data has been aligned categorizes transcripts into different classes, one of them being lncRNAs. On that basis, we identified lncRNAs expressed in our samples and performed differential expression analysis between BMP and NOG samples at both timepoints (Figure 4-14). The heatmap of significantly (FDR < 0.05) DEx lncRNAs shows that the majority (>75%) are upregulated in the NOG samples. The reason for such skewing is not known and has not been observed with PCGs in our results.

Currently, very few lncRNAs are well described with a known function. Out of 753 DEx lncRNAs, 429 were designated 'novel transcripts', with some of them having information as to location (e.g.: intronic or intergenic), direction (sense or anti-sense) or association with a PCG. Interestingly, some of the most upregulated lncRNAs in BMP samples are running anti-sense to DEx TFs: TBX2-AS1, GATA2-AS1 or GATA3-AS1. Both TBX2 and its associated lncRNAs, TBX2-AS1 are one of the most upregulated transcripts in the BMP24 samples.

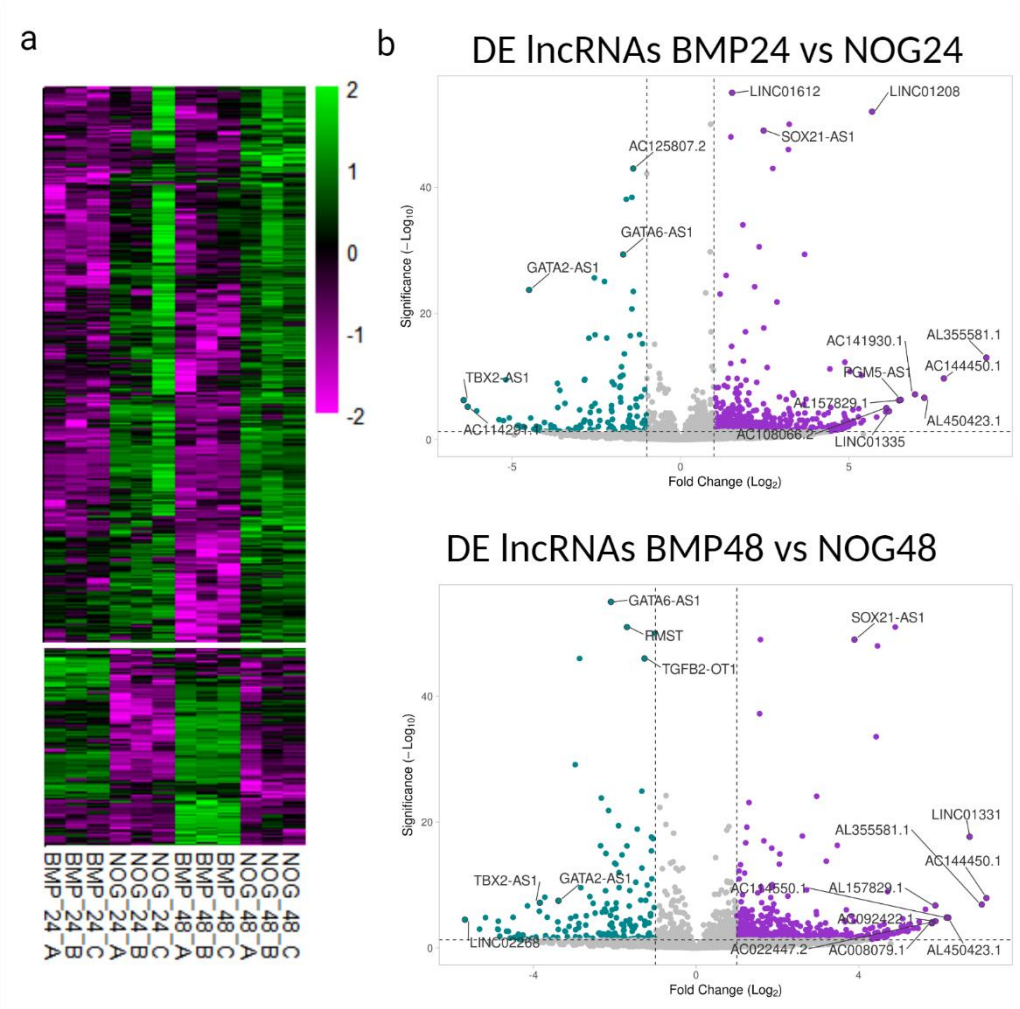

Figure 4-14 BMP4 signalling regulates the expression of lncRNAs.
a) RNA sequencing gene expression heatmap showing DEx lncRNAs between BMP and NOG samples at both time points; b) Volcano plots showing DEx lncRNAs between BMP vs NOG at 24h and 48h.Significantly changed lncRNAs with $\log_2FC > 1$ are highlighted in colour. BMP upregulated: cyan; NOG upregulated: magenta.

4.4. Discussion

In this chapter, we apply our hiPSCs based model of human development that allows reproducible and synchronous LCP differentiation, to understand the mechanism by which BMP4 signalling induces hepatic fate in the FE cells. We show that BMP4 singling inhibition during LPC specification virtually abolishes the expression of liver specific genes, such as HNF4α, ALB or TTR, at the transcript level, consistent with reports on animal models (Rossi et al., 2001). We start by examining changes in the transcriptome of FE cells exposed to BMP4 signalling. RNA sequencing provided us with an insight into the gene groups activated upon this signalling and allowed identification of DEx TFs that may be effectors of BMP4 signalling and important regulators of hepatic specification. We also attempted to identify direct targets of SMAD1/5/8 TFs by chromatin immunoprecipitation (ChIP) followed by sequencing, in order to understand which genes can be directly activated or repressed by these BMP4 signalling mediators. However, our ChIP sequencing experiment has proved unsuccessful with no peak enrichment identified in our ChIP samples compared to input. The small number of differentially expressed peaks all located near the centromeric regions of the chromosomes near assembly holes that indicate highly repetitive regions of the genome. Such peak distribution is unlikely to be reflective of TFs binding activity as TF most frequently bind DNA regions of promoters, near transcription start sites (Whitfield et al., 2012). ChIP is a technically challenging technique that can be affected by almost every aspect of sample preparation. One of the most important elements of successful ChIP experiment is the antibody directed against the protein of interest. For our ChIP experiment, ChIP grade antibody that showed previous success with immunoprecipitation was selected (Baas et al., 2016). However, quality can vary greatly between different lots of the same antibody (Egelhofer et al., 2011). It is recommended to test ChIP antibodies using flow cytometry, fluorescent microscopy or Western Blot. However, due to time limitations imposed by the COVID-19 pandemic we were unable to conduct these tests. We were also unable to optimize the antibody concentration used for each immunoprecipitation to achieve the best signal-to-

noise ratio. Ineffective antibody or wrong antibody concentration will result in inefficient selection of the chromatin and result in non-specific signal. Alternatively, epitopes targeted by the antibody can sometimes be masked in the cross-linked chromatin (an issue less likely to occur with polyclonal antibodies) or the antibody may bind cross-linked chromatin less efficiently. Considering that we selected a polyclonal, ChIP grade antibody with proven success with this technique makes these two issues unlikely. The length of formaldehyde treatment can also affect the success of this technique and may require optimization. Finally, the antibody-protein interaction can become disrupted during washing. If the antibody is tested and deemed to be sensitive and capable to bind the antigen, this issue can be overcome by reducing stringency of washes or using beads with higher affinity for the specific antibody isotype (Carey et al., 2009). Although we cannot be sure without conducting further testing, we think that the most likely reason for our experiment failure was ineffective antibody. In the future, to save costs of the sequencing, a small sample of the ChIP library can be run on the sequencing analyser and bioinformatics analysis performed to check for peak enrichment.

4.4.1. BMP4 induced transcriptome changes during LPC specification

The RNA sequencing has proved successful with good quality reads of which over 97% aligned uniquely to the human genome. The overview of the sequencing data showed that BMP4 signalling regulates transcriptome changes during the LPC specification. We can see the upregulation of known liver specification associated genes such as HNF4α, ALB, TTR or AFP. Apart from the expected genes, we identified a high number of differentially expressed transcripts. The analysis of the upregulated and downregulated genes by ORA and GSEA allowed grouping of these genes into sets that can give us a better understanding of how BMP4 induced hepatic fate in FE cells.

Not surprisingly, one of the top upregulated terms in KEGG analysis of BMP samples was TGFβ pathway. BMP4 signalling belongs to a wider family of TGFβ growth factors and, therefore, a lot of genes it activated are in common with the wider family. BMP4 inhibitors listed in this category, such as BAMBI, SMAD6,

SMAD7 and Noggin are well known to be co-expressed with BMP4 during mouse and Xenopus development, as they are direct targets of this signalling pathway (Karaulanov et al., 2004, Ishida et al., 2000) and aid proper BMP4 signalling interpretation during development (Paulsen et al., 2011). Other pathways enriched at 24h included ones directly related to hepatic functions such as vitamin digestion and absorption and fat metabolism. However, at this point the enrichment was not statistically significant. This changes at 48h where those pathways become significantly enriched, and additionally, a pathway involved in another hepatocyte function, protein digestion and absorption, is also significantly enriched. These pathways are directly related to hepatocyte functions (Si-Tayeb et al., 2010a). Although LPC are bipotential and have the ability to differentiate towards both hepatocytes and biliary epithelial cells (BEC), our results suggest that the upregulation of genes associated with liver function happens already at an early bipotential stage. However, another explanation of this could be that our differentiation protocol favours hepatocyte fate over the biliary one. In Chapter 3, we show that our LPC express markers of both hepatocytes and BECs (Figure 3-12) and this is also reflected in the RNA sequencing data (Figure 4-6). However, how the gene expression profile and level corresponds to the LPC potential to acquire either fate is hard to predict. Further characterization of our differentiation platform, focusing on the bipotential nature of LCPs may be necessary.

GO of biological processes revealed broad categories of embryo and tissue development or tissue morphogenesis. Genes included in such GO terms would be common to many developmental processes and as such, do not provide specific insight into the impact of BMP4 signalling on hepatic specification. Other enriched terms were associated with mesenchymal and cardiac development. Those terms have many genes in common, such as GATA family, HEY1 and HEY2, TBX2, TBX3 and TBX20, as well as MSX1 and MSX2. These are TFs and, so far, only TBX3 has been previously connected to liver development (Lüdtke et al., 2009, Suzuki et al., 2008). The other TFs are involved in many developmental processes which will be discussed further shortly. At 48h, the GO

analysis identified biological processes much more related to the functions of the liver. From the top ten most enriched terms, first 8 relate to the metabolism of fatty acids (FA). Liver is central to lipid homeostasis. Hepatocytes store FA in the form of triglycerides, process FA coming from dietary and endogenous sources and convert excess glucose into FA in a process of de *novo* lipogenesis. Hepatocytes are also responsible for packaging TG into chylomicrons for transport to muscle and adipose tissue (Alves-Bezerra and Cohen, 2017). Additionally, the GSEA analysis identified enrichment of a few more hepatocyte related functions such as coagulation or haem metabolism in the first 24h of specification. Although at 24 hours the enrichment does not reach statistical significance yet, it points at early activation of hepatocyte functions regulated by BMP4. Taken together, the ORA and GSEA analysis indicate that BMP4 signalling activates genes related to hepatocyte specific functions at quite an early stage in LPC specification. Due to ChIP, failure, we cannot say if these genes are directly regulated by BMP4 effectors or if their upregulation is caused by downstream targets of BMP4 signalling.

Some studies showed that knockout of genes important for hepatic specification may skew cell differentiation towards a different fate e.g.: pancreatic development (Mukherjee et al., 2021) or biliary development (Suzuki et al., 2008, Seth et al., 2014). ORA analysis of the genes upregulated in NOG samples has revealed few clues as to the fate these cells take when BMP4 signalling is blocked. Genes involved in negative regulation of developmental process were significantly enriched among the upregulated PCGs in NOG samples at 24h. Similarly, the 48h NOG samples showed enrichment of genes involved in the negative regulation of cell development and developmental process. This may result from Noggin blocking the progression of FE cells towards the hepatic fate. Considering the terms identified in the KEGG and GO analysis, it is unlikely that FE cells progressed along an alternative developmental pathway towards a different cell fate. Although the GO of biological processes analysis of NOG samples at 48h identified some genes involved in neurogenesis and generation of neurons, the same analysis also identified genes involved in the negative

regulation of nervous system development. The nervous system develops from neuroectoderm and it has been shown in Xenopus that the specification of this tissue is under the control of BMP signalling inhibitors (Lamb et al., 1993, Hemmati-Brivanlou and Melton, 1994). This has been applied to hPSCs differentiation protocols where inhibition of BMP signalling with Noggin is sufficient to induce neuronal progenitors (Gerrard et al., 2005). It is possible that BMP4 blocking by Noggin alters the gene expression/suppression dynamic in a similar way as during neurogenesis and it is the cause of the appearance of the neurogenesis terms in our KEGG analysis of NOG48 upregulated genes rather than a genuine fate transition. This underlines the need for careful interpretation of ORA results.

Interestingly, both ORA and GSEA identified gene sets associated with cell adhesion, locomotion and ECM-receptor interactions. Shortly after hepatic specification the laminin-rich basal layer that surrounds the hepatic endoderm breaks down and LPCs begin delamination and migration into the STM. The process of migration involves cell adhesion molecules and interactions with the ECM (Zorn, 2008). One of the genes upregulated in NOG samples (downregulated in BMP) is E-cadherin (CDH1). This gene was shown to be abnormally upregulated in Prox-1 knockout and double knockout of HNF6/OC2. In both of these mice models, there was a failure of hepatoblast migration into the STM and it is thought that the failure to downregulate E-cadherin expression is one of the main reasons for it (Sosa-Pineda et al., 2000, Margagliotti et al., 2007).

Another example, ICMA-1, an intercellular adhesion molecule, is crucial for cell-cell adhesion, cell-ECM interactions and cytoskeleton rearrangement (Benedicto et al., 2017). It is expressed in endothelial cells around the body, where it facilitates cell-cell attachment. In mice, silencing of ICAM-1 reduces the adhesion of endothelial cells (Lv and Fan, 2020). ICAM-1 is also necessary for immune functions, where for example, it facilitates neutrophil arrest on brain microvascular endothelial cells during inflammation (Gorina et al., 2014). In the liver, it is expressed by LSEC, hepatocytes, Kupffer cells and hepatic stellate cells

and has a major role during initiation of liver metastasis. ICAM -1 is also involved in tumour cell adhesion and cancer metastasis in other cancers and blocking its expression can prevent tumour spread to brain or lung (Benedicto et al., 2017).

Claudins, also downregulated in BMP samples, are important components of tight junctions and play an essential role in regulating the flow of molecules in the intercellular spaces of epithelium. However, they also have been shown to both impair and promote cell motility. Claudin 18 has been reported to supress cell motility in human lung adenocarcinoma cells (Shimobaba et al., 2016), while Claudin 10 expression is associated with reduced migration of clear cell renal carcinoma (Yang et al., 2021). Claudin 4 has been shown to both inhibit and promote cell migration in breast cancer cells (Webb et al., 2013, Levine and Ogunwobi, 2021).

Low expression of IGSF9 (Immunoglobulin superfamily 9), a cell adhesion molecule, correlates with metastasis in breast cancer. IGSF9 knockdown promoted migration of breast cancer cell lines and its overexpression had an inhibitory effect (Li et al., 2022).

The cited studies detail the role of cell adhesion molecules in the process of cell migration. Downregulation of these proteins in the BMP4 samples suggests that this signalling may regulate cell migration in order to facilitate LPC migration of out the hepatic bud and invasion of the STM.

4.4.2. Signalling cross-talk during the LPC specification

An interesting observation was the upregulation of all four FGFR receptors in BMP4 samples. Both BMP and FGF signalling are necessary for hepatic specification and blocking of either of these pathways results in failure of liver bud formation in mice (Rossi et al., 2001, Jung et al., 1999). FGF has been shown to also be necessary for differentiation of LPC from human PSCs (Twaroski et al., 2015) and we have shown that blocking BMP4 signalling during that stage virtually abolishes mRNA expression of liver specific genes (4.3.1). How these two pathways interact with each other to induce hepatic gene expression is not well understood. BMP and FGF co-regulation of developmental processes is

frequently observed and, depending on the context, it can be synergistic or antagonistic in nature. In liver specification setting, the pathways cooperate but BMP4 negatively regulates FGF7 and FGF10 signalling during nephrogenesis (McMahon, 2016). In birds, BMP and FGF cooperation is also involved in LPC fate decision between hepatocyte and biliary epithelial cell (BEC) differentiation. FGF2 and FGF7 were shown to promote BEC fate in cooperation with BMP4 and ECM proteins (Yanai et al., 2008). In avian eye development, differentiation of lens fibre cell from epithelial cells is also regulated by BMP and FGF signalling. The study shows that BMP signalling inhibition affected the responsiveness of epithelial cells to FGF signalling and that this effect was receptor mediated. However, if there is decreased expression of these receptors on the surface of epithelial cells as a result of BMP blocking is not clear as it was not possible to quantitatively assess FGFR protein. QPCR examination of the FGFR1 and FGFR3 mRNA levels showed normal transcript levels between noggin treated epithelial cells and controls (Boswell and Musil, 2015). However, mice knockout models with 50% reduction in transcript levels of FGFR1, 2 and 3 did not form lens fibre cells showing that transcript level changes of these receptor can influence FGF mediated differentiation (Zhao et al., 2008). Additionally, the avian study also show that FGF signalling in turn promotes the expression of BMP target genes showing how interconnected and mutually regulating the two signalling pathways can be (Boswell and Musil, 2015). In our experiments, FGFR transcript levels are upregulated upon BMP4 signalling which could point to the mechanism of how these two singling pathways cooperate to induce hepatic genes expression.

The changes in genes associated with WNT signalling were not so clear. Wnt signalling blocks the expression of at least one of the hepatic factors, HHex, therefore antagonists of this signalling pathway need to be expressed in the FE for correct liver specification (McLin et al., 2007). Our RNA sequencing showed expression of some WNT antagonists, such as NKD1 or FRZB. NKD1 has been identified as one of direct targets of FGF signalling during human PSC differentiation to LPC. NKD1 knockout hiPSC line showed significantly decreased

expression of hepatic markers following the differentiation towards LPCs and this suppression was corrected by addition of XAV939, a small molecule WNT inhibitor (Twaroski et al., 2015). Upregulation of NKD1 in our BMP4 samples indicates that BMP4 signalling also has a role in regulating this pathway. However, more studies are needed to understand the mechanism of this regulation. To complicate the picture, two others well know inhibitors of this pathway, DKK1 and WIF1, are downregulated in the BMP samples. Similarly, the expression changes in WNT receptors are not straightforward. FZD2 and FZD5 are downregulated in BMP samples while FZD4 is upregulated. Kinetic binding assays revealed that WNT ligands have varying binding affinities for different members of FZD receptors family. WNT-3A binds strongly to FZD2/4/5 but WNTs 4, 5A or 5B are more restricted in FZD receptor activation (Dijksterhuis et al., 2015). Therefore, on the basis of our results, it is difficult to gain an understanding of how BMP4 signalling might influence the WNT pathway.

4.4.3. BMP4 signalling induces TF changes

We were interested in the changes in TFs during the specification of LPCs. To date, only three TFs have been shown to be essential to hepatic specification. In double Foxa1 and Foxa2 knockout mice, there is no liver bud formation. These two factors act in a redundant fashion as single knockout of either of them did not result in big abnormalities in liver development (Lee et al., 2005). Similarly, Hnf1β knockout mice show no liver bud formation or hepatic gene expression. Pancreas development is also affected suggesting a wider defect in the competence of FE (Coffinier et al., 2002). HNF4α has been shown to be a critical factor for hepatic specification in hESC differentiation (Delaforest et al., 2011), however in in *vivo* experiments on knockout mice liver specification occurred and a defect appeared at the stage of hepatocyte differentiation (Parviz et al., 2003). Our RNA sequencing experiment identified a number of DEx TFs between cells differentiated in the presence or absence of BMP4 signalling. Little is known about the role of these TFs in hepatic development but many of them are involved in differentiation and specification of different tissues around the body.

HEY1 and HEY2 proteins are well known direct targets of Notch signalling in cardiovascular system development and mostly function as transcriptional repressors (Weber et al., 2014). However, both TFs have been shown to be activated by various BMP ligands in different developmental settings. HEY1 is regulated by BMP9 signalling during osteoblast lineage differentiation of mesenchymal stem cells. ChIP analysis identified HEY1 as a likely direct target of SMAD proteins (Sharff et al., 2009). HEY2 is induced by BMP signalling in the late phase differentiation of the retina (Kuribayashi et al., 2014). The two factors were shown to share many targets and binding sites, inducing mainly gene suppression. They have roles in cardiac development and disease, vascular development, myogenesis, bone development and homeostasis and neural development (Weber et al., 2014).

Similarly, MSX1 and MSX2 are involved in several aspects of human development. They have been shown to function in a redundant manner in limb development (Lallemand et al., 2005) and MSX2 is a downstream mediator of BMP4 signalling during the EMT of differentiating hESCs (Richter et al., 2014). Considering that our RNA sequencing screen showed that some EMT associated genes are downregulated in the BMP samples and that these two TFs are involved in many other developmental processes, they are interesting candidates for further study of their role in hepatic specification.

TBX3 is already a known TFs involved in hepatic development (Suzuki et al., 2008) and TBX2 and TBX3 have redundant roles in mice lung development (Ludtke et al., 2016). TBX2 is a transcriptional suppressor involved in heart, brain, eye, bone, limb and mammary gland development and it has also been implicated in cancer, including liver cancer (Abrahams et al., 2009).Search for publications on TBX2 or TBX20 involvement in liver development has not returned any results.

The redundant roles of some of the TFs identified in our DEx analysis must be taken into consideration when planning functional studies. Single knockouts frequently do not induce an abnormal phenotype and the involvement of the TFs in specific developmental program may be missed. Therefore, double or

triple knockdown or knockouts experiments need to be considered for further study of these TFs.

4.4.4. BMP4 induced lncRNA changes

Our RNA sequencing has also identified a number of DEx lncRNAs. This class of transcripts (non-coding, longer than 200nt) has gained much attention since the advent of high throughput signalling techniques and the realisation that almost every nucleotide of the DNA is transcribed under certain conditions and that only around 2% of DNA codes for proteins (Djebali et al., 2012). A well-known role for lncRNA is X-chromosome inactivation in female mammals by Xist (Penny et al., 1996), but these transcripts have now been implicated in the control of pluripotency and differentiation as well as cancer and other diseases. In development, two lncRNAs have been shown to control TFs important for endoderm differentiation. DIGIT controls the expression of Gooscoid, while DEANR1 has been implicated in SMAD2/3-mediated FOXA2 transcription (Jiang et al., 2015, Daneshvar et al., 2016). The analysis of DEx transcripts identified a high number of DEx lncRNAs. TBX2-AS1 was one of the most strongly upregulated lncRNAs in BMP24 samples. Similarly, three other TFs associated lncRNAs showed up in the top ten most upregulated hits: GATA2-AS1, GATA3-AS1 and GATA6-AS1 lncRNAs. All of those lncRNAs are anti-sense and overlapping their respective TFs. All associated TFs were also hits in the DEx TF screen. It is an interesting question if the expression of these TFs is somehow regulated by lncRNAs and how the expression of lncRNAs is controlled by BMP4 signalling.

Chapter 5. Generation and validation of inducible hiPSC lines for the functional study of candidate transcription factors upregulated by BMP4 signalling during hepatic specification

5.1. Introduction

In chapter 4 we identified TFs that are upregulated very early upon the activation of BMP4 signalling during the specification of LPCs from foregut endoderm. We hypothesise that those TFs are either direct or indirect effectors of BMP4 signalling and therefore can be of critical importance for early liver development. We have selected 9 candidates for functional study. Initially, we planned to study the effects of gene knockdowns, knockouts and overexpression on the progression of the hepatic specification. However, the effects of COVID pandemic have severely limited the time available in the laboratory and the scope of experiments we were able to conduct. Therefore, we focused on the study of the loss-of-function of our candidate genes as the building of plasmids necessary for this part of the project was most advanced at the time of COVID interruption. As shown in Chapter 3, the strategies for transfection and nucleofection of foregut endoderm monolayer were suboptimal. To circumvent the issue we identified a method of conditional manipulation of gene expression in hiPSC and its differentiated derivatives via inducible shRNA expression (Bertero et al., 2018).

5.1.1. Conditional expression of shRNAs against candidate TFs in hiPSCs

The protocol published by Bertero et al., details an OPTimized inducible Knockdown (OPTiKD) method for rapid genetic modification of hPSCs with a customisable transgene that enables conditional expression of specific shRNA against single or multiple gene/s of interest. The transgene is targeted to an AAVS1 locus designated as a 'safe harbour' (Figure 5-1a). Such locus is reported to be expressed in majority of human tissues, to be resistant to silencing during differentiation and can be modified without negative effects on normal

functioning of the cellular activity (Sadelain et al., 2012). The cells are genetically modified by introduction of double strand breaks (DSB) in the targeted AAVS1 region, triggering cellular DNA repair mechanisms: non-homologous end joining (NHEJ) or homology directed repair (HDR) (Chatterjee and Walker, 2017). The latter repair mechanism can facilitate the introduction of a transgenes into the site of the DSB (Figure 5-1b). The template for the repair is provided with the pAAV_puro_siKD (siKD) plasmid. The introduced transgene contains all the functional elements necessary for the OPTiKD method between the homology arms (Figure 5-1a). The transcription of the shRNAs is controlled via the TET OFF system (Figure 1-1). In the absence of tetracycline, the transcription of shRNAs is blocked by the binding of tetracycline-sensitive repressor protein (TetR) to the tet operone sequences (TO) within the H1 promoter. The addition of tetracycline induces confirmational change in the TetR protein and its dissociation from the H1 promoter, enabling the transcription of the shRNAs to proceed. This method allows the temporal control of knockdown in dose-responsive and reversible fashion (Kallunki et al., 2019). The inducible nature of the system was advantageous for our study as it allowed the avoidance of knockdown effects at the earlier stages of differentiation. TFs are known to have dual functions during development (Kyle and Lim, 2011, Wang et al., 2012), therefore it was crucial to ensure any observed effects on hepatic specification are due to gene knockdown at the time of the specification and not a consequence of earlier disturbance to differentiation.

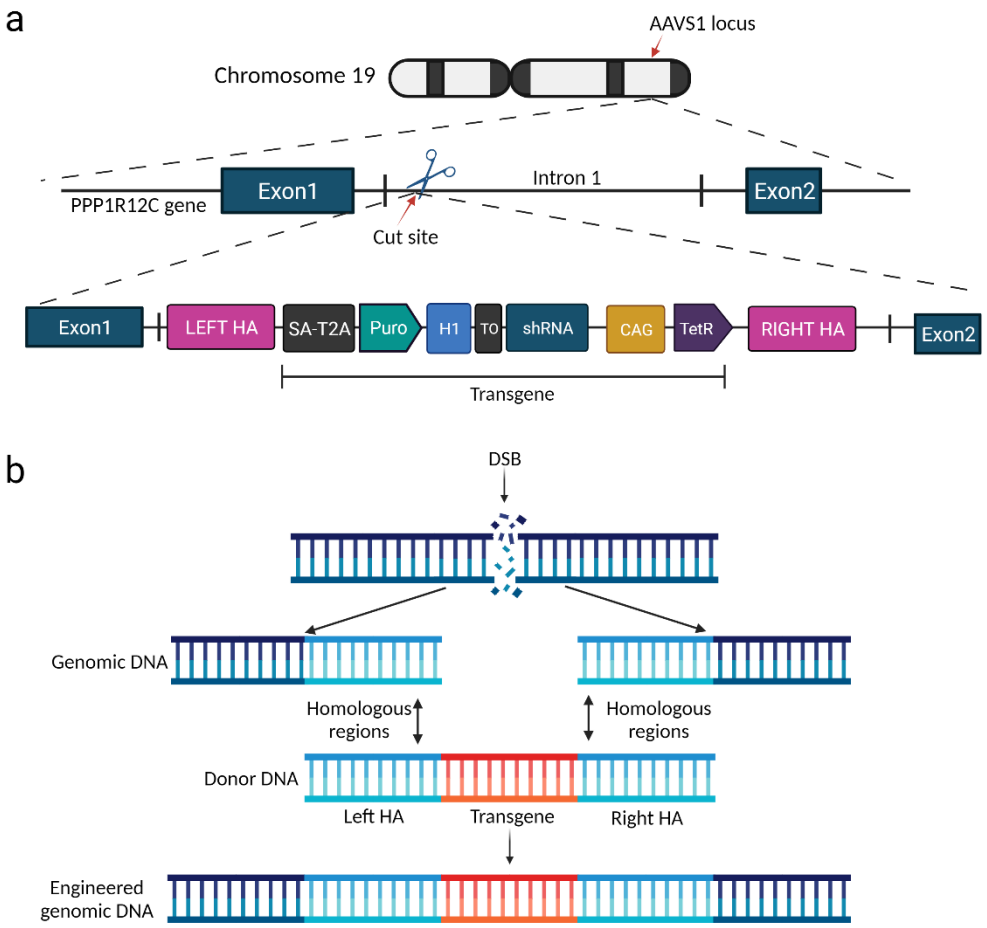

Figure 5-1 AAVS1 locus targeting for inducible expression of shRNAs.
a) Schematic of AAVS1 locus and targeting site within the intron of PPP1R12C gene and components of the transgene cassette; b) Mechanism of DSB repair by homology directed repair.

5.1.2. CRISPR/Cas9 targeting of the AAVS1 locus

The Bertero protocol introduces DSB in the AAVS1 locus with the use of Zinc finger nucleases, a class of programmable nucleases that can be targeted to the locus of interest. In our study, we decided to use CRISPR/Cas9 system, a most recent method to genetically modify cells. We had access to previously designed and validated guide RNAs targeting the required area of the AAVS1 locus. Additionally, the traditional Cas9 enzyme inducing DSB was replaced with its mutated version, Cas9 D10a, with nickase activity. Wild type (WT) Cas9 enzyme has been shown to produce multiple off target mutations as up to 5 mismatches are tolerated within the guide RNA (Fu et al., 2013). DNA cuts in random locations can introduce mutations when the DNA is fixed by NHEJ, and an insertion/deletion (indel) is created. Cas9D10a creates SSB in the DNA and to

create DSB for HDR to occur it needs to be directed to the cut site by two different guide RNAs. This significantly reduces the chances of random DSB occurring somewhere else in the genome (Shen et al., 2014).

5.2. Chapter aims and objectives

The aim of this chapter was the generation of several inducible knockdown hiPSC lines. Each line was engineered to contain conditionally inducible transgene in a 'safe harbour' locus capable of expressing shRNAs targeting our selected TFs. Modified hiPSC cells lines were differentiated to hepatic progenitors and the effect of specific genes knockdowns on specification of hepatoblasts were determined by morphology and gene expression analysis.

5.3. Results

5.3.1. Preparation of donor template and CRISPR plasmids

To create hiPSC cells lines with inducible secretion of shRNAs designed to knockdown our TFs of interest we obtained 4 plasmids. Targeting plasmid for AAVS1 locus, pAAV_puro_siKD, that can easily be modified to contain required shRNAs, was designed and created by (Bertero et al., 2016). It was purchased via addgene.org website (Plasmid #86695) and delivered as a bacterial stab (Figure 5-2a). The plasmid was purified and quality checked by Sanger sequencing of the H1+TO and multiple cloning site (MSC) to ensure correct sequence for modifications of the plasmid (Figure 5-2 b). Additionally, diagnostic digest with EcoRI and PsiI restriction enzymes was performed to quality check the plasmid backbone. The double digestion yielded three bands of roughly the right size at 3938bp, 3551bp and 1998 bp (Figure 5-2 c). Further, the plasmid was linearised by double digestion with BglII and SalI enzymes and dephosphorylated with alkaline phosphatase in preparation for cloning with shRNA oligos. The double digestion of the plasmid creates DNA fragments of 21bp and 9466bp. The smaller fragment cannot be visualised on the same agarose gel as the large one and therefore, to confirm the successful linearisation of the digested plasmid, it was run on the gel alongside its circular,

undigested form (Figure 5-2 d). The supercoiled DNA can travel faster in the gel and can help identify the fully digested plasmid.

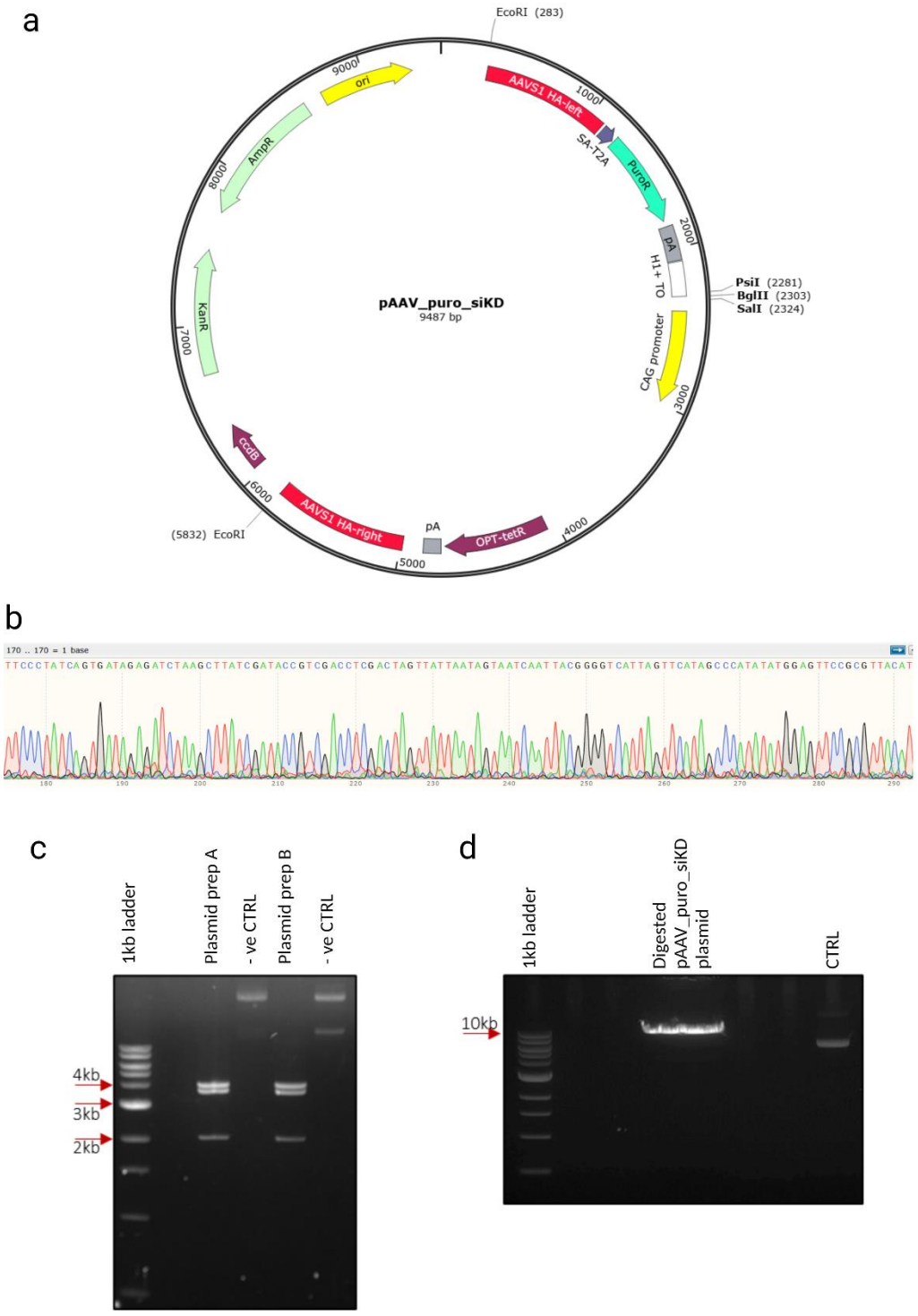

Figure 5-2 Quality control and preparation of the pAAV_puro_siKD plasmid.
a) Plasmid map detailing elements of the targeting vector; b) Sequencing check of the promoter, Tet operon (H1+TO) and MSC regions of the plasmid; c) Diagnostic digestion of the plasmid with EcoRI and PsiI restriction enzymes; d) Plasmid linearization using BglII and SalI restriction enzymes.

The CRISPR/Cas9 plasmids for targeting of the intron of the PPP1R12C gene within the AAVS1 locus were kindly provided for our experiments in purified form by the Denning lab (Figure 5-3 a, c). Guide RNAs were designed and produced by Dr Jamie Bhagwan. Quality check by Sanger sequencing of the guide RNA and RNA scaffold region confirmed correct sequences (Figure 5-3 b). Partial sequence of the Cas9 D10a was also performed as a quality check (data not shown).

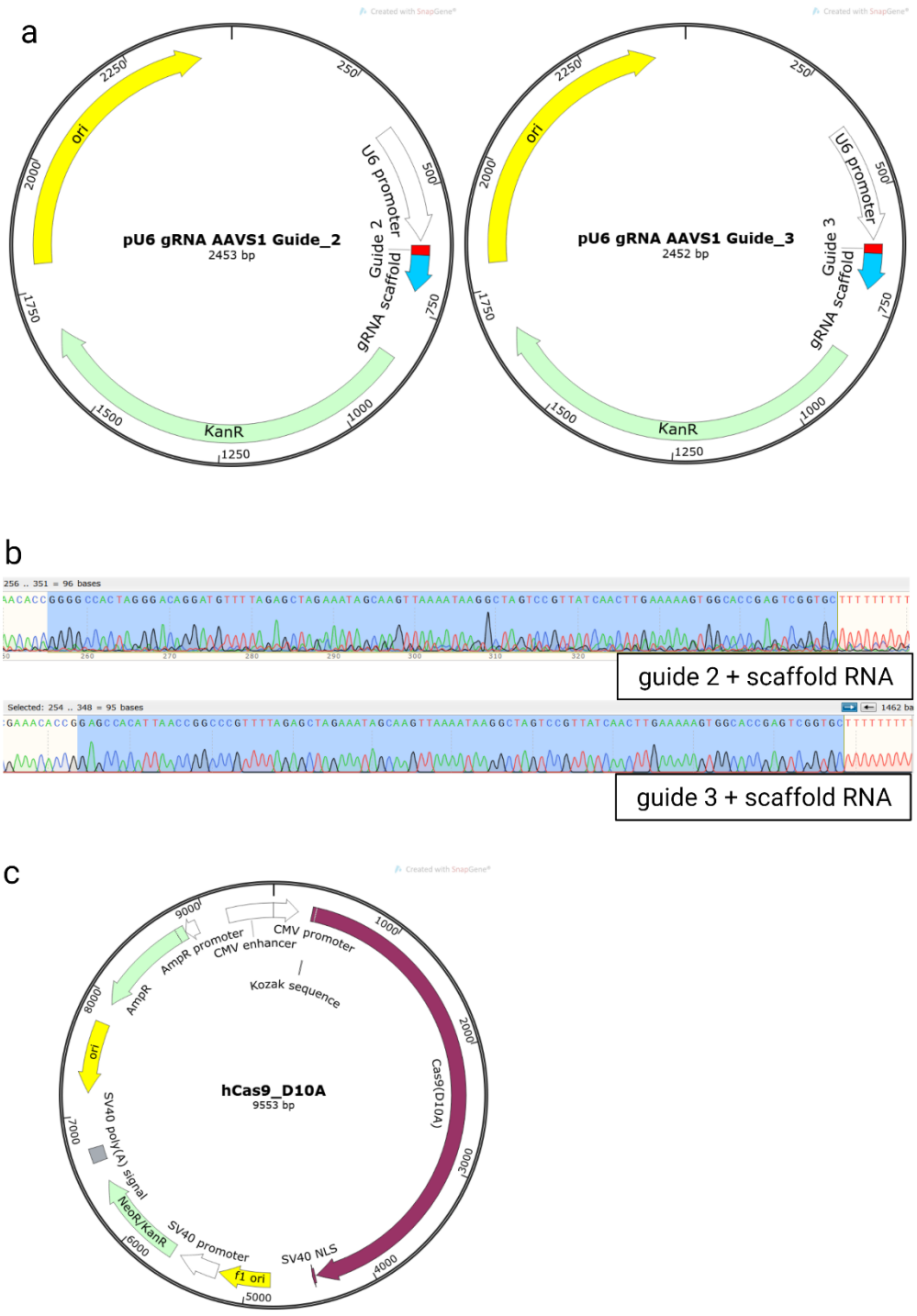

Figure 5-3 CRISPR/Cas9 plasmids for targeting of the AAVS1 locus.
a) Plasmid maps of guide RNAs: b) Sanger sequencing check of the guide RNA and scaffold RNA; c) Plasmid map of Ca9 D10A containing plasmid.

5.3.2. Creation of individual pAAV_puro_siKD plasmids (siKD)

The initial step to create Multi pAAV_puro_siKD (MsiKD) plasmid with shRNAs against multiple genes was to create the individual pAAV_puro_siKD plasmids with individual shRNAs targeting each of our selected TF. Figure 5-4 a, b, c, d explains the generic cloning strategy on the example of constructs with shRNAs against TBX genes. DS shRNA oligos were ligated into an empty siKD vector. The ligated constructs were transformed in bacteria and individual bacterial colonies were PCR-screened. Insertion of the DS oligo into the empty vector produces a DNA fragment that is 37bp longer than one produced from an empty plasmid. The change is subtle but clearly visible (Figure 5-4 e). Positive colonies were expanded and sent for Sanger sequencing to confirm correct insertion and sequence of each shRNA (Figure 5-4 f). SiKD constructs for shRNAs against GATA, MSX, MAF and HEY genes were created following the same procedure. SiKD_SCR (Scramble shRNA) construct was also produced. (Figure 5-5;Figure 5-6).

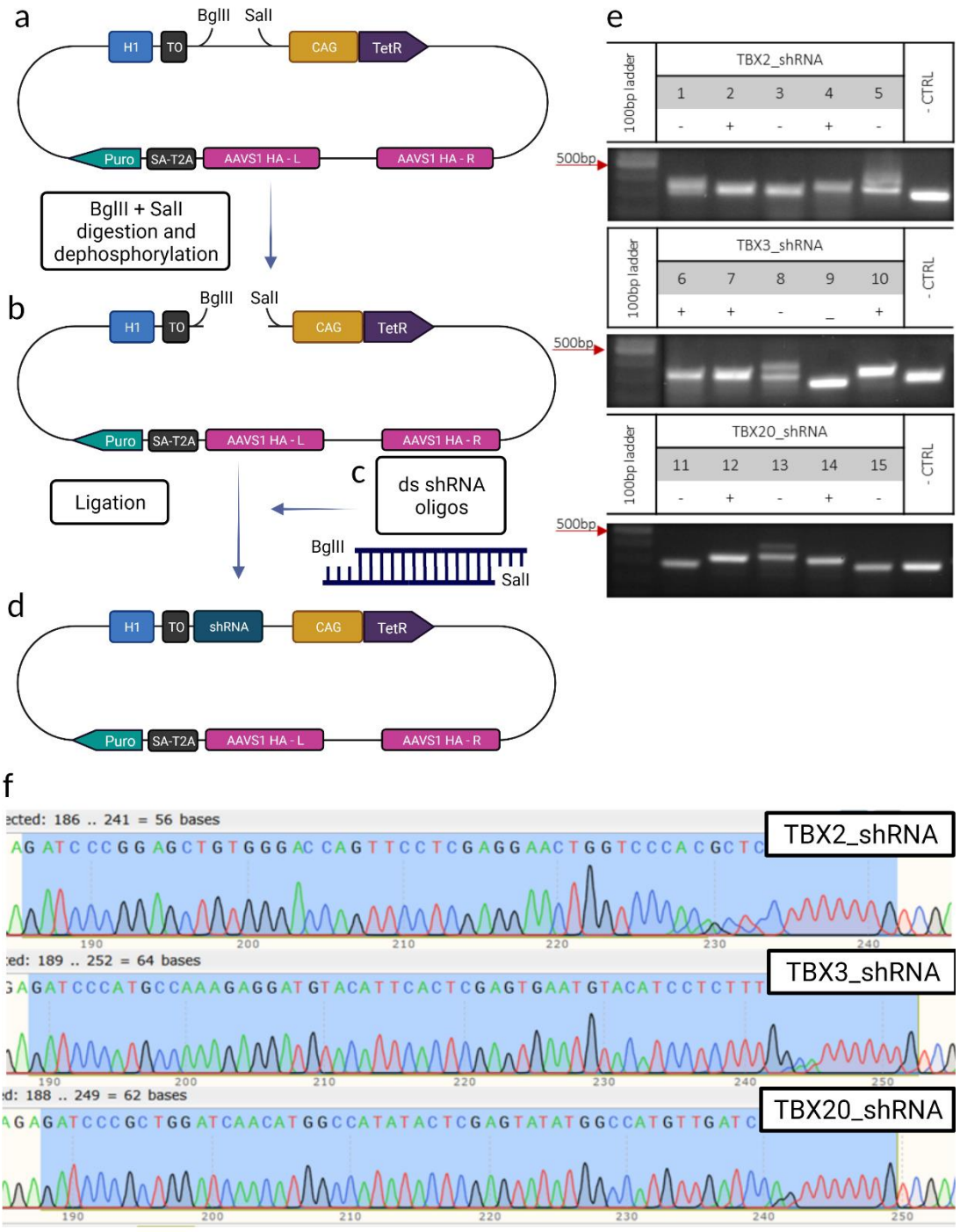

Figure 5-4 Building of the pAAV_puro_siKD plasmids for individual TBX genes.
a) Empty pAAV_puro_siKD plasmid; b) pAAV_puro_siKD plasmid digested with BglII and SalI restriction enzymes; c) Annealed DS shRNA oligos with sticky ends complimentary to BglII and SalI restriction sites; d) pAAV_puro_siKD plasmid with ligated DS shRNA oligo; e) PCR screen of bacterial colonies following ligation of the plasmid and DS shRNA oligos for individual TBX genes; f) Sequencing confirmation of correct insertion and sequence of inserted shRNAs.

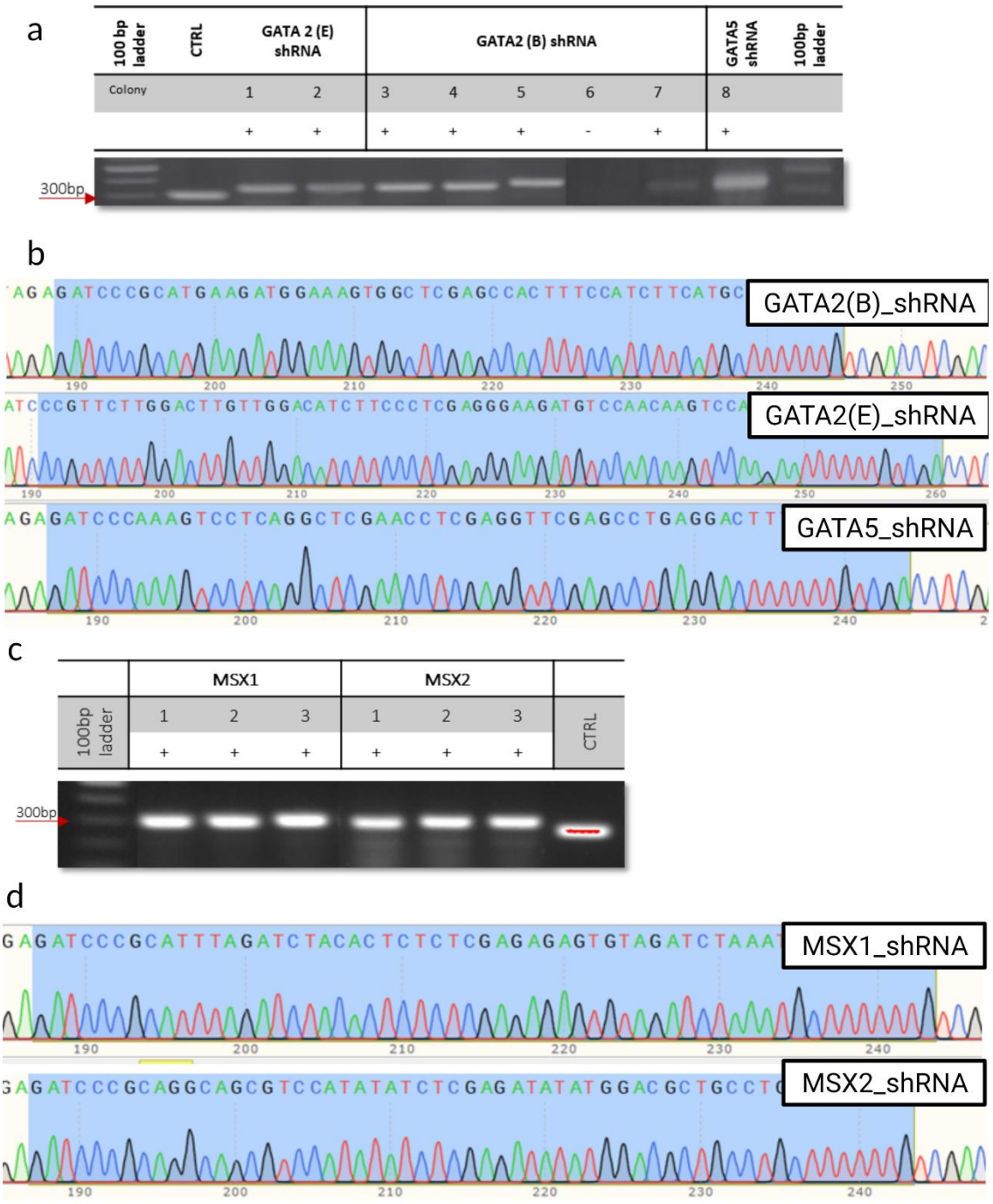

Figure 5-5 Building of the pAAV_puro_siKD plasmids for individual GATA and MSX genes.
a),c) PCR screen of bacterial colonies following ligation of the plasmid and DS shRNA oligos; b),d) Sequencing confirmation of correct insertion and sequence of inserted shRNAs.

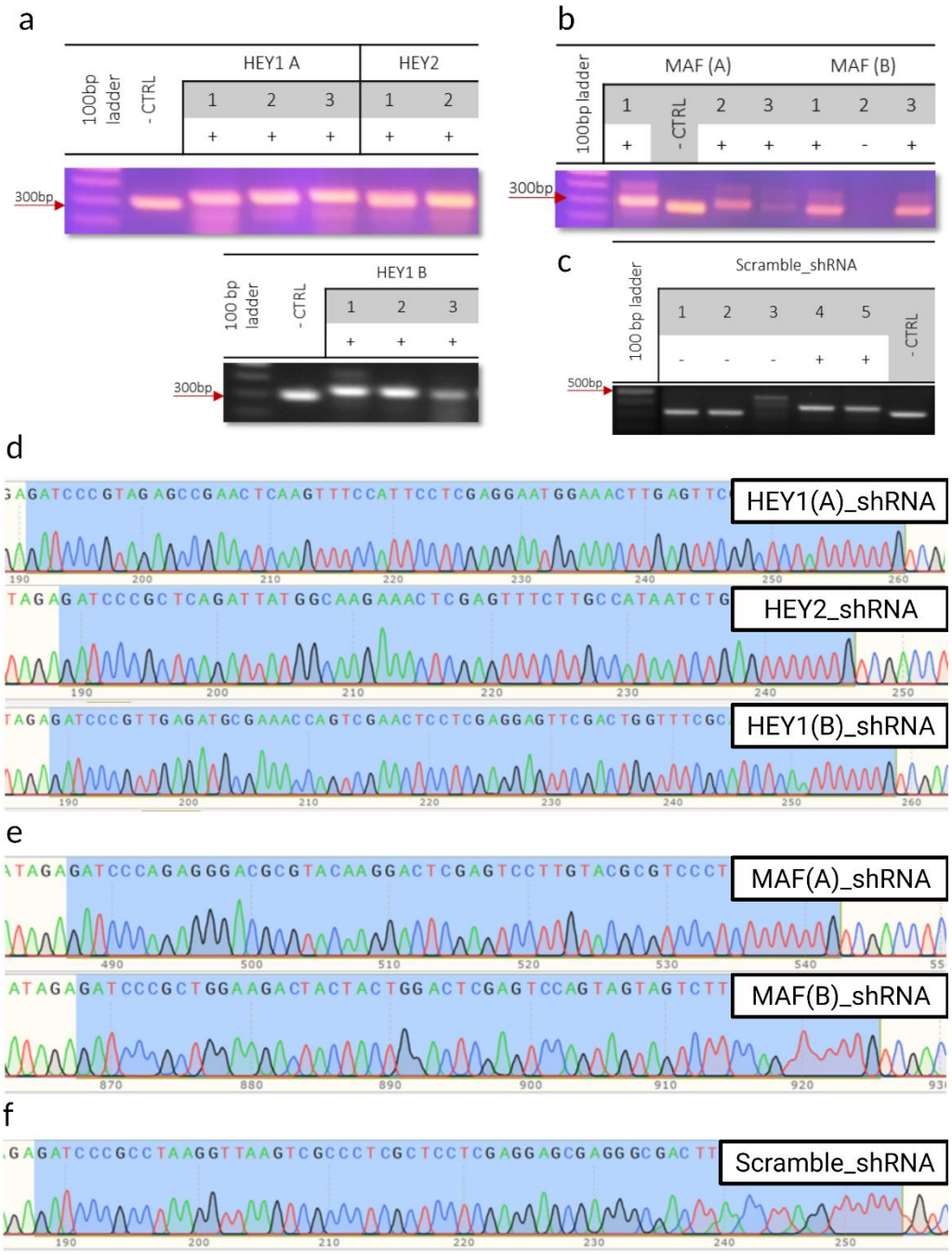

Figure 5-6 Building of the pAAV_puro_siKD plasmids for individual HEY and MAF genes and Scramble shRNA.

a),b) and c) Bacterial colony screen following ligation of the plasmids and DSs oligos for HEY, MAF and Scramble constructs, respectively; d),e) and f) Sequencing confirmation of correct insertion and sequence of HEY, MSX and Scramble DS shRNA oligos, respectively.

5.3.3. Creation of Multi pAAV_puro_siKD plasmids (MsiKD)

The next step was to create constructs that contained 2 or 3 shRNAs against 2 or 3 members of the same gene family. Figure 5-7 (a, b, c) explains the cloning

strategy for triple shRNA assembly for TBX plasmid. Individual siKD TBX plasmids were amplified with specific PCR primers to create shRNA cassettes. Those cassettes contained the H1 promoter, TO, shRNA and sequences complimentary to, either the plasmid backbone, or each other (Figure 5-7a). The cassettes were cloned into an empty pAAV_puro_siKD plasmid, previously digested with BstBI and HincII restriction enzymes (Figure 5-7b), by Gibson assembly creating MsiKD_TBX plasmid (Figure 5-7c). Successful cloning was identified by bacterial PCR. The assembly of three shRNA cassettes created a band of around 1.6kb (Figure 5-7d). Positive colonies were sent for sequencing with several primers to confirm insertion and correct sequence of each TBX shRNA cassettes (Figure 5-7e). MsiKD_GATA and MsiKD_HEY were produced following the same procedure (Figure 5-8). Creation of MsiKD_MSX involved assembly of two shRNA cassettes with the empty vector (Figure 5-9a, b, c). PCR screen produced a band of around 1.2kb (Figure 5-9d). Positive colonies were sent for Sanger sequencing to check the insertion and correct sequence of each MSX cassette (Figure 5-9e). MsiKD_MAF was produced following the same procedure (Figure 5-10).

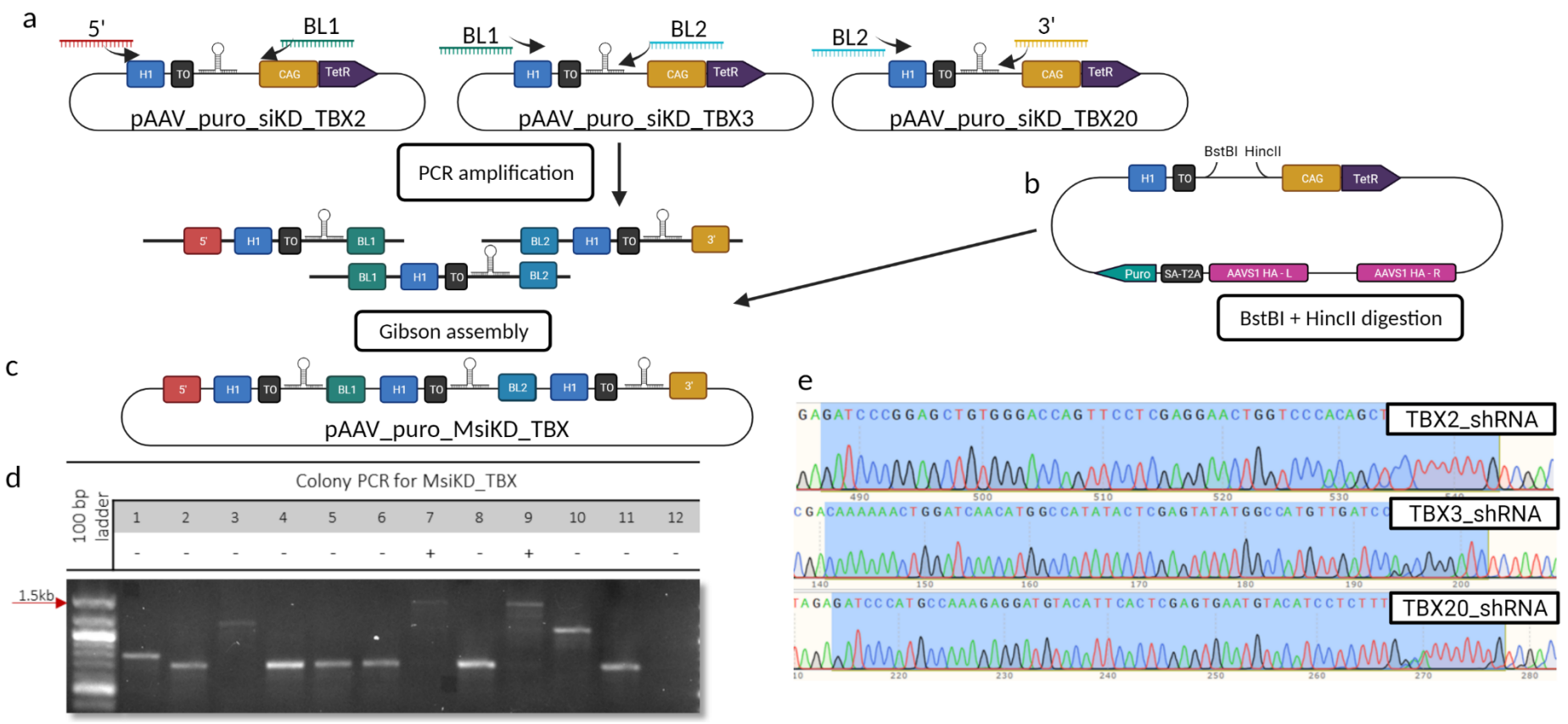

Figure 5-7 Cloning strategy for building of the MsiKD_TBX (triple assembly).

a) Individual siKD plasmids are PCR amplified with specific primers containing overlap regions with either the plasmid backbone or each other, to create shRNA cassettes; b) Empty siKD plasmid is linearised by digestion with BstBI and HincII restriction enzymes; c) The shRNA cassettes and digested plasmid are cloned using Gibson assembly and produce MsiKD_TBX plasmid; d) Bacterial colony screen for triple assembly of TBX shRNA cassettes; e) Sanger sequencing of colony 7.

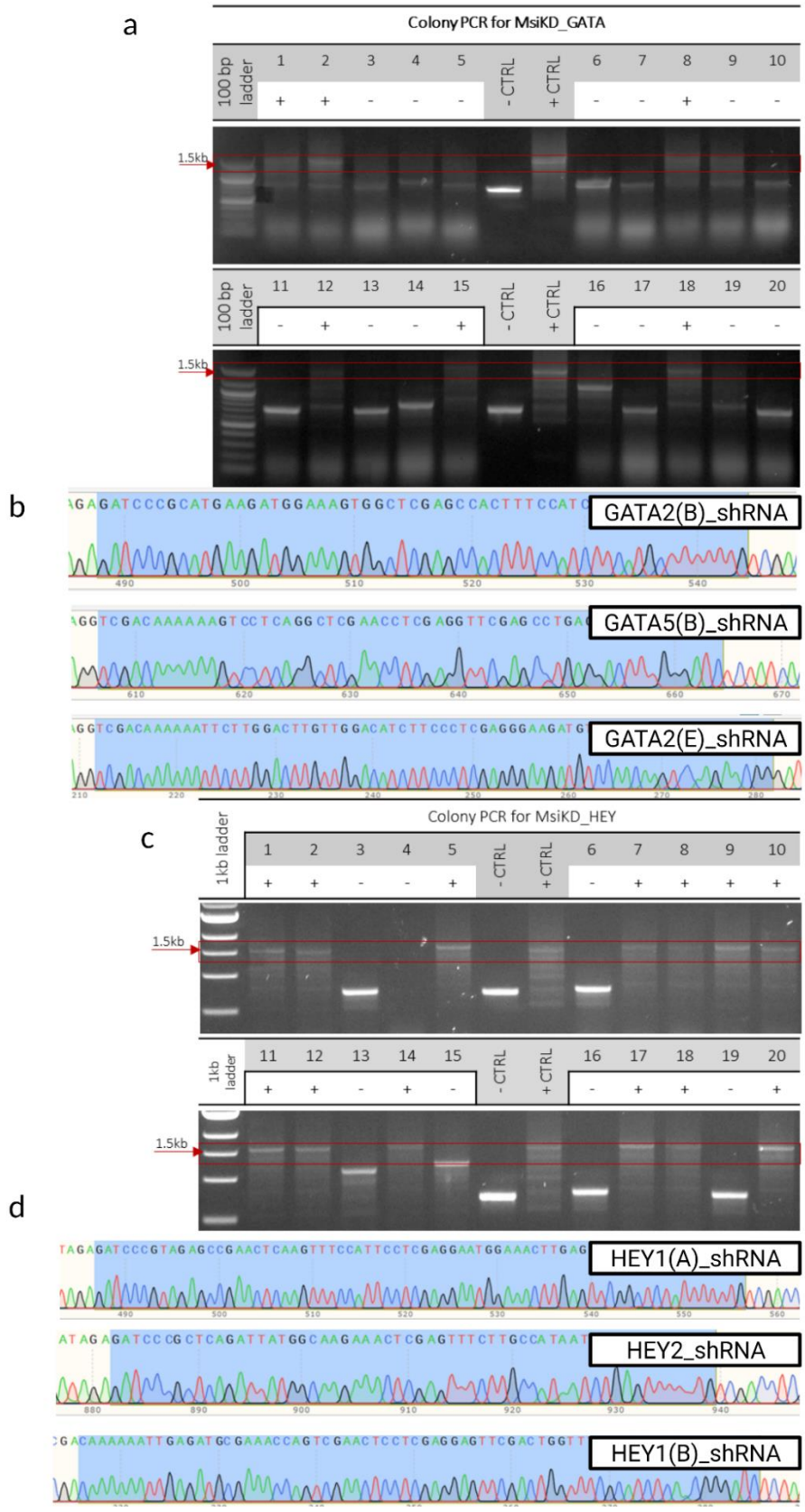

Figure 5-8 Building of the MsiKD for GATA and HEY genes.
a),c)PCR screen following Gibson assembly of GATA and HEY shRNA cassettes, respectively; b)c) Sanger sequencing confirmation of correct integration and sequence of MsiKD_GATA and MsiKD_HEY, respectively.

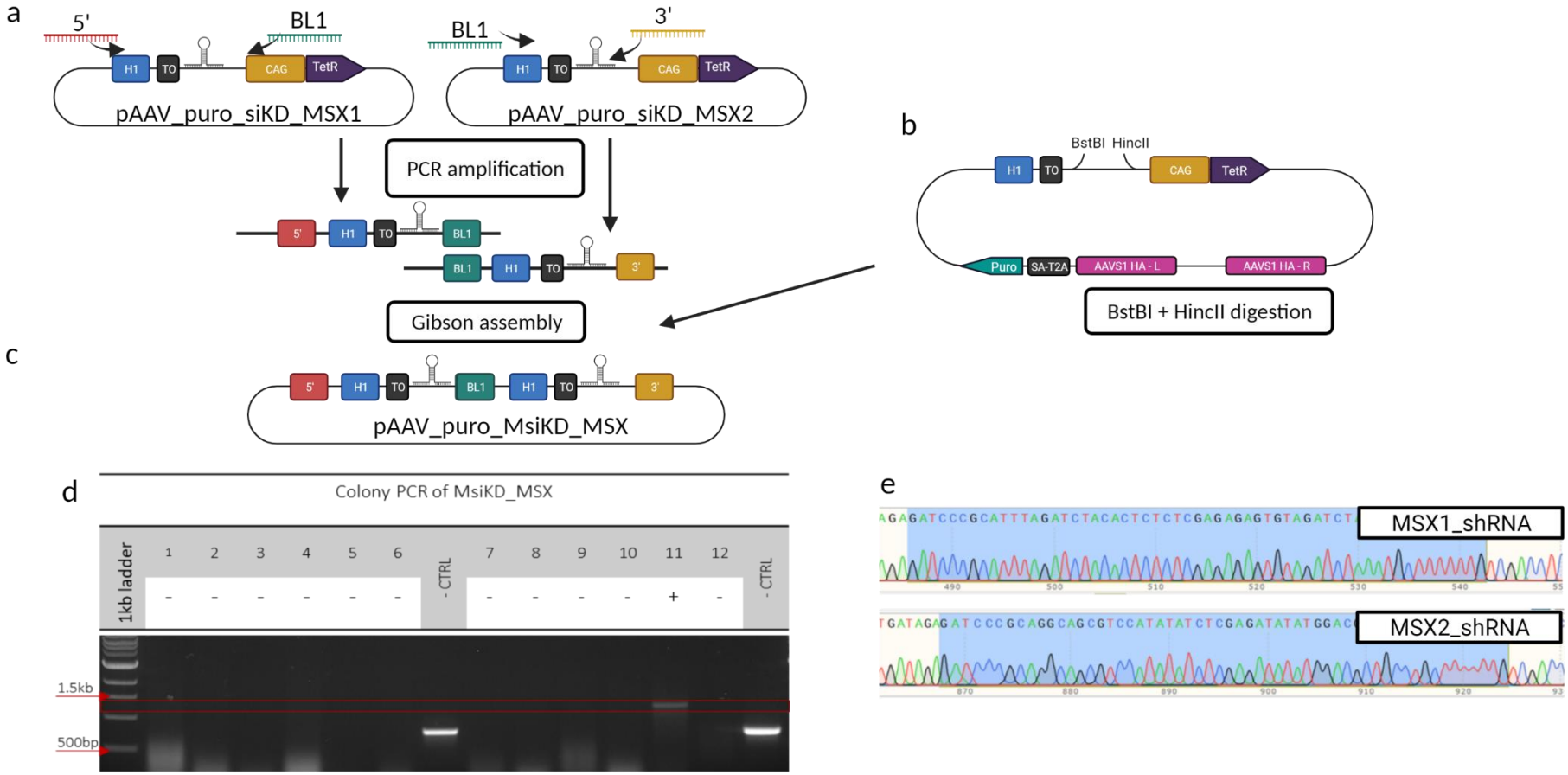

Figure 5-9 Cloning strategy for building of the MsiKD_MSX (double assembly).

a) Individual siKD plasmids are PCR amplified with specific primers containing overlap regions with either the plasmid backbone or each other, to create shRNA cassettes; b) Empty siKD plasmid is linearised by digestion with BstBI and HincII restriction enzymes; c) The shRNA cassettes and digested plasmid are cloned using Gibson assembly and produce MsiKD_MSX plasmid; d) Bacterial colony screen for double assembly of MSX shRNA cassettes; e) Sanger sequencing of colony 11.

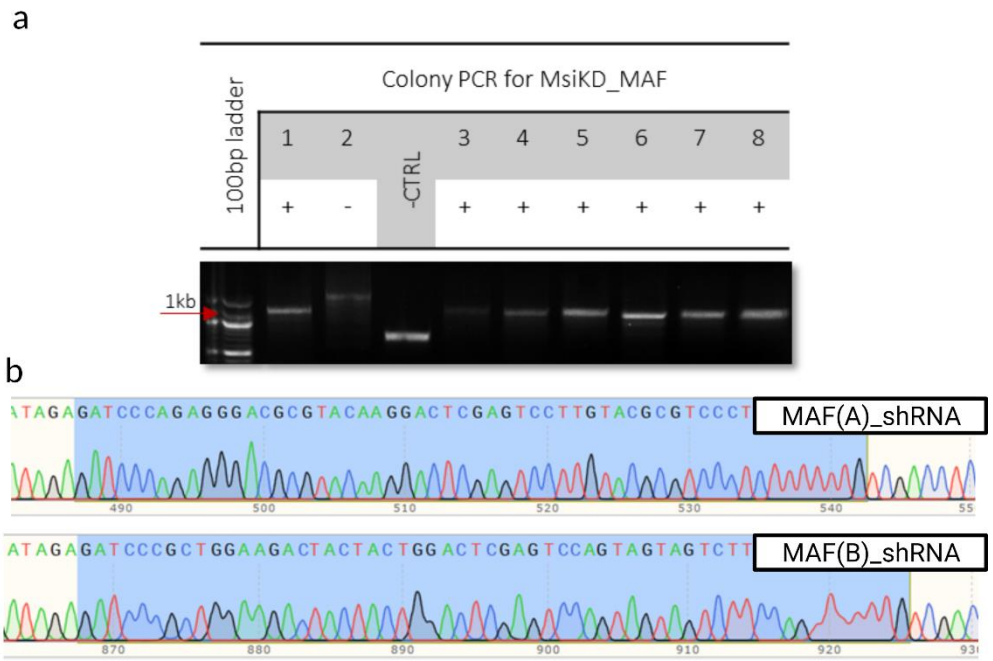

Figure 5-10 Building of the MsiKD MAF gene.
a) PCR screen following Gibson assembly MAF shRNA cassettes; b) Sanger sequencing confirmation of correct integration and sequence of MsiKD_MAF

5.3.4. HiPSC transfection and selection

Once construction of all targeting vectors was completed, we proceeded to genetic modification of our hiPSC cell line R-PAT M. The outline of the transfection strategy to create inducible knockdown cell lines is shown in Figure 5-11.

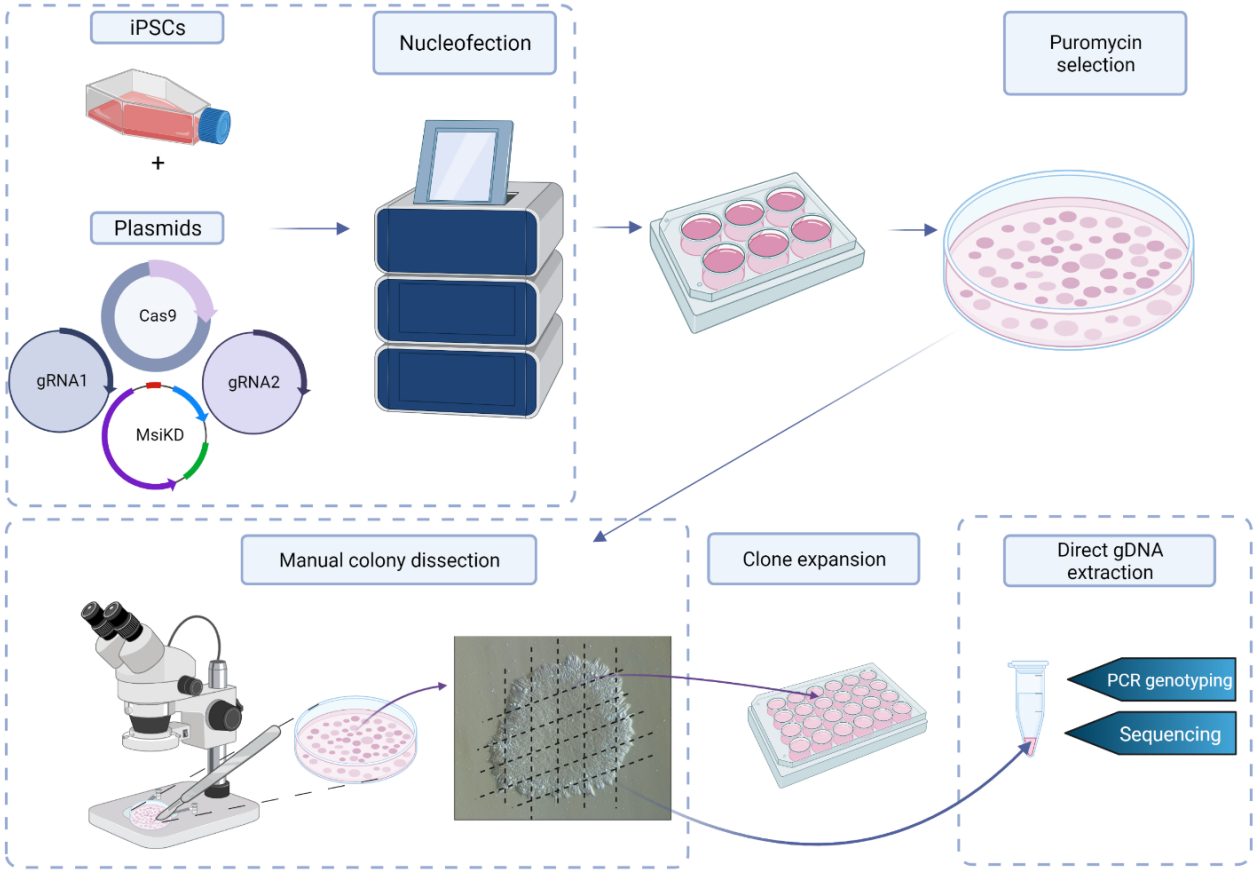

Figure 5-11 HiPSC targeting steps.

Firstly, the selection step of our strategy needed optimization. The integrated transgene contained antibiotic resistance gene puromycin-N-acetyltransferase (PAT) which inhibits protein synthesis inhibitor puromycin. Therefore, puromycin treatment of transfected cells can enrich for successfully targeted cells. Puromycin resistance test was performed on our unmodified hiPSC line to identify puromycin concentration that kills 100% of cells at 48hrs. Puromycin concentration of 0.3μg/ml was shown to be sufficient to eliminate virtually all cells within 48hrs (Figure 5-12).

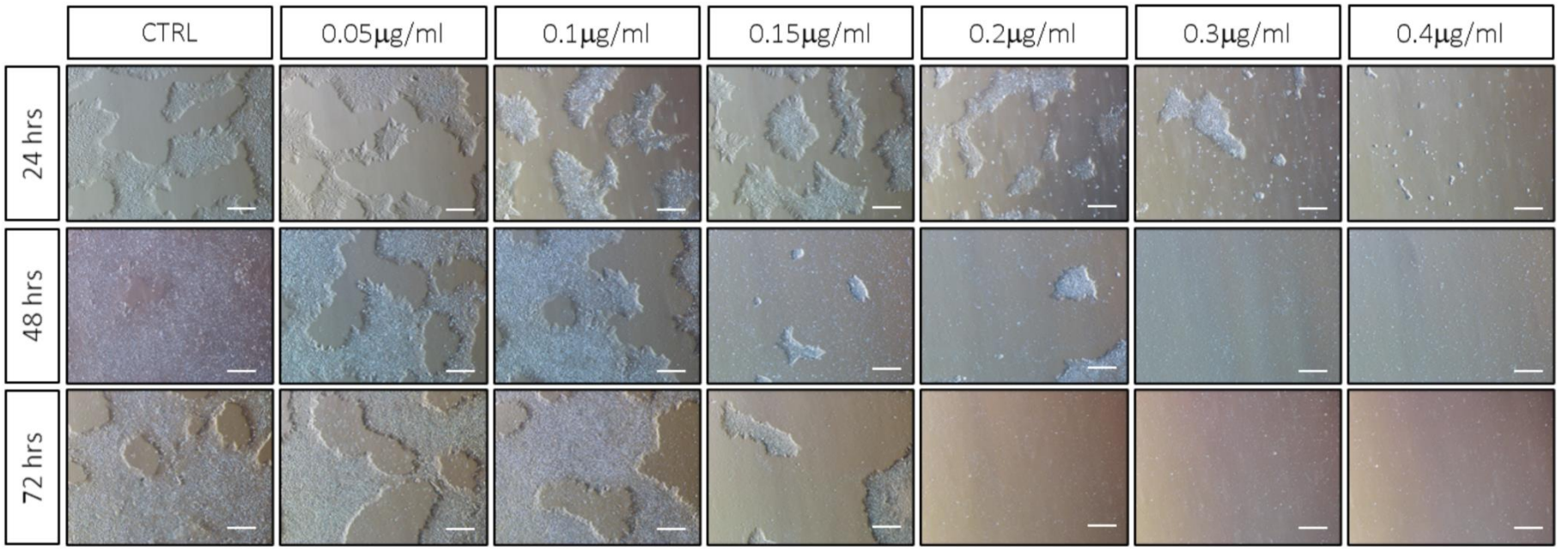

Figure 5-12 Puromycin resistance test of R-PAT M cell line.
Cells were seeded at 20k/cm^2 density and allowed to proliferate for 48hrs.Specified concentration of puromycin was added every 24hrs with the media change. Representative microscopy images show changes in cell confluence in increasing puromycin concentration over 72hrs (n=2). Scale bar = 200µm.

Once the optimal concentration of puromycin was identified for selection, we proceeded with targeting of the R-PAT M hiPSC line with all the MsiKD constructs. CRISPR and MsiKD plasmids were complexed and delivered to the cells using Lonza nucleofection kit with Amaxa nucleofector. Transfection efficiency was monitored for each targeting with flow cytometry analysis of GFP expression. Post nucleofection, cells were seeded into the 6 well plate and allowed to recover and expand for 48hrs. Next, the cells were seeded into a 53cm^2 petri dish for puromycin selection using 0.3μg/ml concentration. Following selection, cells were manually dissected using a stem cell cutting tool. A small fragment of each colony was collected for direct gDNA extraction and PCR genotyping, and the remaining fragments were placed in a well of a 24 well plate for further expansion (Figure 5-13). The manual dissection method gave an excellent survival of the picked clones (95.1%±6.3%). Clones determined as successfully targeted on the basis of PCR genotyping were expanded through successive splitting from 24wp to-12wp to T25 format.

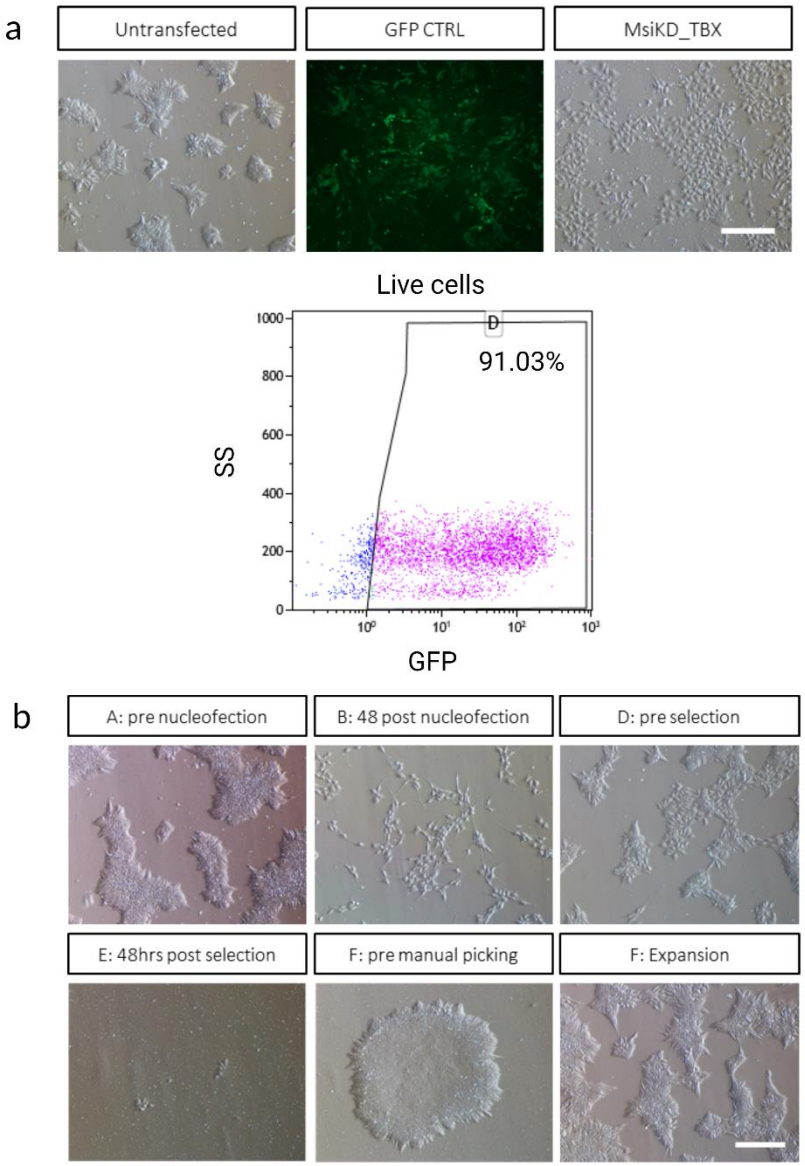

Figure 5-13 Targeting of R-PAT M hiPSC with MsiKD_TBX plasmid.
a) R-PAT M cells were transfected using Lonza nucleofection kit. The efficiency of each nucleofection was monitored by fluorescent microscopy. The percentage of live cells positive for GFP was assessed using flow cytometry and was used to monitor the efficiency of each transfection; b) Following nucleofection, cells were transferred to a petri dish and were selected by a 48hrs puromycin treatment. Surviving cells were allowed to grow into colonies before manual dissection. Positive clones were expanded into a monolayer in a T25 format. Scale bars = 250μm.

5.3.5. Genotyping of MsiKD clones

To establish if the picked clones were successfully targeted, i.e.: the OPTiKD cassette was integrated at the right position in the genome, with no spare plasmid copies and homozygous/heterozygous manner, five PCR reactions were performed following Bertero protocol (Bertero et al., 2018).

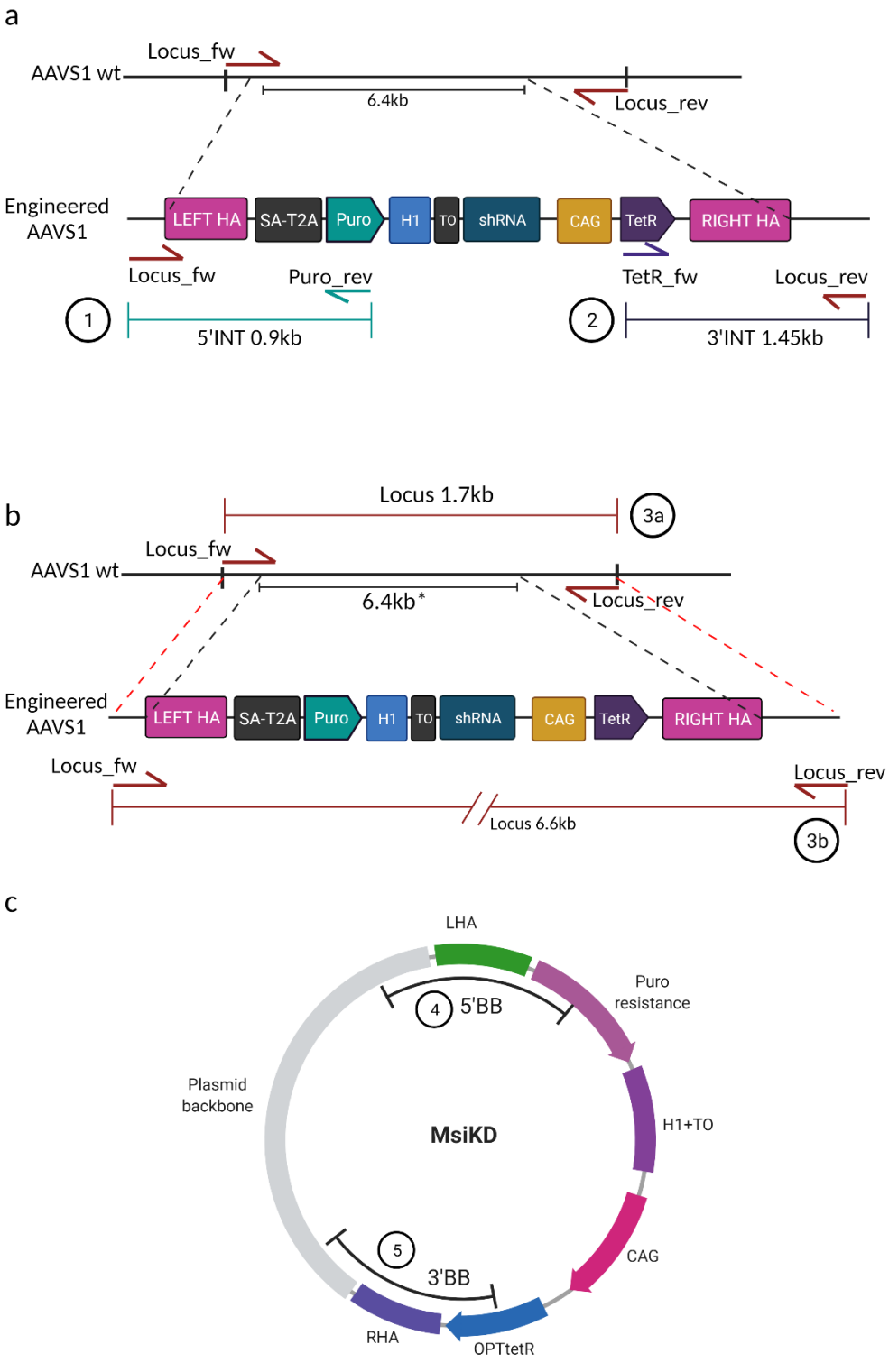

Figure 5-14 Strategy for genotyping of MsiKD clones.
a) PCR reaction 1 and 2 identify successfully inserted OPTiKD cassette within the AAVS1 locus; b) PCR reaction 3 amplifies a region from outsides of both homology arms. When the OPTiKD cassette is correctly inserted the DNA fragment between two primers becomes over 6kb long and there is a failure of amplification. The reaction allows identification of negative clones or heterozygous clones (presence of gel band of 1.7kb (3a)) and homozygous clones (no gel band present (3b)); c) Reactions 4 and 5 identify the presence of extra copies of the plasmid within the genome.

Primers for reactions 1 (5'INT) and 2 (3'INT) covered the region of the genome from outside of the homology arms to within the OPTiKD cassette (Figure 5-14a). Amplification of the product indicates successful integration of the cassette within the AAVS1 locus of the cell's genome.

Primers for PCR reaction 3 (Locus) were complimentary for regions in the AAVS1 locus just outside the homology arms (Figure 5-14 b). PCR reaction of negative or heterozygous clones produces a gel band of 1.7kb in length. For homozygous clones, no band would be produced due to amplification failure as the insertion of the transgene changes the length of the DNA between the primers to over 6kb.

PCR reactions 4 (5'BB) and 5 (3'BB) detect the presence of extra copies of the targeting plasmid. The primers are complimentary to regions of the plasmid backbone outside of the homology arms and should not be present in correctly targeted clones without random integration of extra plasmid copies (Figure 5-14c).

To speed up the genotyping process we introduced some changes to the PCR set up of the reactions 1 and 2. We used direct tissue gDNA extraction kit by Thermofisher. The kit is designed for rapid gDNA extraction form a small colony fragment enabling us to start the genotyping process as soon as the colonies were picked. Additionally, Phire polymerase included in the kit has a faster amplification rate compared to Long Amp polymerase (20s/kb vs 50s/kb), which shortens the PCR reaction time significantly. Finally, Phire master mix contains loading dye which speeds up the loading of gel electrophoresis. Those features have significantly shortened the time needed for genotyping of the picked clones and most have been screened by the time they were expanded beyond 24wp format. This has saved significant amount of time and resources and made the process of clones handling less labour intensive. The use of a different polymerase required optimisation of primer annealing temperatures. A range of temperatures for gradient PCR was selected on the basis of recommendations in the Phire kit. PCR programme was adjusted for the Phire polymerase. For 5'INT all the selected annealing temperatures produced single band of correct

length (Figure 5-15a) The temperature of 65.1°C was selected for genotyping. For 3'INT primers, the annealing temperatures of 69.4°C and 71.2°C produced single bands of correct length. The annealing temperature of 69.4°C was used for genotyping (Figure 5-15b). Additionally, as colony fragments collected for gDNA extraction were small, there was a risk that not enough gDNA was extracted for a PCR reaction to occur. To ensure that enough gDNA was present in each sample, the Phire kit supplies a pair of primers amplifying a highly conserved non-coding region upstream of SOX21 gene in a wide range of vertebrate species. Based on kit recommendations, we selected two annealing temperatures to check the primers' performance. Both temperatures amplified a band of correct length for all samples (Figure 5-15c) and the temperature of 72.6°C was selected for screening.

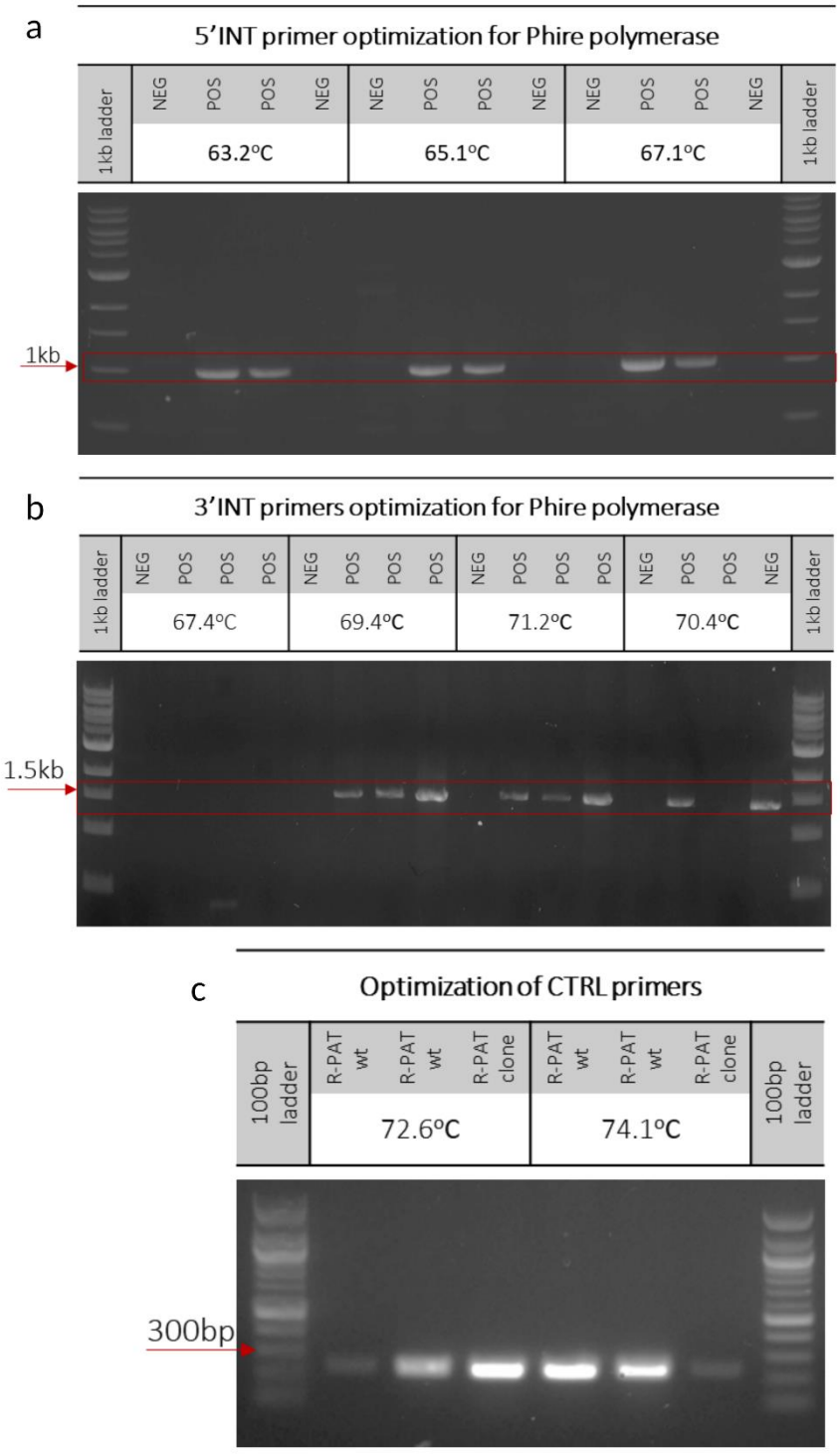

Figure 5-15 Optimization of primer annealing temperature for Phire polymerase.
a) Optimization of genotyping 5'INT primer pair; b) Optimization of genotyping 3'INT primer pair; c) Optimization of CTRL primers.

To streamline the genotyping process, only clones showing amplification in 5'INT and 3'INT reactions were taken for further screening.

R-PAT M cells were targeted using six different constructs with shRNAs against our TFs of interest. Genotyping was performed after clones were picked form each targeting and only positive clones (homozygous or heterozygous) were expanded and banked for further study.

Genotyping of the MsiKD_TBX clones showed 13/48 clones with OPTiKD cassette integration at the AAVS1 locus (Figure 5-16) . Further screening showed that 8/13 were homozygous and 5 out of the 8 homozygous clones had no extra copies of randomly integrated plasmid (Figure 5-17).

Genotyping of MsiKD_MAF clones showed 10/24 of clones with OPTiKD cassette integration at the AAVS1 locus. Further screening showed 7/10 were homozygous and of those 7, 5 had no random plasmid integration (Figure 5-18).

Genotyping of MsiKD_GATA clones showed 8/17 clones with OPTiKD cassette integration at the AAVS1 locus. Further screening showed that 4/17 were homozygous and of those 4 only 1 had no extra copies of randomly integrated plasmid (Figure 5-19).

Genotyping of MsiKD_HEY clones showed 13/23 clones with OPTiKD cassette integration at the AAVS1 locus. In this targeting, all clones showed presence of randomly integrated plasmid. Locus screen was not performed (Figure 5-20).

Genotyping of MsiKD_MSX clones showed 8/13 clones with OPTiKD cassette integration at the AAVS1 locus. One clone was homozygous. The screen for random plasmid integration was not performed for this targeting (Figure 5-21).

Genotyping of SiKD_SCR clones (Scramble CTRL) showed 13 clones with OPTiKD cassette integration at the AAVS1 locus. 8/13 clones were homozygous and 4 of those 8 had no extra copies of the plasmid (Figure 5-22).

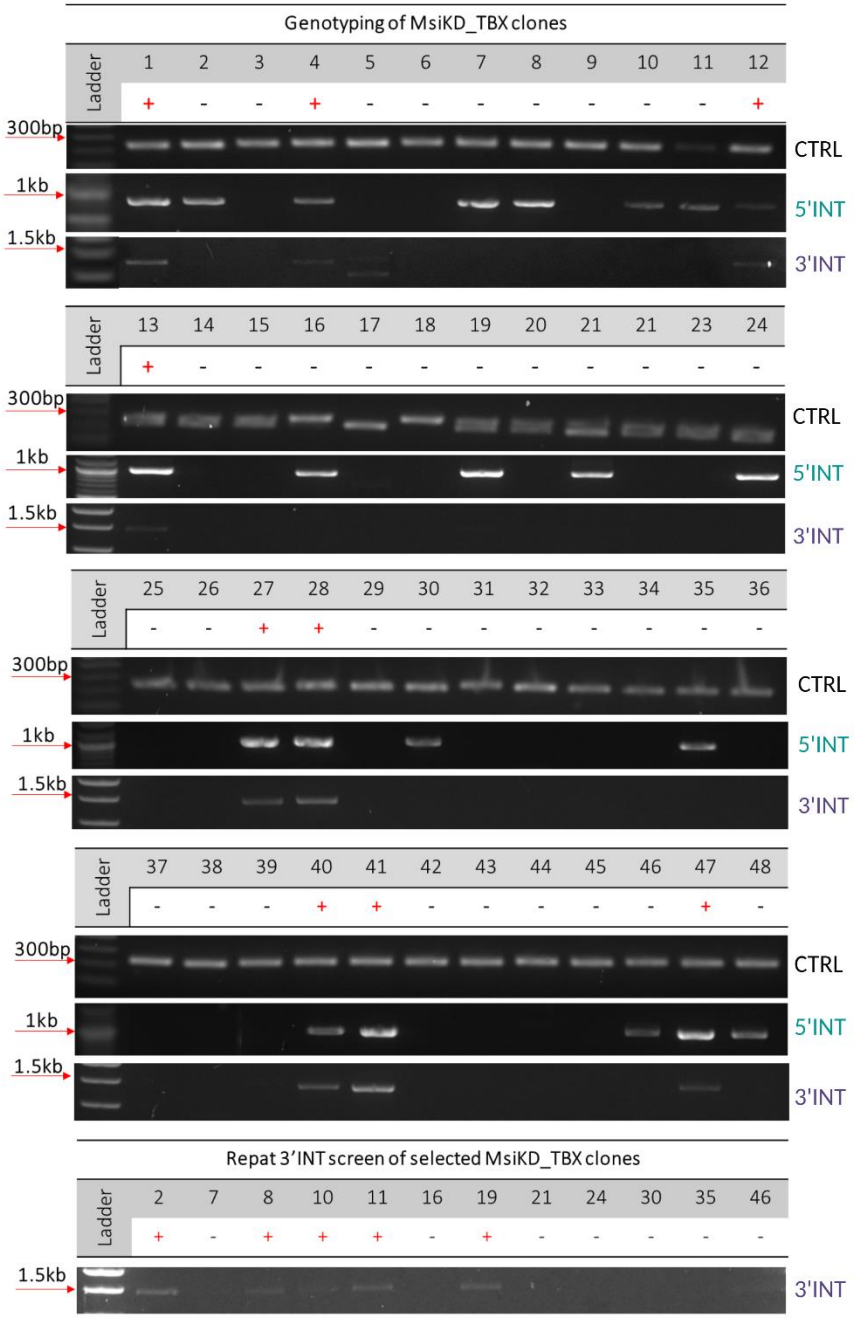

Figure 5-16 Gel images of PCR reactions for genotyping of MsiKD_TBX clones.
22 clones were positive for 5'INT integration and 14 for 3'INT integration. 13 clones were positive for both 5'INT and 3'INT and were taken for further genotyping.

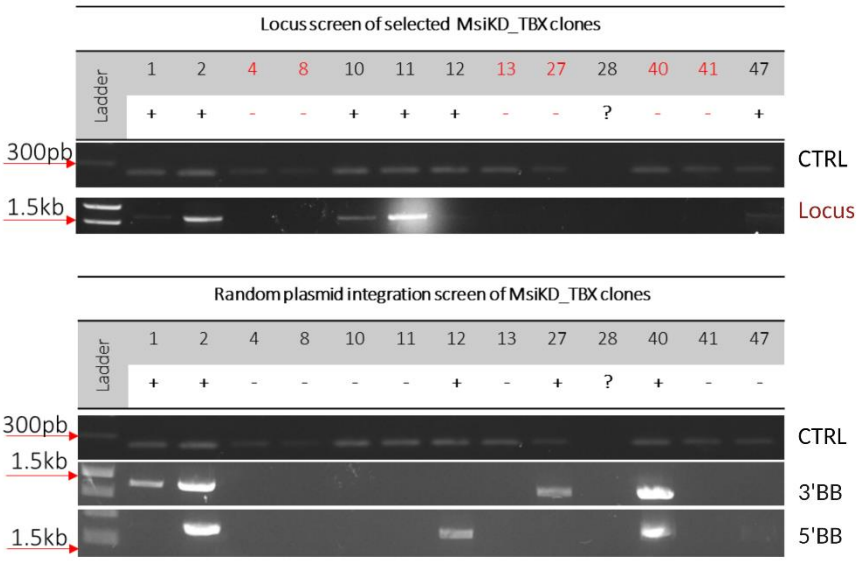

Figure 5-17 Gel images of PCR reactions for genotyping of MsiKD_TBX clones.
Out of 13 clones taken for further genotyping, 7 showed no band in the Locus screen indicating homozygosity and 8 had no random plasmid integration. 5 clones were identified as homozygous and clean of random plasmid integration: 4,8,13,41,47.

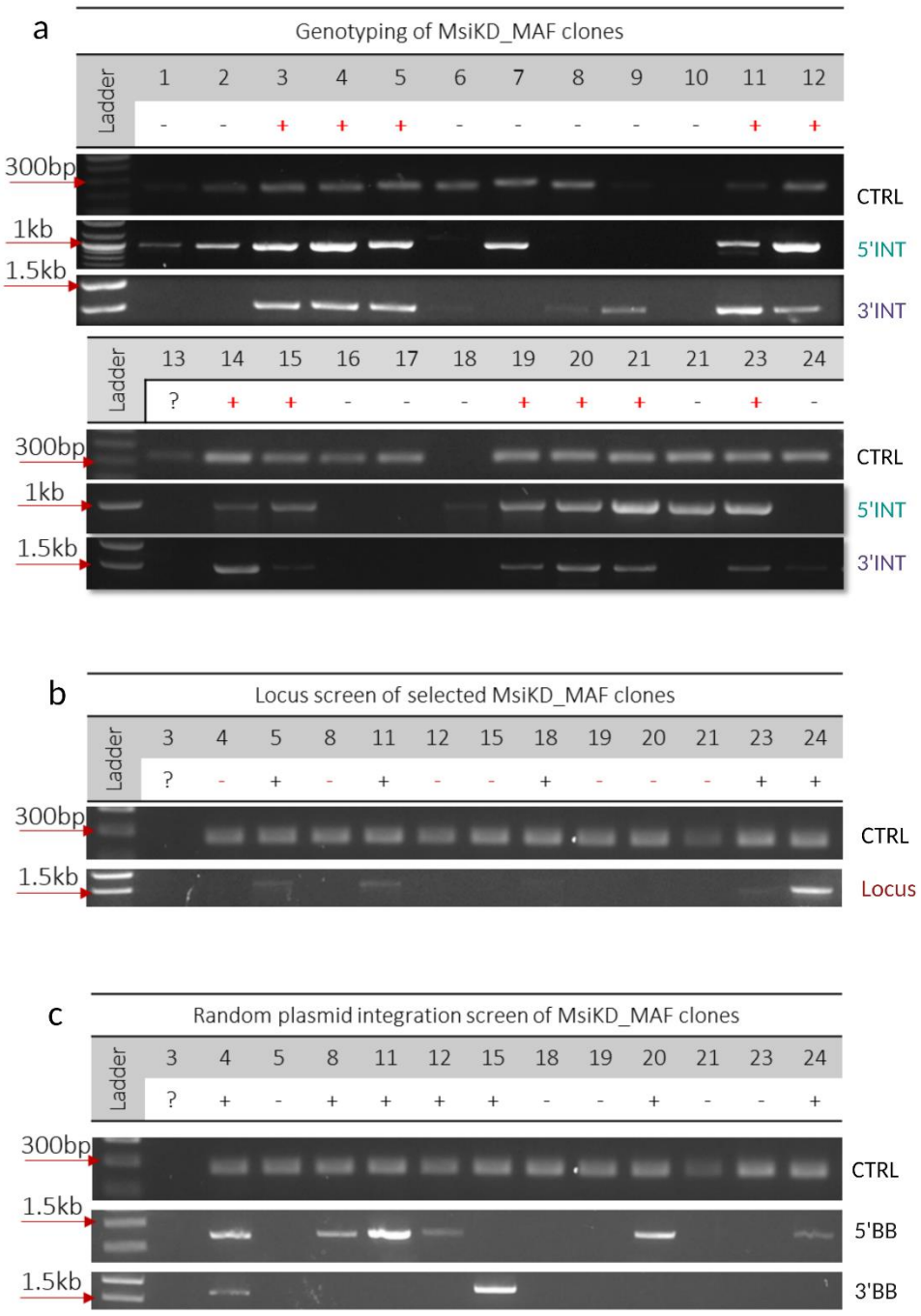

Figure 5-18 Gel images of PCR reactions for genotyping of MsiKD_MAF clones.
a) 16 clones produced a band in 5'INT screen and 14 in 3'INT screen, 10 produced a band in both screens; b) 13 clones were taken for further genotyping and 7 produced no band in the Locus screen indicating a homozygous clone; c) 5 clones showed no extra copies of randomly integrated plasmid. Clones 18,19 and 21 we homozygous and with no extra plasmid copies.

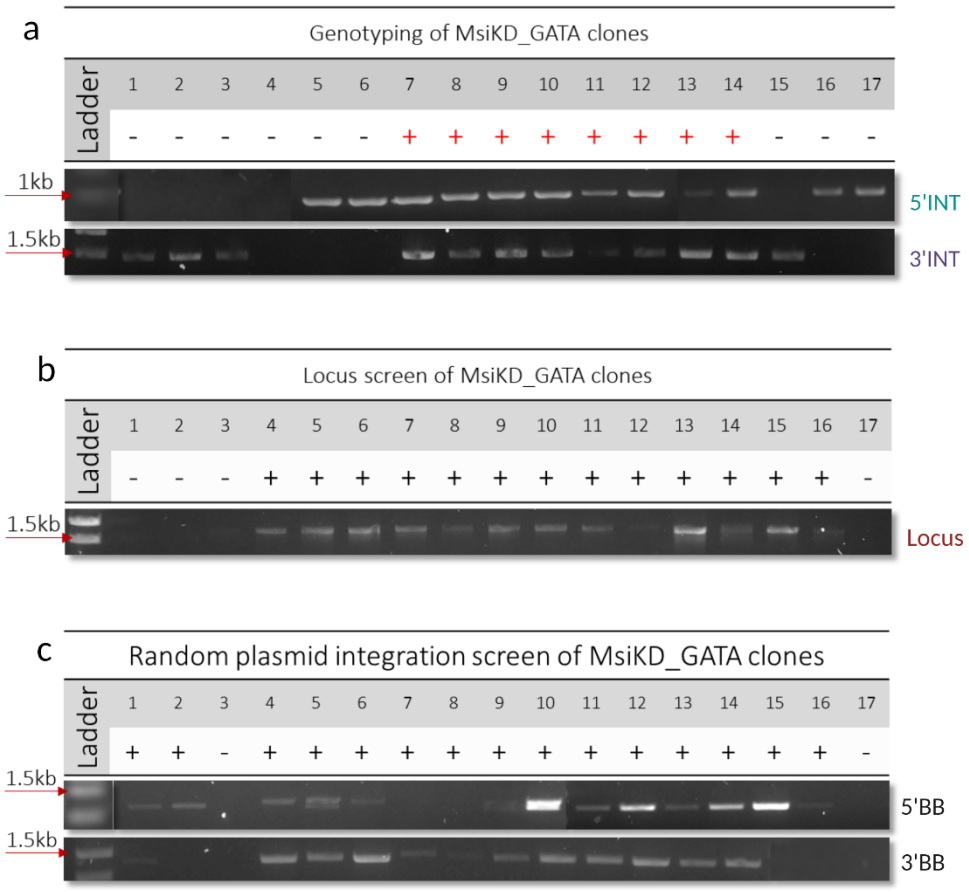

Figure 5-19 Gel images of PCR reactions for genotyping of MsiKD_GATA clones.
a) 12 clones showed a band in the 5'INT screen and 12 in 3'INT screen. 8 clones were double positive; b) 4 clones showed no amplification in the locus screen; c) 2 clones had no extra copies of randomly inserted plasmid. Clone 17 was the only homozygous clone with no randomly integrated plasmid.

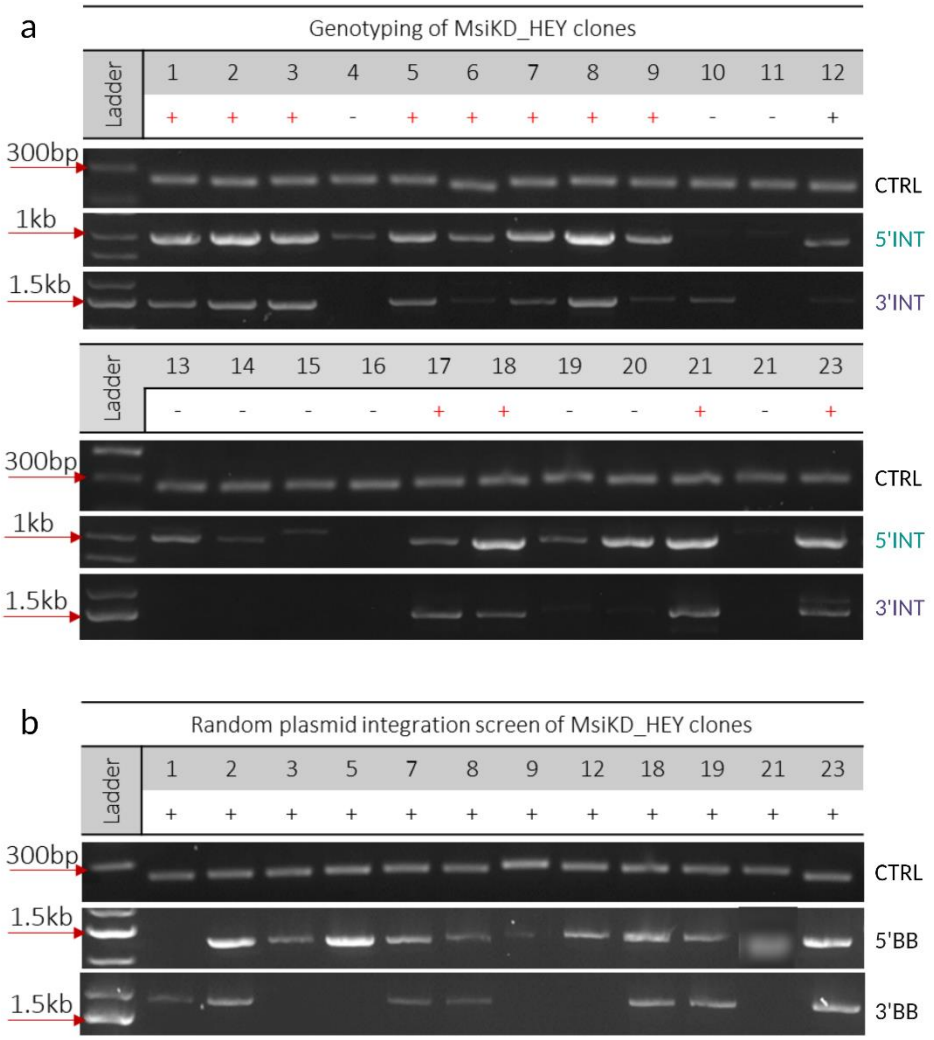

Figure 5-20 Gel images of PCR reactions for genotyping of MsiKD_HEY clones.
a) 19 clones showed a band in the 5'INT screen and 14 in the 3'INT screen. 13 were double positive; b) None of the picked clones was clean of randomly integrated plasmid.

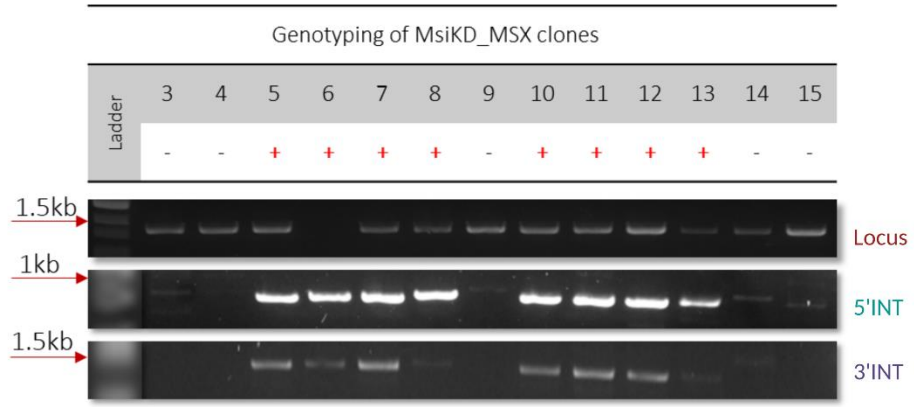

Figure 5-21 Gel images of PCR reactions for genotyping of MsiKD_MSX clones.
8 clones produced a band in both 5'INT and 4'INT screen. Clone 6 produced no band in the Locus screen and was the only homozygous clones in this targeting.

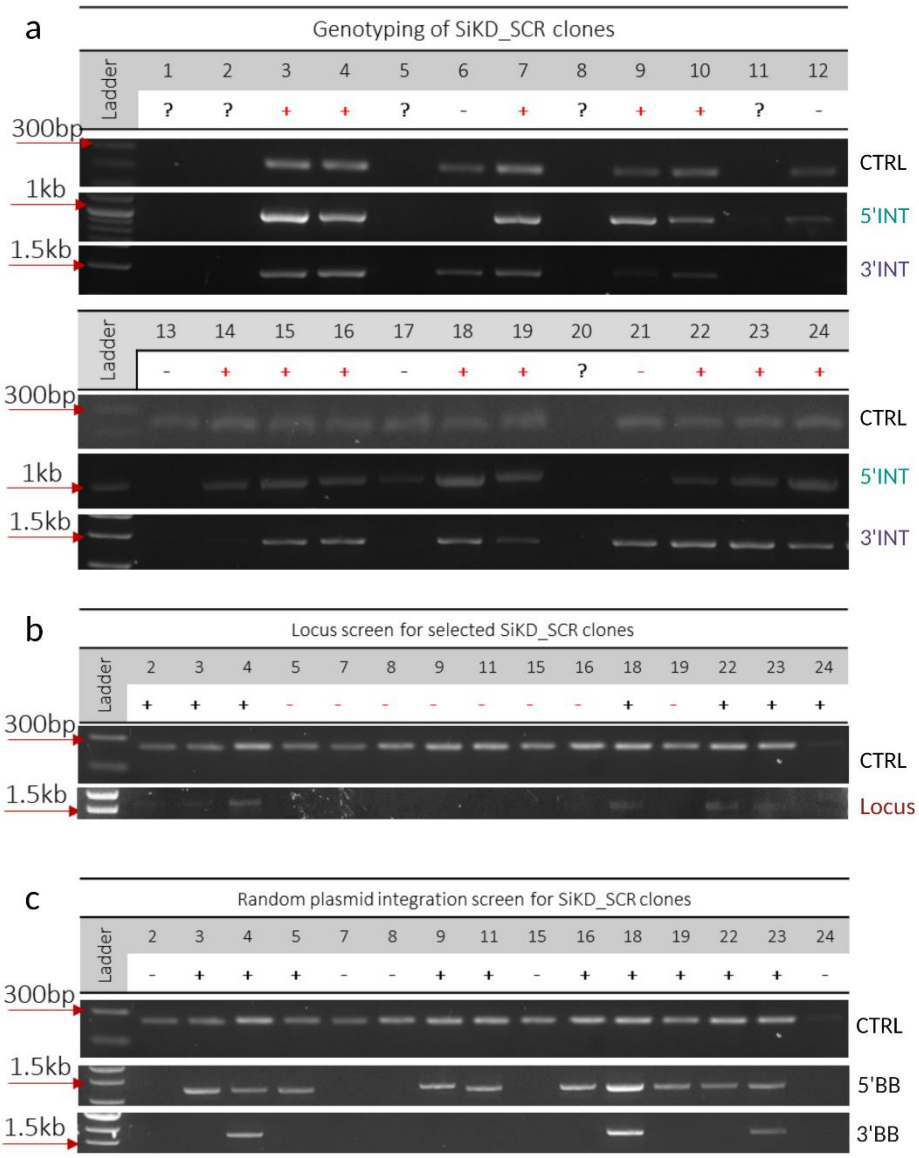

Figure 5-22 Gel images of PCR reactions of Scramble clones.
a)15 clones produced a band in the 5'INT screen and 14 in the 3'INT screen. 13 clones were double positive; b) 8 clones produced no band in the locus screen; c) 5 clones had no extra copies of randomly integrated plasmid. Clones 7,8,15 and 24 were homozygous with no extra plasmid copies.

In summary, the slightly modified Bertero protocol enabled rapid creation (approx. 4 weeks) of cell lines in which conditional manipulation of TFs expression is possible. On average, 95% of picked clones survived the picking process and 57% of picked clones were identified with correct insertion of the OPTiKD cassette into the AAVS1 locus. Of those correctly targeted clones, 27%

were identified as homozygous but only on average 16% were clean of extra copies of the plasmid backbone within the genome (Figure 5-23).

Where possible, homozygous clones without random integration were taken for validation of the knockdown experiments. In some cases, heterozygous clones were also used for validation. For MsiKD_GATA and MsiKD_HEY targeting, clones with random integration of the plasmid were also tested in further experiment due to either no clone clean of random integration or clones that were not checked for random integration of the plasmid.

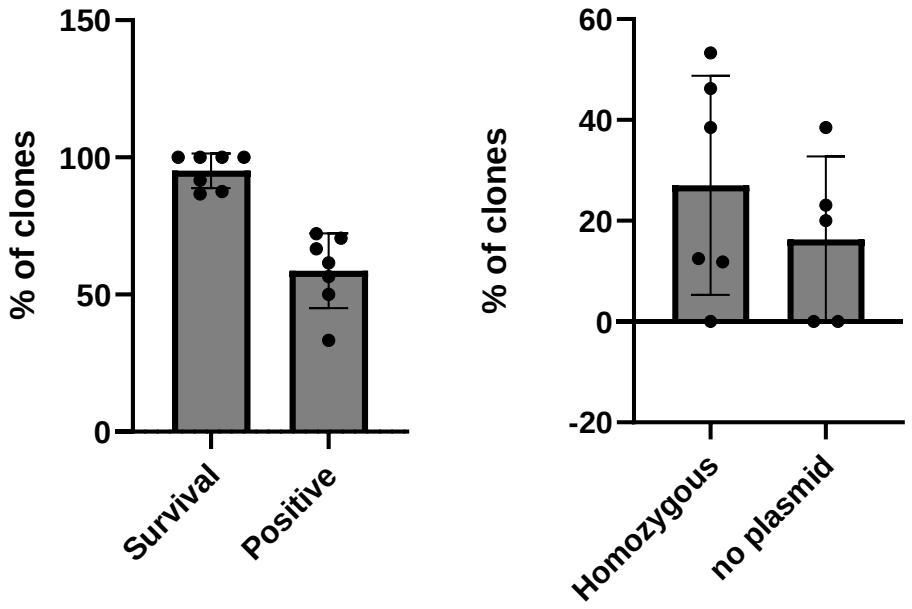

Figure 5-23 Summary of the targeting strategy for creating inducible hiPSC cell lines. a) Percentage survival calculated as number of clones that survived 5 days after manual dissection. Percentage positive clones calculated as the number of picked clones with correctly inserted OPTiKD cassette. Mean +/-SD. (n=7); b) Positive clones were tested for homozygosity and presence of extra copies of the plasmid backbone in the genome. Mean +/- SD (n=5).

5.3.6. Validation of the MsiKD cell lines

Where possible, three clones with each construct were selected to identify one with the best knockdown efficiency. As most of our candidate TFs are not expressed at the pluripotent stage, each clone was differentiated to the liver

progenitor cell stage at which point the knockdown efficiency was checked. Production of shRNAs was induced by addition of tetracycline (TET) at the last day of foregut differentiation. TET was continuously added to cell culture media for the duration of liver progenitor specification. Cell morphology during the differentiation was assessed and RNA was collected to check the knockdown efficiency by qPCR. Additionally, AFP or TTR levels were checked as an initial screen of knockdown effects on the specification of hepatocytes.

Three MsiKD_GATA clones were selected for screening: clone 8, 10 and 13. There were no morphological differences during the specification of LPCs in any of the clones. Analysis of the knockdown efficiency by qPCR has shown no knockdown or changes in the expression levels of AFP or TTR genes (Figure 5-24, Figure 5-25, Figure 5-26). As no knockdown was observed, no further experiments were conducted on the MsiKD_GATA clones.

Three MsiKD_HEY clones were selected for screening: clone 1, 3 and 9. Clone 1 showed no morphological differences between -TET and +TET cells (Figure 5-27a). The qPCR analysis of mRNA expression showed a good knockdown of HEY1 gene with 74%, 81% and 76% decrease in HEY1 mRNA levels on hepatic specification days 1, 2 and 4, respectively. HEY2 mRNA levels showed significantly decreased levels on day1 of hepatic specification (48%). TTR mRNA levels showed 68% reduction on day 1 of specification but remained unaffected as the specification continued (Figure 5-27b).
Clone 3 showed no morphological differences between -TET and +TET cells (Figure 5-28a). The qPCR analysis showed a good knockdown of both HEY1 and HEY2 mRNA levels, with a stronger knockdown of HEY1 mRNA of 61%,75% and 67% on days 1, 2 and 4 of specification, respectively. HEY2 mRNA levels decreased by 42%, 21% and 37% on day 1,2 and 4, respectively. There was no statistically significant reduction in the levels of the TTR mRNA expression (Figure 5-28b). Although there is a significant knockdown of both genes, the initial screen indicated that there was no disruption of the hepatic specification as shown by unchanged TTR levels. This is, however, only an initial screen and further experiments are necessary to determine if knockdown of HEY genes

affects hepatic specification. Additionally, the success of the knockdown needs to be checked on the protein level by Western blot analysis. MsiKD_HEY clone 9 showed no morphological differences between the -TET and +TET cells (Figure 5-29a). The qPCR analysis of the mRNA levels showed HEY1 knockdown on day 2 only of hepatic specification (68% reduction). HEY2 levels were unaffected. There was a small reduction in TTR mRNA levels on day 1 of specification (Figure 5-29b).

MsiKD_HEY clone 3 was identified as the best candidate for further study of the possible role of HEY1 and HEY2 genes in the specification of hepatoblasts.

Three MsiKD_MSX clones were tested: 5, 8 and 11. There was no differences in morphology upon the induction of shRNA against MSX1 and MSX2 secretion in any of the selected clones (Figure 5-30a, Figure 5-31a, Figure 5-32a). There were virtually no differences in the mRNA expression levels of MSX1, MSX2 or TTR in any of the clones (Figure 5-30b, Figure 5-31b, Figure 5-32b). Clone 8 showed 46% reduction in mRNA levels of MSX1 on day 4 of hepatic specification, but on days 1 and 2 it showed a reverse trend where MSX1 levels were significantly higher in the TET induced cells vs TET non-induced (Figure 5-31b). The screen has not identified any suitable clone for further experiments.

Of the three MsiKD_TBX clones tested: 4,11 and 41, clone 4 has showed morphological differences during the differentiation (Figure 5-33a). From day 2 of liver progenitor specification, the cells show a less compact monolayer which remained so until the end of specification at day 4 (Figure 5-33b). QPCR analysis of mRNA levels showed no significant difference in the mRNA levels of TBX2 gene but a good knockdown of TBX3 gene of 74%, 59% and 87% at day 1, 2 and 4 of liver specification, respectively. TBX20 mRNA levels also showed decreased expression of 78%,57% and 34% on day 1, 2 and 4 respectively, although the reduction on days 2 and 4 did not reach statistical significance. The levels of TTR mRNA were also affected. On day 1 of specification there was a 84% reduction in the mRNA levels of TTR gene and on day 4 the levels were reduced by 42% (Figure 5-33). This is an indication that the knockdown of the TBX3 and TBX20 can possibly influence the molecular network governing hepatic specification.

MsiKD_TBX clone 8 showed no difference in morphology between -TET and + TET cells (Figure 5-34a). TBX2 mRNA levels were unaffected but there was a decrease in TBX3 mRNA levels of 59%, 28% and 67% on days 1,2 and 4, respectively. TBX20 mRNA levels were also affected on day 1 and 2, by 64% and 46% respectively. There was a reduction in TTR mRNA levels on days 1 and 2 by 50% and 20%, respectively (Figure 5-34b). This indicates a successful TBX3 and TBX 20 knockdowns that possibly affect early hepatic specification. Clone 41 showed no differences in morphology between -TET and + TET cells and there was no significant knockdown of any of the TBX genes (Figure 5-35). MsiKD_TBX clone 4 was selected for further experiments to understand the role of TBX TFs in the specification of liver progenitors.

Due to technical difficulties only one MsiKD_MAF clone was screed for the efficiency of the knockdown. Clone 21 showed no morphological differences between -TET and +TET cells (Figure 5-36a). QPCR analysis showed reduction in MAF mRNA levels of 46% and 31% on days 2 and 5 of specification, respectively. TTR mRNA levels were reduced by 39%, 30% and 40% on days 1, 2 and 4 of the specification, respectively (Figure 5-36b). The results indicate a functional knockdown of the MAF TF with possible consequence for the specification of liver progenitor cells and the clone can be taken for further experiments.

Lastly, the MsiKD_SCR cell line was created as a control cell line. In si/shRNA experiments scramble control is routinely used to determine that observed effects are due to knockdown and not the activation of the RNAi pathways within the cell. Additionally, in our system it can also determine that the observed effect is not due to the consequences of tetracycline addition on the physiology of the cell. We have differentiated one MsiKD-SCR clone to hepatic progenitor cells and observed no difference in cell morphology between -TET and +TET cells (Figure 5-37a). We also analysed the mRNA expression levels of all TFs selected from the RNA seq. screen for which the knockdown cells lines were created and observed no effect on the mRNA levels (Figure 5-37b). We also looked at the mRNA levels of major hepatic genes to ensure the addition of TET or induction of the RNAi pathways does not influence their levels and

observed no effect (Figure 5-38). These results support the assertion that any observed effect on the cell's morphology or mRNA levels expression in our inducible cell lines are not due to the tetracycline activity or RNAi pathway induction.

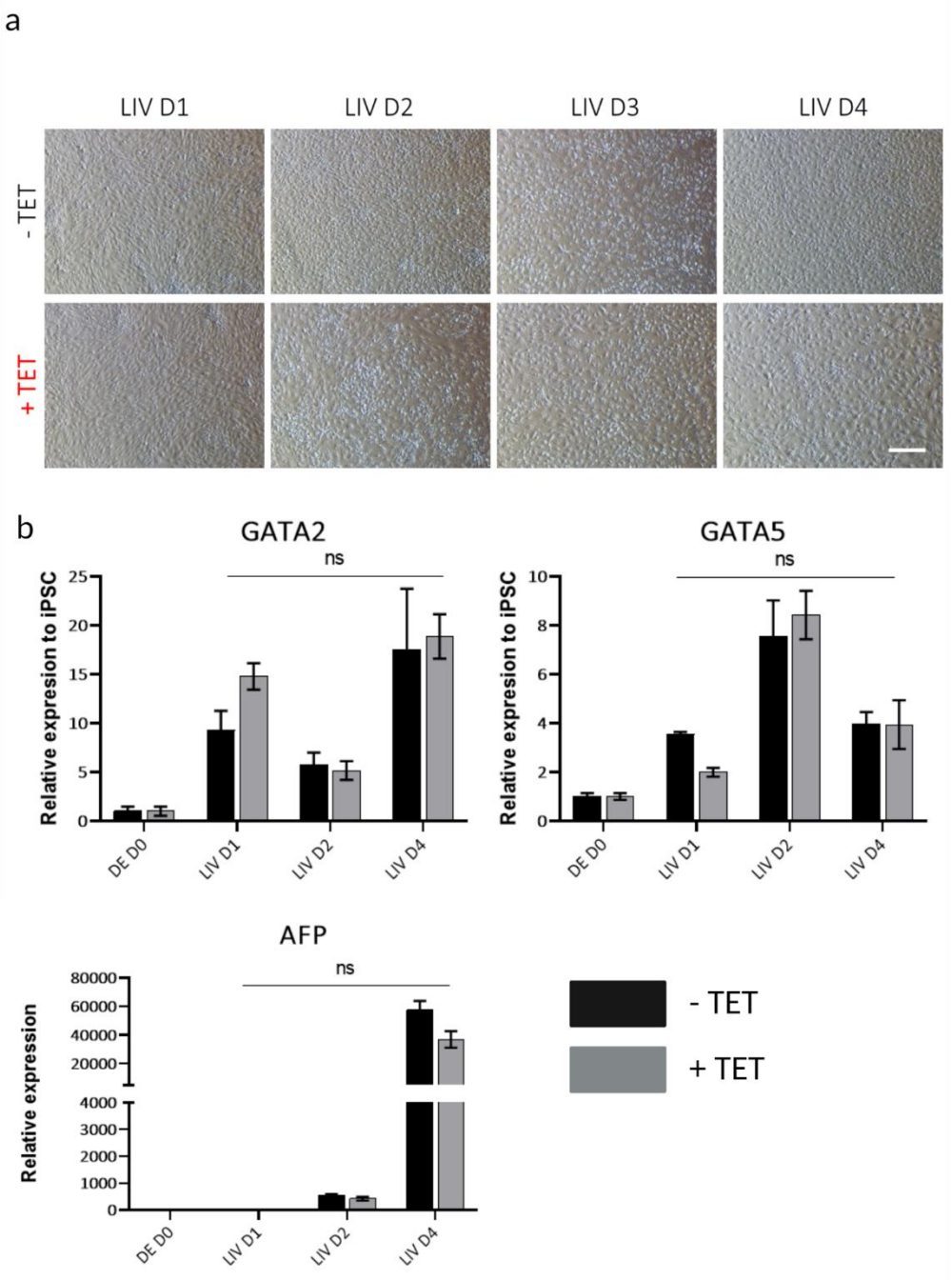

Figure 5-24 Knockdown analysis of R-PAT M MsiKD_GATA clone 8.
a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of GATA genes knockdown and AFP expression Mean ± SD (n=1).

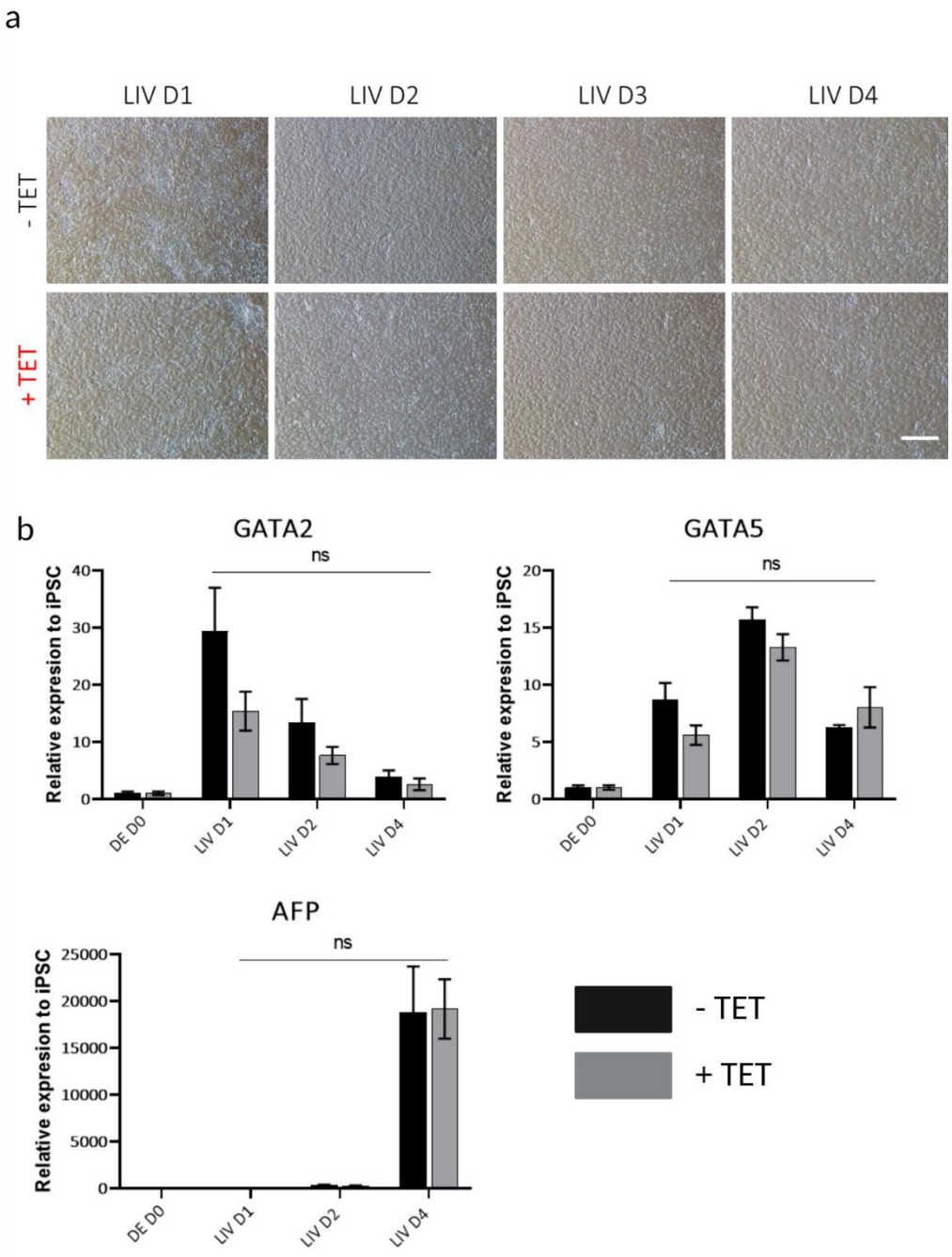

Figure 5-25 Knockdown analysis of R-PAT M MsiKD_GATA clone 10.
a) Microscopy images of the last stage of differentiation to liver progenitor cells; Scale bar = 200μm; b) qPCR analysis of GATA genes knockdown and AFP expression Mean ± SD (n=1).

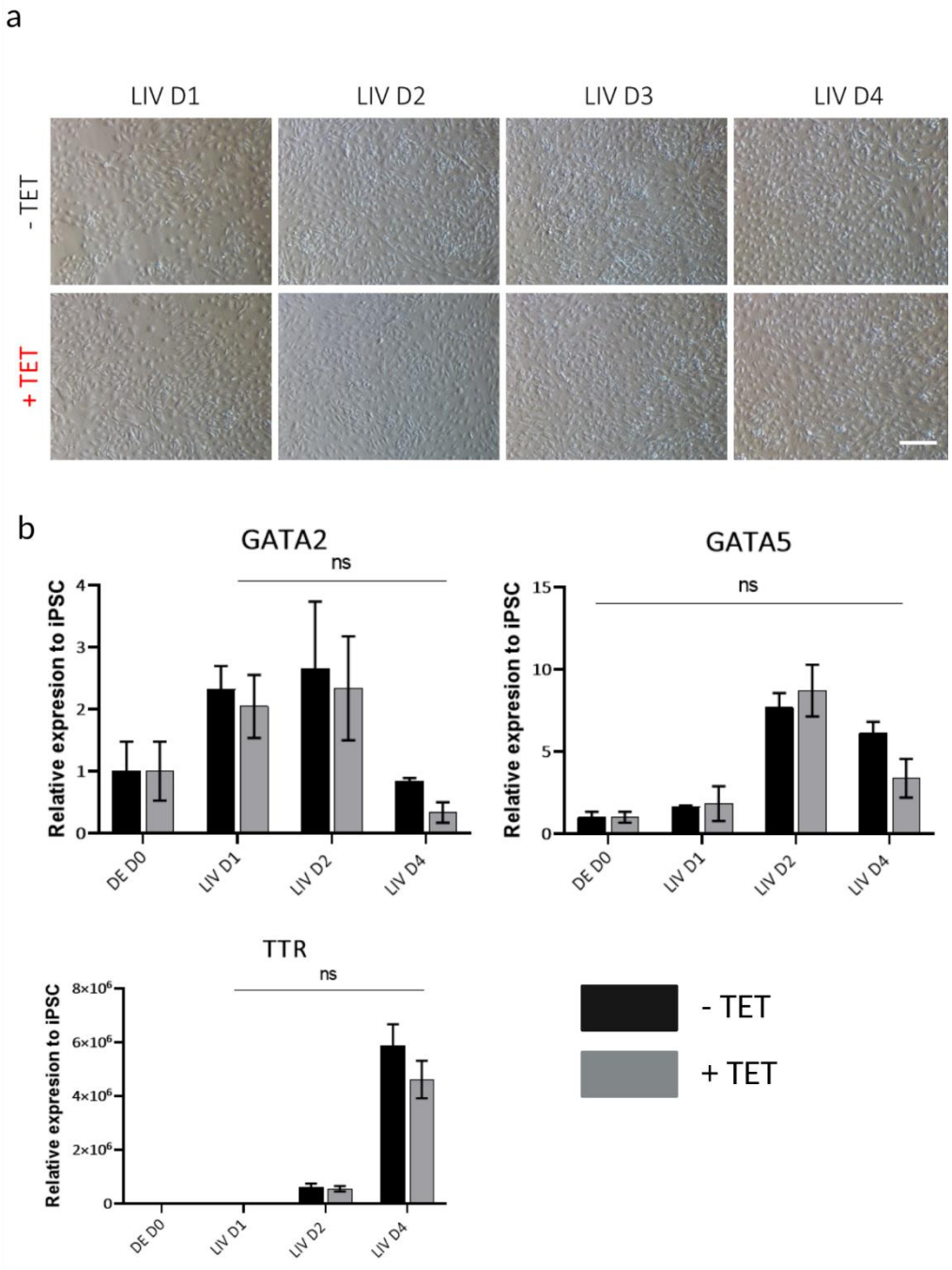

Figure 5-26 Knockdown analysis of R-PAT M MsiKD_GATA clone 13.
a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200µm; b) qPCR analysis of GATA genes knockdown and TTR expression. Mean ± SD (n=1).

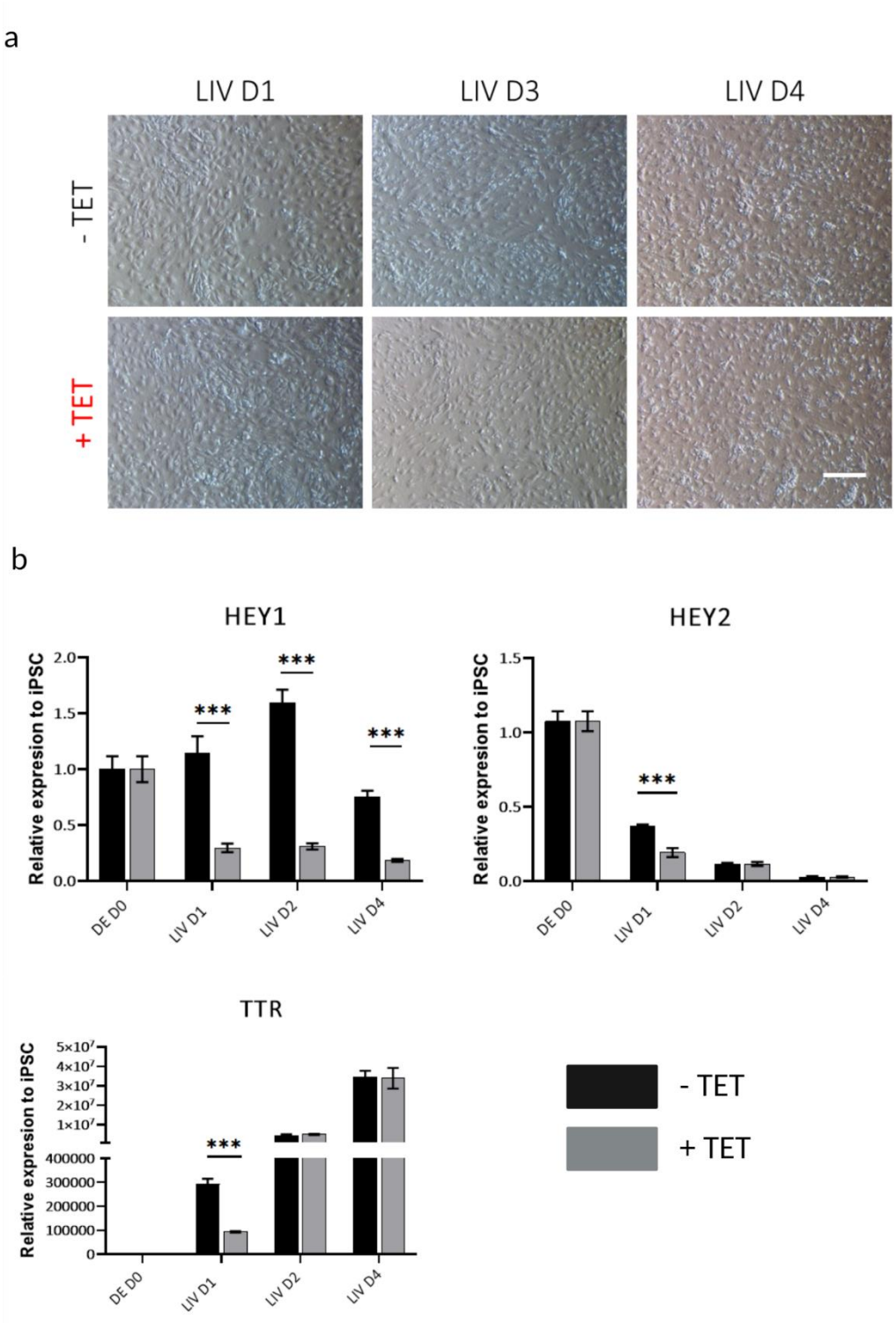

Figure 5-27 Knockdown analysis of R-PAT M MsiKD_HEY clone 1.
a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of HEY genes knockdown and TTR expression. Mean ± SD (n=1).

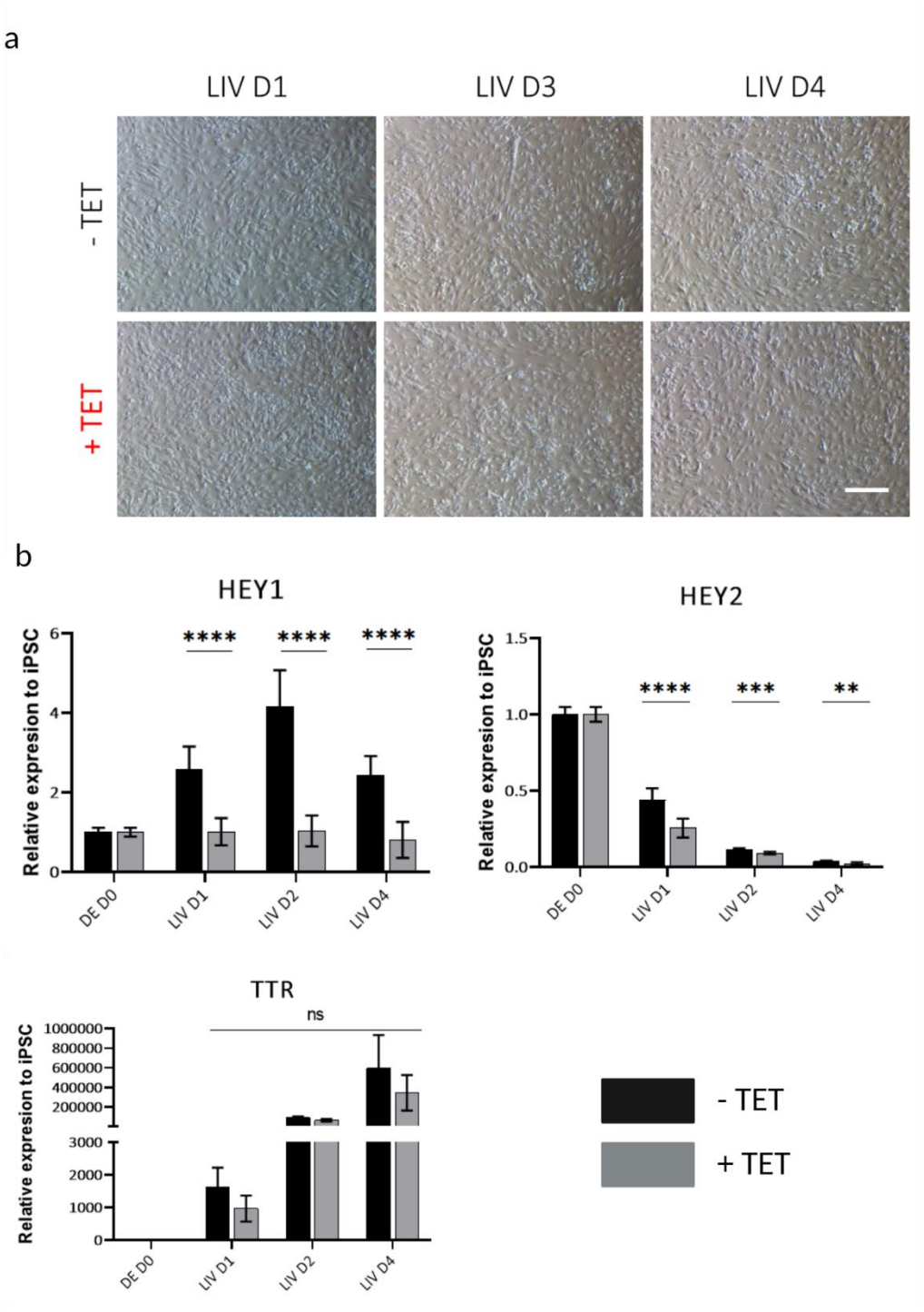

Figure 5-28 Knockdown analysis of R-PAT M MsiKD_HEY clone 3.
a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of HEY genes knockdown and TTR expression. Mean ± SD(n=3).

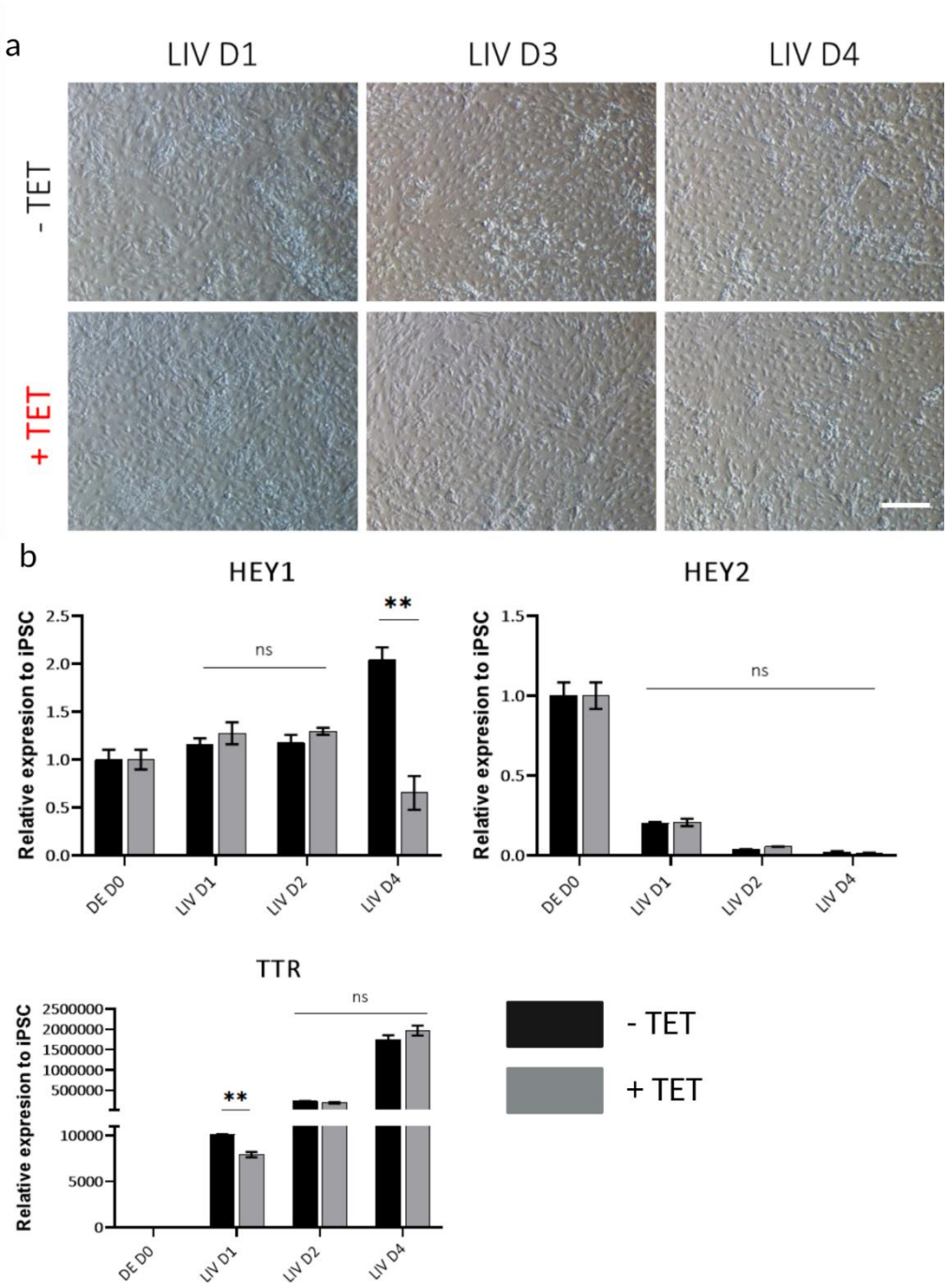

Figure 5-29 Knockdown analysis of R-PAT M MsiKD_HEY clone 9

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of GATA genes knockdown and TTR expression. Mean ± SD (n=1).

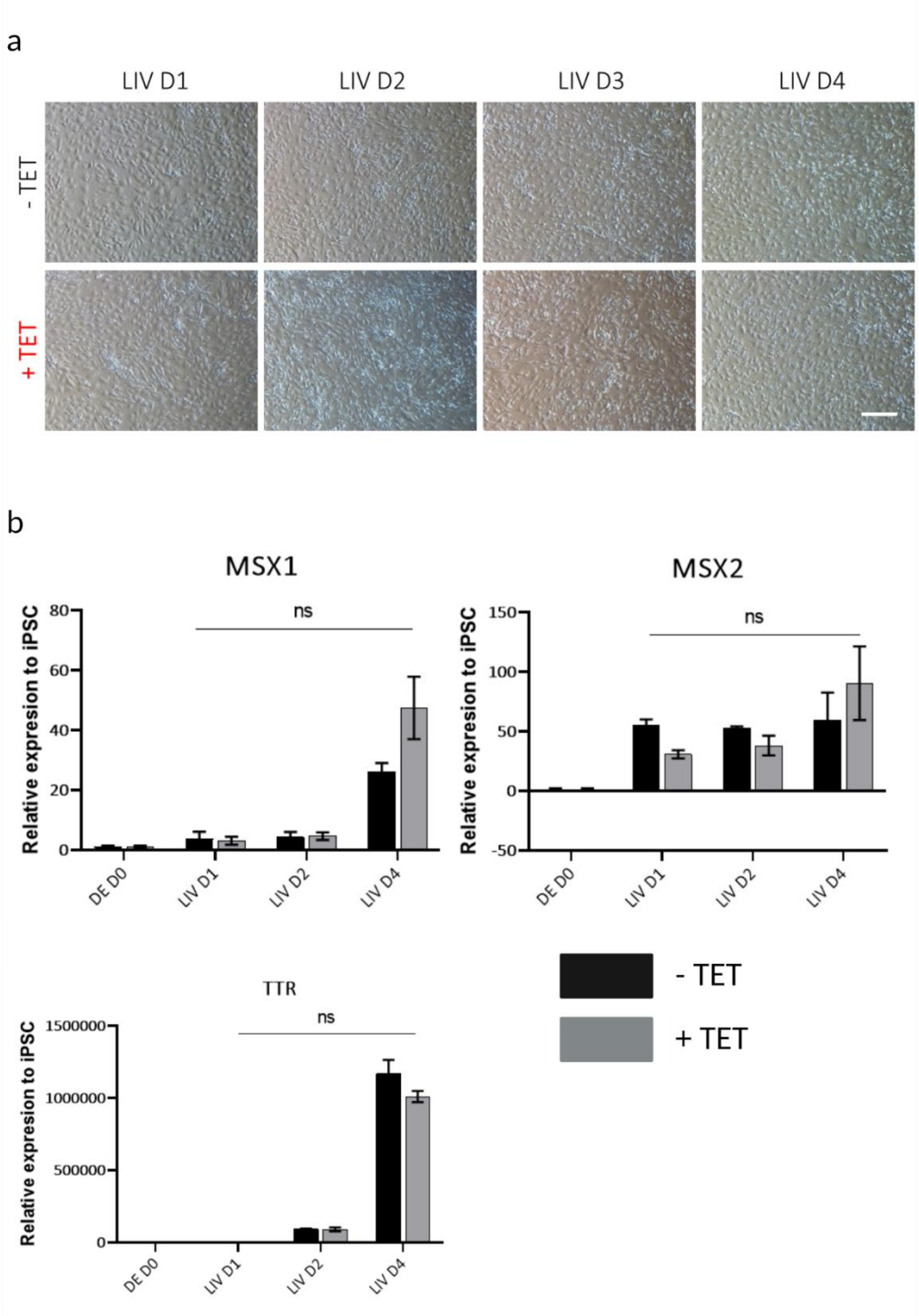

Figure 5-30 Knockdown analysis of R-PAT M MsiKD_MSX clone 5

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of MSX genes knockdown and TTR expression. Mean ± SD (n=1).

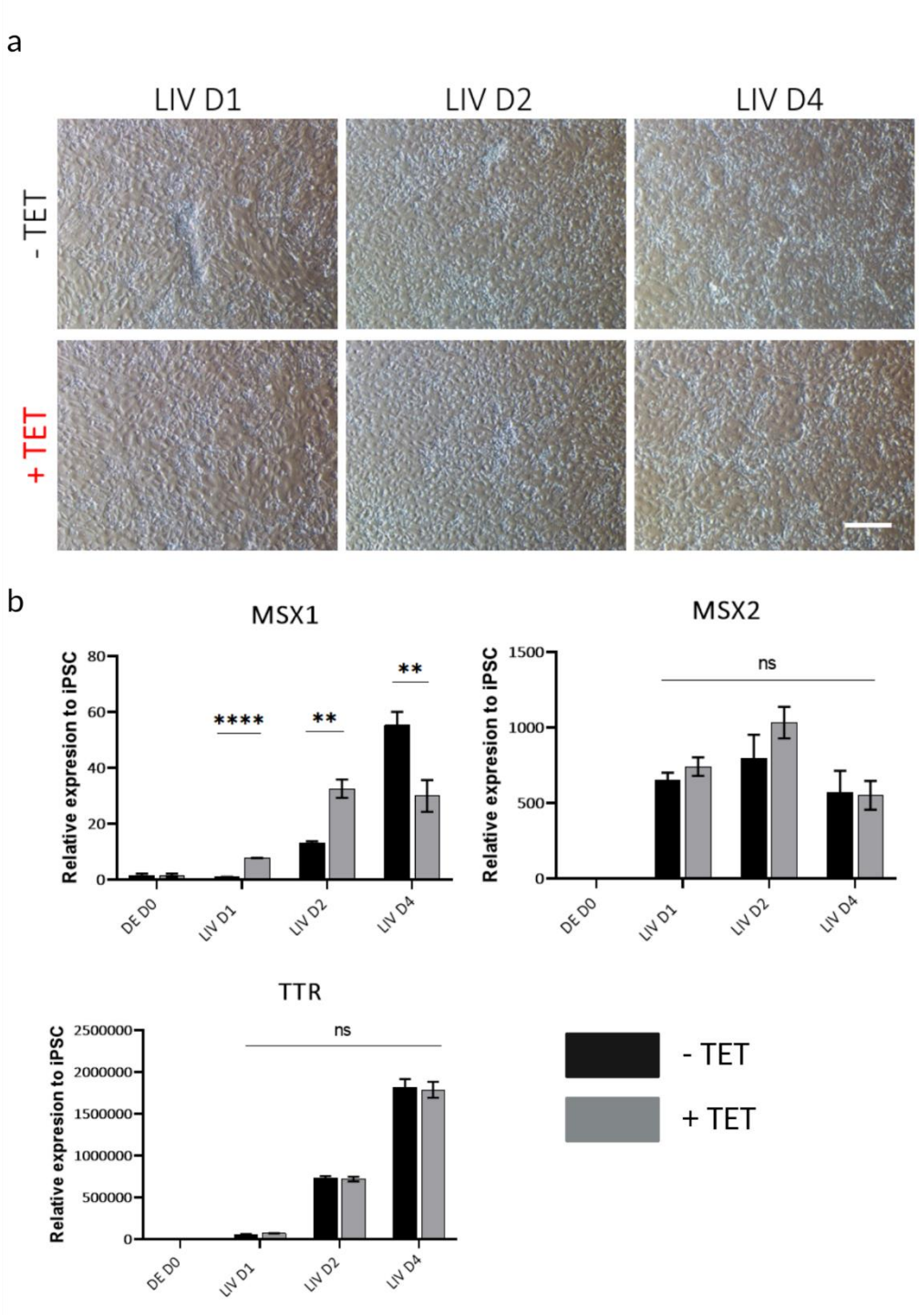

Figure 5-31 Knockdown analysis of R-PAT M MsiKD_MSX clone 8

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of MSX genes knockdown and TTR expression. Mean ± SD (n=1).

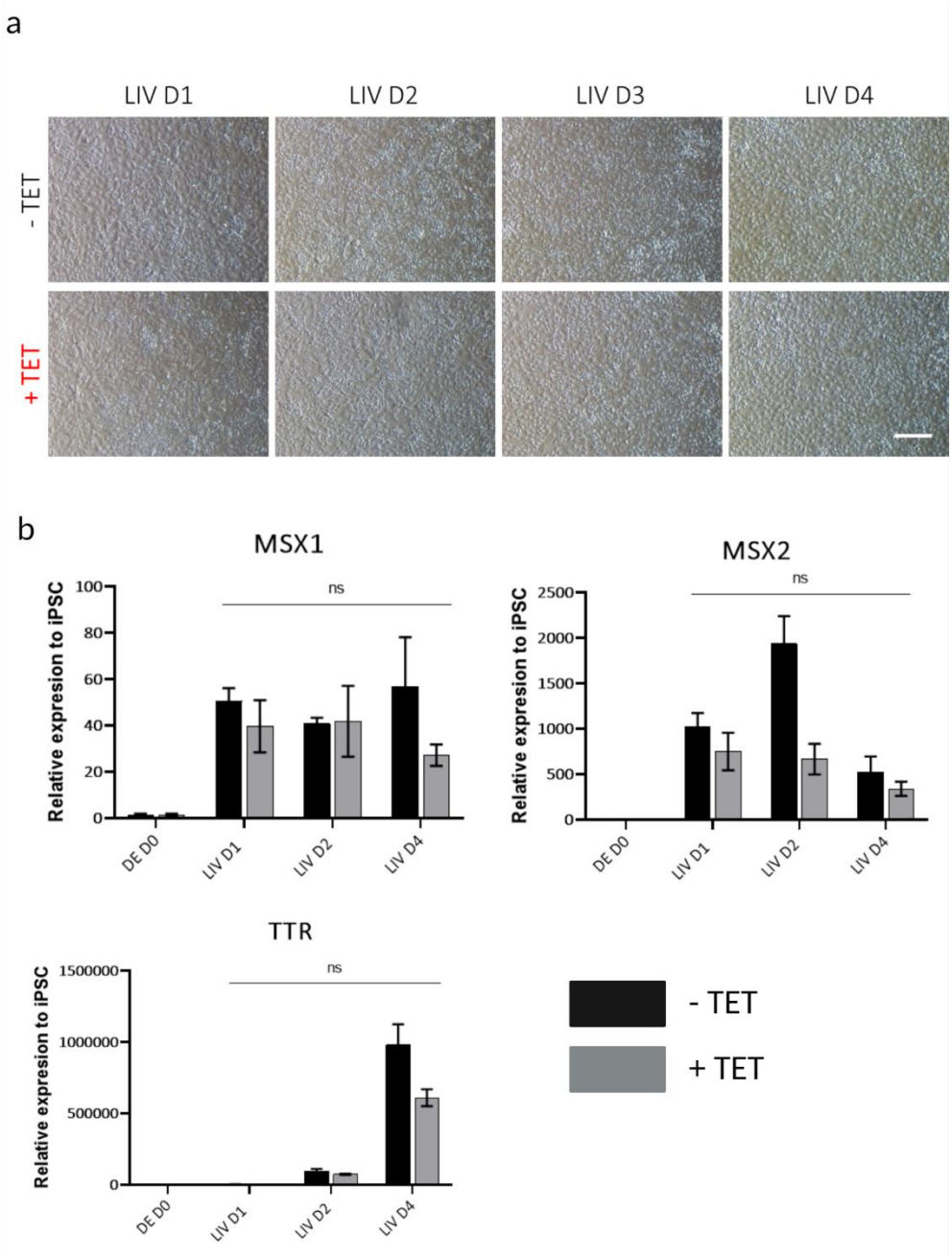

Figure 5-32 Knockdown analysis of R-PAT M MsiKD_MSX clone 11
a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of MSX genes knockdown and TTR expression. Mean ±SD (n=1).

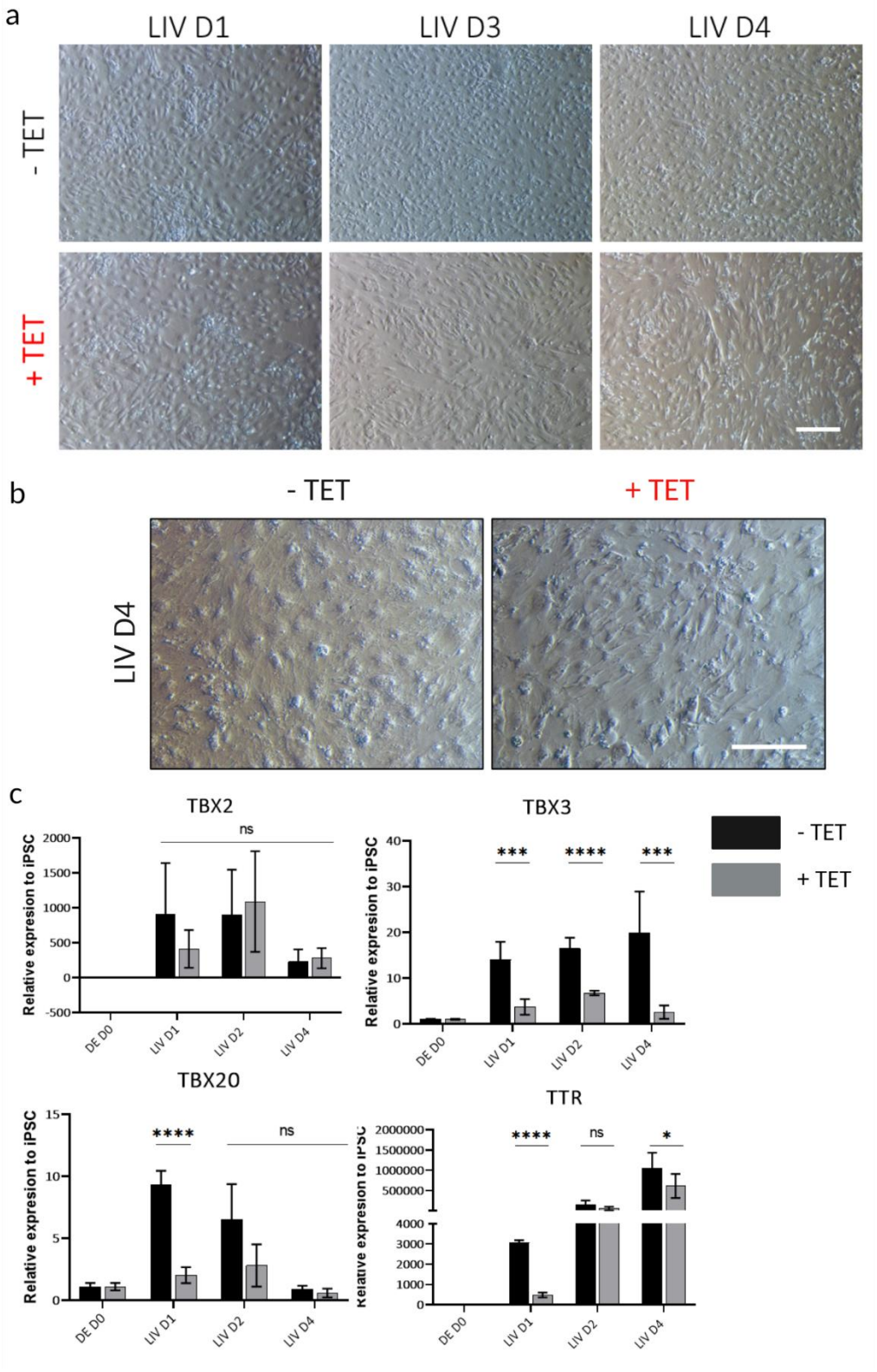

Figure 5-33 Knockdown analysis of R-PAT M MsiKD_TBX clone 4

a),b) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; c) qPCR analysis of TBX genes knockdown and TTR expression. Mean ±SD (n=2).

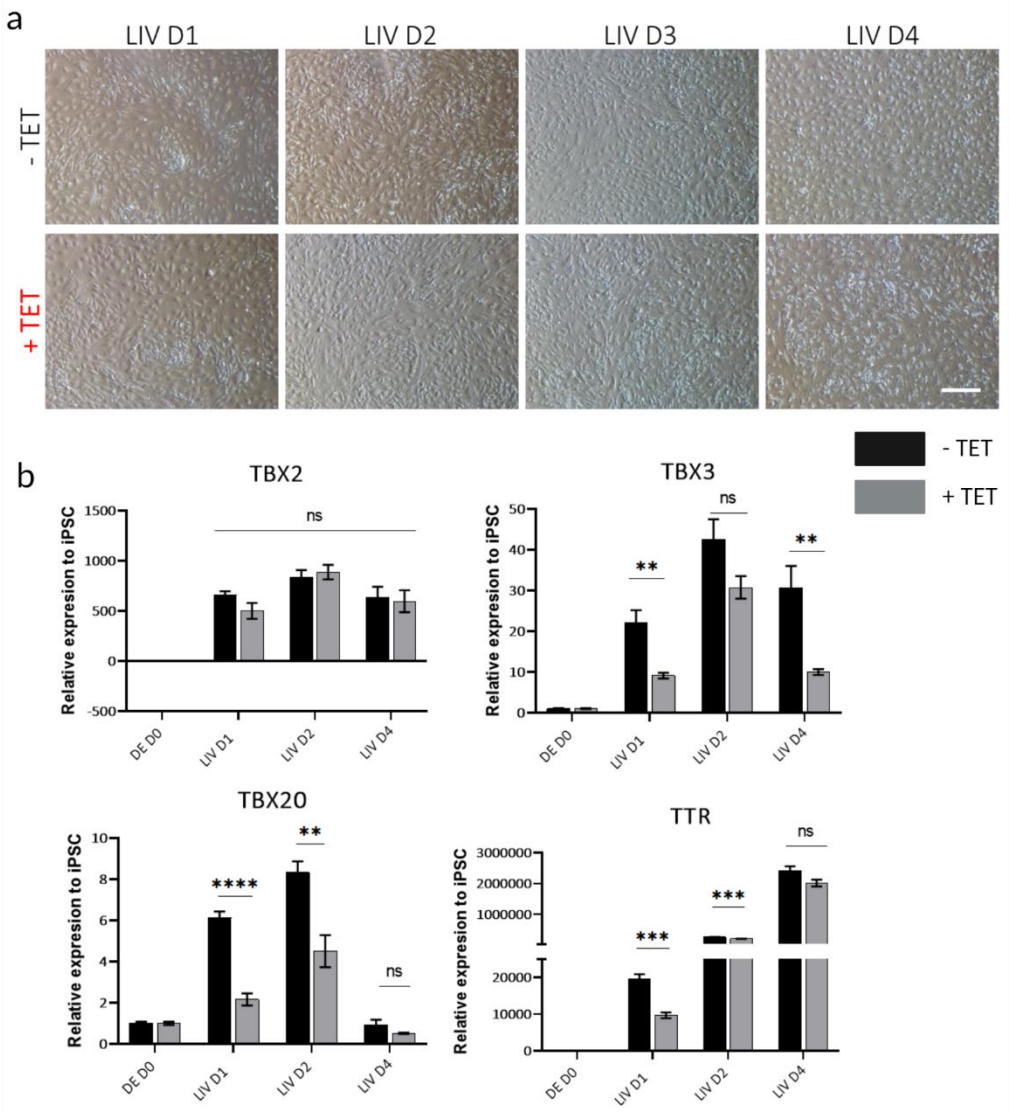

Figure 5-34 Knockdown analysis of R-PAT M MsiKD_TBX clone 8

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of TBX genes knockdown and TTR expression. Mean ±SD (n=1).

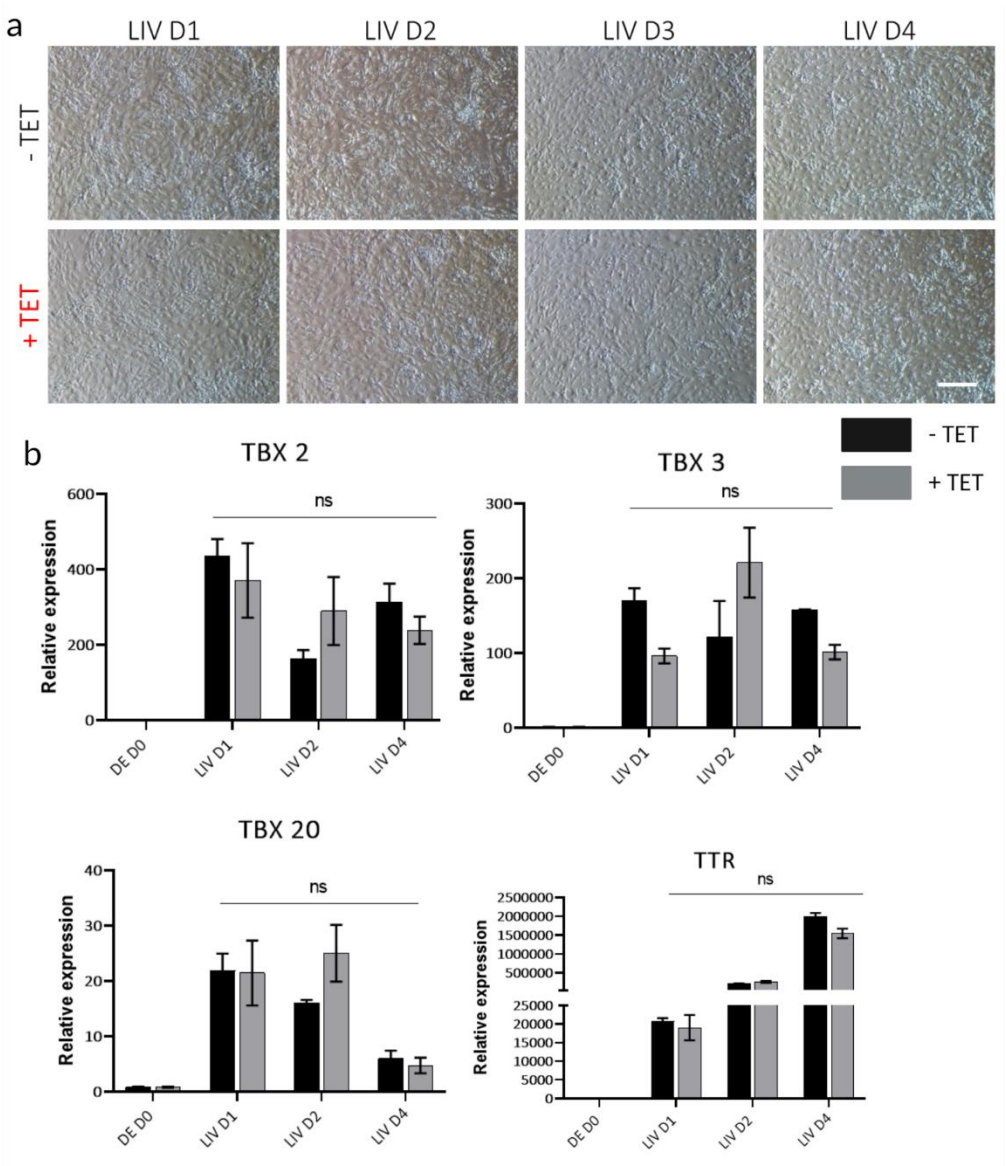

Figure 5-35 Knockdown analysis of R-PAT M MsiKD_TBX clone 41

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of MSX genes knockdown and TTR expression. Mean ±SD (n=1).

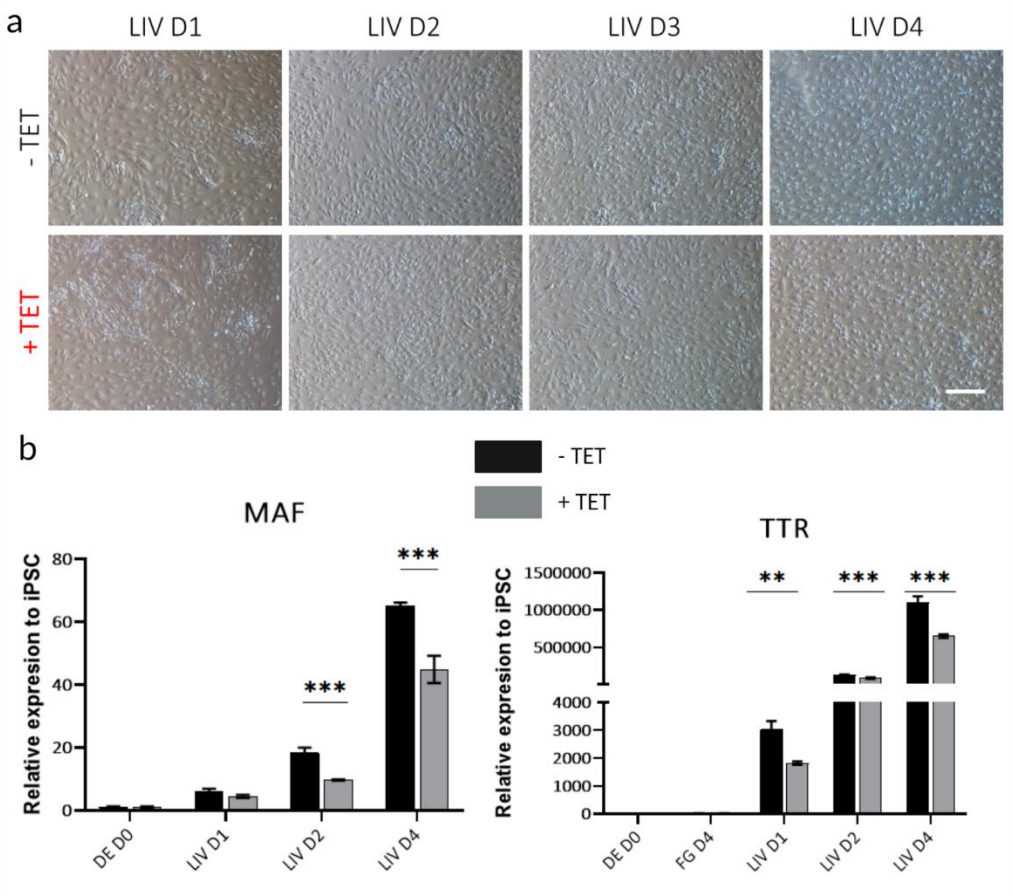

Figure 5-36 Knockdown analysis of R-PAT M MsiKD_MAF clone 21

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of MAF gene knockdown and TTR expression. Mean ±SD (n=1).

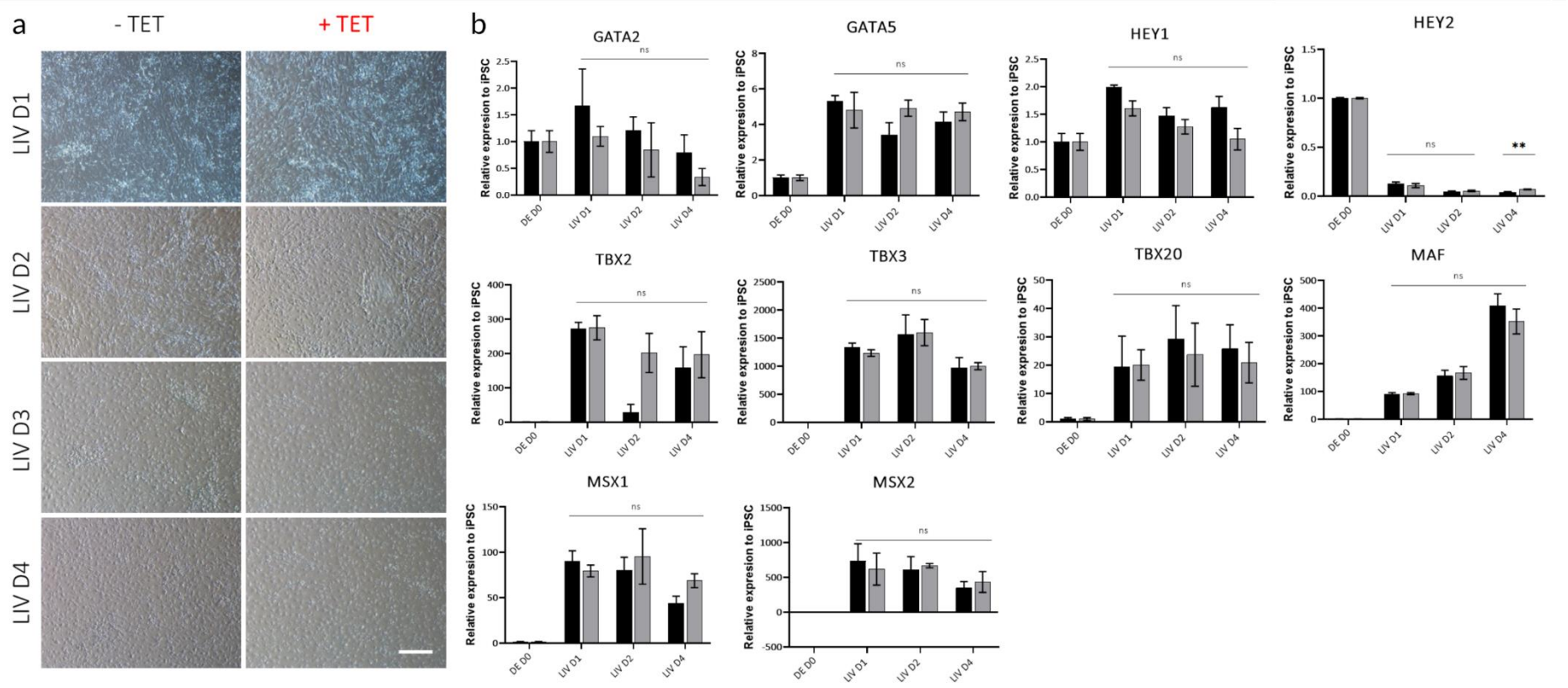

Figure 5-37 SiKD_SCR clone 8 differentiation to liver progenitor cells with TET induction.

a) Microscopy images of the last stage of differentiation to liver progenitor cells (LIV). Scale bar = 200μm; b) qPCR analysis of expression of all candidate TFs during the specification to liver progenitor cells. Mean ± SD (n=1).

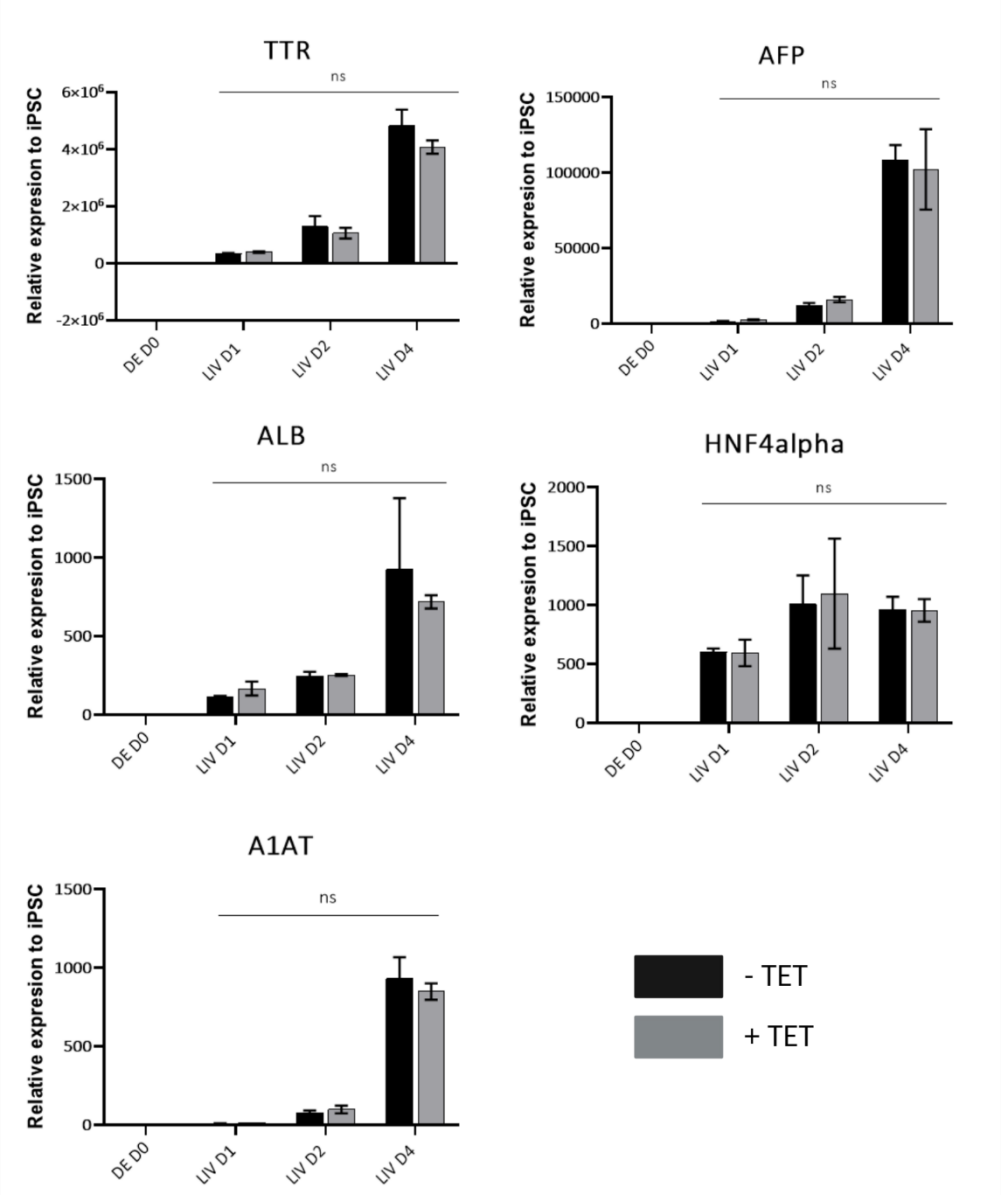

Figure 5-38 Expression of liver specific genes in SiKD_SCR clone during liver progenitor differentiation.

Gene expression measured by qPCR. Mean ±SD (n=1).

In summary, the initial screen of the correctly targeted MsiKD clones has identified the best available clones for further knockdown verification and experiments aimed at understanding the consequences of specific TF knockdown on the specification of hepatic progenitor cells. Before any further experiments are carried out the cell lines need to be fully characterized to ensure maintenance of pluripotency and normal karyotype.

5.3.7. TBX3 and TBX20 knockdown disrupts hepatic specification

Once a MsiKD cell line with efficient inducible knockdown have been identified for a candidate gene, further validation of the cell line was conducted and initial gene expression analysis was done by qPCR.

The MsiKD_TBX cell line has been designed to conditionally induce a production of shRNAs targeting three members of the TBX TF family: TBX2, TBX3 and TBX20. All three genes have been identified in our RNAseq. as upregulated in BMP4 induced liver progenitor cells. MsiKD_TBX cell line showed a significant decrease in mRNA levels of TBX3 and TBX20, but not TBX2. However, the cell line could still be used to investigate the consequences of the reduction in the two TBX genes, especially considering observable changes in hepatic progenitor cell morphology upon differentiation in the presence of TET. Gene expression analysis has not shown any changes in the expression of major hepatic TFs: HNF4alpha, C/EBP alpha or HNF1alpha during the four days of specification (Figure 5-39a). However, it showed a decrease in the mRNA levels of liver specific genes such as TTR, AFP, AAT and ALB (Figure 5-39b). This could indicate that the role of TBX3 and TBX20 was downstream of major hepatic TFs. Next, we wanted to check if TBX3/TBX20 knockdown directed the differentiation towards other endodermal cell lineages. Analysis of mRNA expression levels of major markers of foregut (HHEX), midgut (PDX1) and hindgut (CDX2) showed no statistically significant changes in the expression levels between -TET and +TET cells (Figure 5-40a). Although, statistical analysis showed no significant difference in the mRNA levels of PDX1 TF, there seems to be a trend of higher PDX1 expression in the cells with induced TBX3/TBX20 knockdown. There has been a recent report indicating that TBX3 knockdown during hepatic

specification enhances pancreatic gene expression (Mukherjee et al., 2021). We compared the mRNA levels of another two, pancreatic specific genes: HLXB9 and NKX6.1 between -TET and +TET cells. HLXB9 mRNA levels are increased in +TET cells at day 4 of liver specification, compared to -TET cells, but failed to reach significance. NKX6.1 mRNA levels show no significant difference (Figure 5-40b). More biological repeats are necessary to gain a better analysis of pancreatic gene expression in the TBX3/TBX20 knockdown. A cell line achieving a stronger knockdown activity, or a knockout cell line, could possibly show more clearly if TBX3/TBX20 influence pancreatic vs hepatic specification dynamics. Additionally, in mice it has been reported that TBX3 deficient hepatoblasts deviated towards cholangiocyte fate (Lüdtke et al., 2009, Suzuki et al., 2008). We have checked the expression of mRNA levels of cholangiocyte specific genes: CK19, OC1 and OC2 (Figure 5-41b). Only OC2 showed significantly different mRNA levels at day 2 of liver specification between -TET and +TET cells, however the trend is opposite to what would be expected if cholangiocyte fate was favoured. No other significant results were observed in cholangiocyte specific gene expression. This may be due to physiological differences between mice and humans and how the molecular changes are governed in the two species, however a more detailed study with more biological repeats is necessary to substantiate this claim.

Another mechanism by which TBX3 may control hepatic development is by the regulation of genes involved in hepatoblast delamination and epithelial to mesenchymal transition (EMT) (Lüdtke et al., 2009, Mukherjee et al., 2021). Our qPCR screen of mRNA expression levels of PROX1, TWIST and SLUG genes showed significant changes in the levels of SLUG mRNA between -TET and +TET cells, indicating that there may be some disruption to the EMT caused by TBX3 knockdown. However, this observation would also need verification by more biological repeats and a further detailed study.

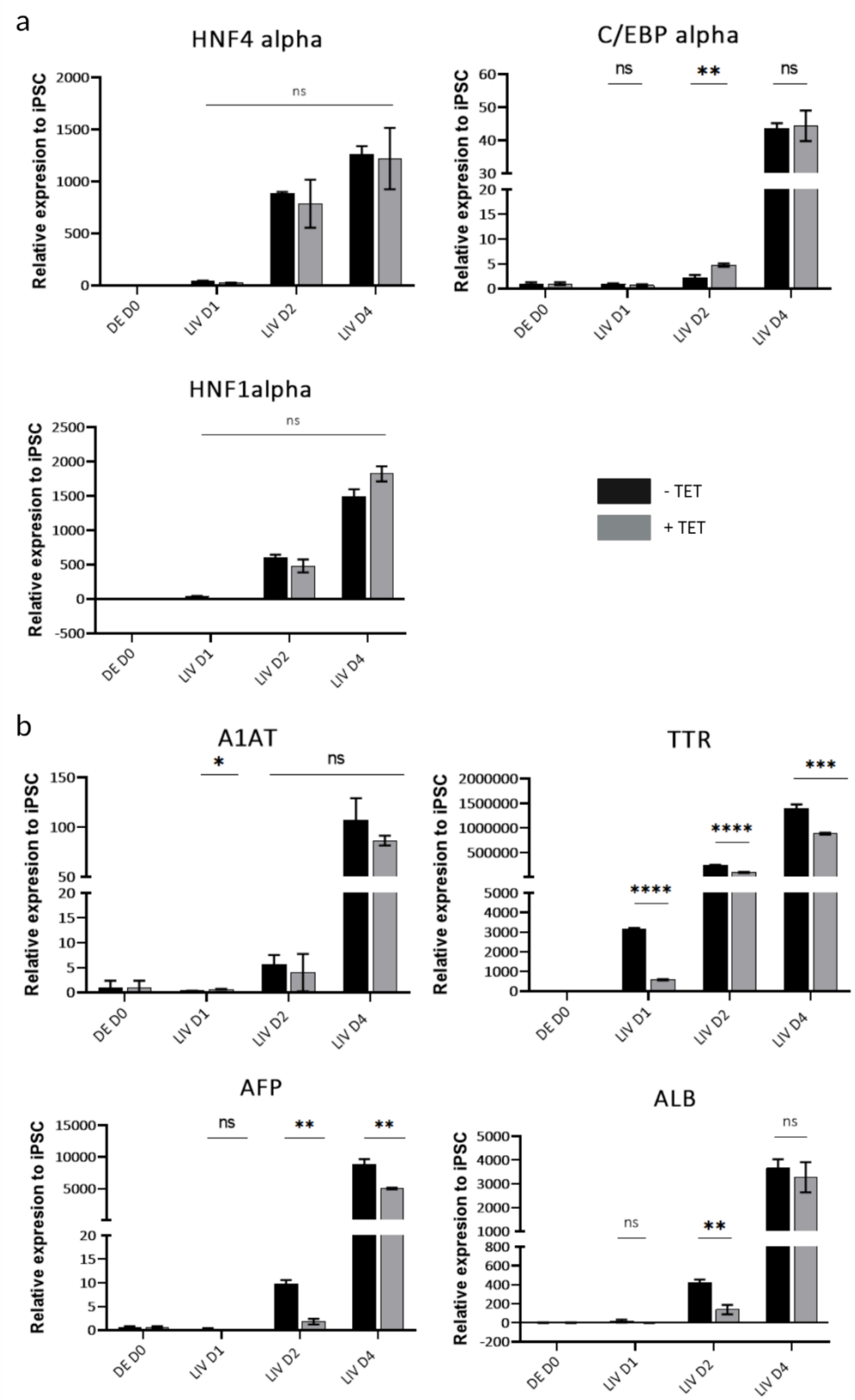

Figure 5-39 QPCR analysis of hepatic gene expression in the MsiKD_TBX cell line.
a) The mRNA levels of major TFs involved in hepatic specification were unaffected by TBX3/TBX20 knockdown; b)The mRNA levels of some hepatoblast specific genes were significantly decreased upon TBX3/20 knockdown. Mean ±SD (n=1).

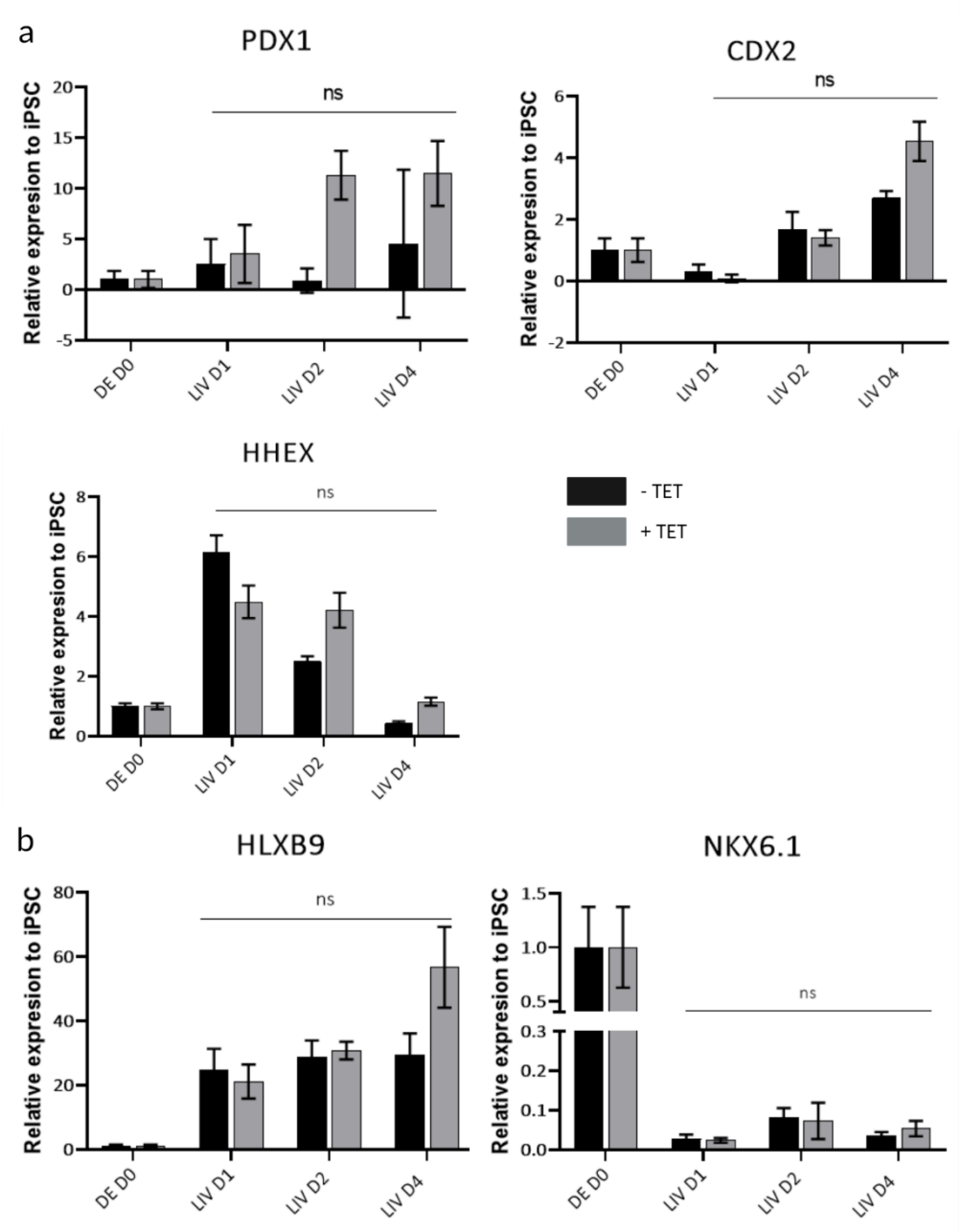

Figure 5-40 QPCR analysis of endodermal lineage gene expression in the MsiKD_TBX cell line.

a) The mRNA levels of major markers of foregut, midgut and hindgut are unaffected by the TBX3/20 knockdown; b) The mRNA levels of pancreatic genes are unaffected by the TBX3/20 knockdown. Mean +/- SD (n-=1).

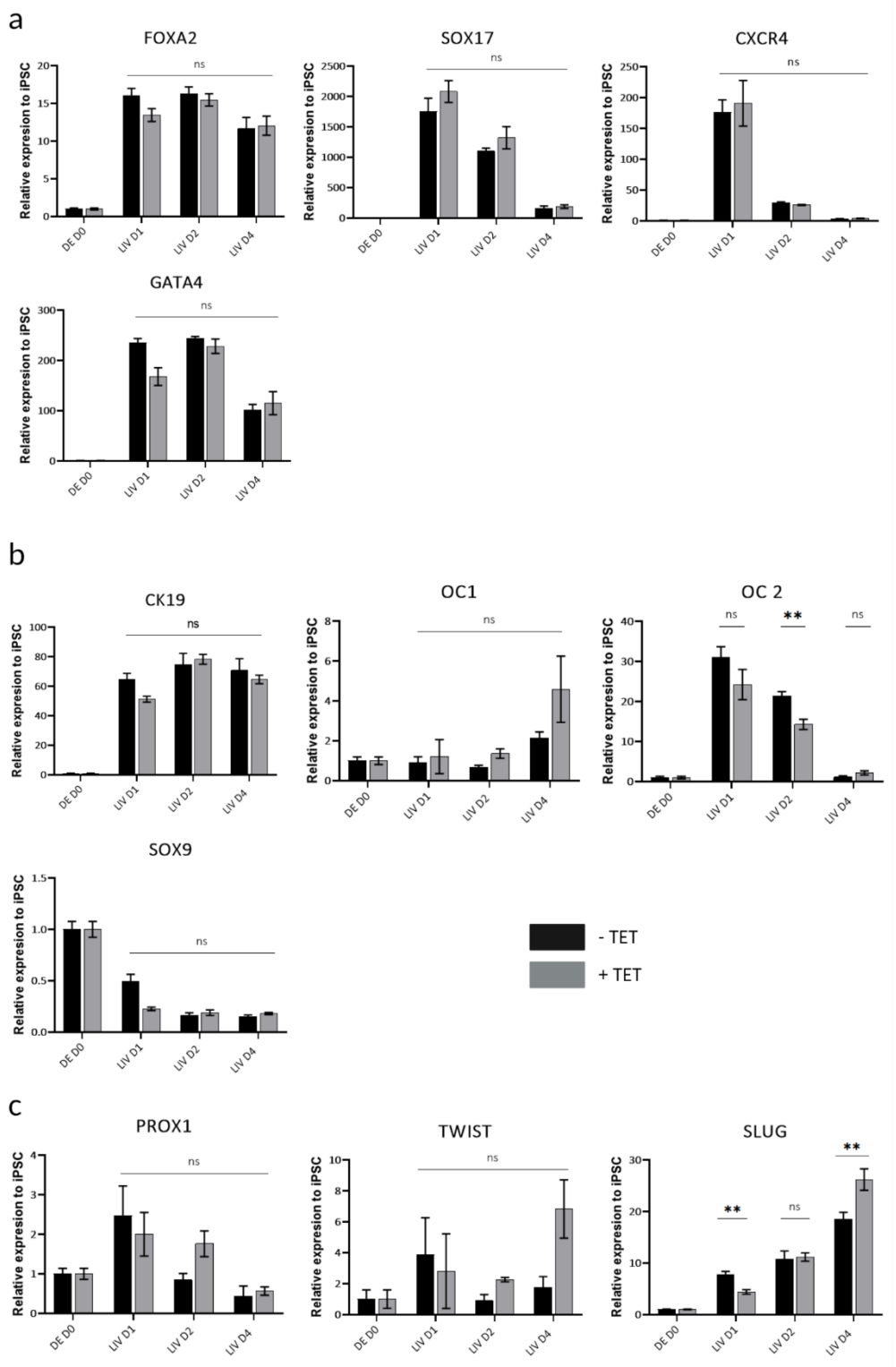

Figure 5-41 QPCR analysis of gene expression in MsiKD_TBX cell line.

a) The mRNA levels of foregut genes remains unaffected by TBX3/20 knockdown; b) The mRNA levels of cholangiocyte specific genes remain mostly unaffected by the TBX3/20 knockdown; c) qPCR analysis of changes in expression levels of TFs involved in hepatocyte migration and EMT. Mean +/- SD (n=1).

5.4. Discussion

In this chapter we described the creation of several inducible knockdown (OPTiKD) cell lines that would allow us the initial investigation into the role our selected TFs have in the hepatic specification.

5.4.1. Application of CRISPR/Cas 9 for AAVS1 locus targeting

CRISPR/Cas9 has rapidly overtaken other nuclease systems, such as zinc finger nucleases (ZNF) and transcription activator-like effectors (TALENS), as the genome engineering tool of choice due to its ease of design and use, higher specificity and efficiency, as well as low cost and possibility for multiplexing (Khan, 2019). However, it has been shown that Cas9 enzyme can tolerate up to five mismatches between the guide RNA sequence and the targeted sequence creating off target mutations around the genome (Fu et al., 2013). To improve the specificity, it was shown that optimization of the guide RNA sequence or the use of a modified version of Cas9 enzyme: Cas9 D10A with nickase activity can virtually eliminate the occurrence of off target mutations (Cho et al., 2014). As we had access to previously designed and validated CRISPR/Cas9 D10A system targeting the same area of the AAVS1 locus (Bhagwan et al., 2020) we decided to use this genome engineering tool instead of the ZFNs specified in the Bertero protocol.

The targeting efficiency was high but variable (58% +/- 13.6%). It was, however, lower than the quoted 95% achieved by Bertero and colleagues. This could have been caused by the use of different genetic engineering tool, user's experience or strictness of puromycin selection. In our experience, R-PAT M cells were more susceptible to puromycin treatment following nucleofection and, when the puromycin concentration determined in the resistance screen was used, all the targeted cells were killed within 48hrs of puromycin treatment despite high efficiency of plasmid delivery as determined by monitoring of GFP expression in CTRL cells. In initial experiments, to obtain any colonies for picking and screening the puromycin concentration had to be halved and maintained so for 48 hrs before gradual increase in the concentration for stricter selection. Therefore, some of the targeted cells were picked and screened following suboptimal

selection step and this could be the reason for lower efficiency and high variability in successfully targeted colonies. Perhaps puromycin toxicity testing would be more reflective if done following a mock nucleofection. On average, the derivation of homozygous clones showed similar efficiency but again showed high variability, with some attempts at targeting resulting in no homozygous clones. Although heterozygous clones can also be used provided a good level of knockdown is achieved, homozygous clones are more likely to ensure a strong and consistent level of knockdown (Bertero et al., 2018). The low numbers of HM clones are most likely caused by suboptimal clonal selection process. The cells post nucleofection were seeded into a petri dish and then treated with puromycin. The puromycin selection was stopped once single cells were remaining and, once colonies formed, they were picked for screening and expansion. Although the plates were observed to exclude colonies forming near each other, the process in not optimal and cannot exclude cell migration or colonies formed from more than one cell and missed during observation. For genuinely clonal cell lines, the cells following puromycin selection should be seeded at one per well dilution into a 96 well plate expanded and screened from a single cell stage. However, this process is extremely time consuming and considering homozygous clones can still be identified from a less strict clonal selection, screening of a higher number of colonies could also increase the chances of identifying a HM clone.

The PCR genotyping also showed that a high proportion of positively targeted clones had an extra plasmid copy/ies integrated in the genome. Although this could be advantageous in terms of induction of higher level of knockdown if shRNAs are secreted from more than one locus in the genome, it is generally undesired for several reasons. Firstly, the randomly integrated plasmid can disrupt unknown genes and confound the interpretation of experimental results. Secondly, if placed in a genomic region that undergoes silencing during differentiation, it could produce unreliable and inconsistent knockdown effects. In our experience, the number of clones positive for random integration of the plasmids was quite high, in case of MsiKD_HEY cell lines, all of the correctly

targeted clones were also shown to have randomly integrated plasmid backbone. One reason for this could be a presence of the plasmid as an episome in the cells during the screen. The protocol states that no residual plasmid should be left in the cells 2 weeks post transfection. To speed up the screening process we were collecting the gDNA at between 13-16 day post nucleofection and it is possible that episomal plasmid was still present in the cells. To ensure plasmid copies are completely gone from the screen clones, the gDNA for the random integration screen should be collected at a later stage, possibly once the clone is split into a T25, which usually was at least 21 days after nucleofection.

As the use of paired Cas9 nickases reduces the occurrence of off targets by up to 1500 fold (Ran et al., 2013) and no indels are created in predicted homology sites for each guide RNA (Cho et al., 2014) we have not performed off target screenings of any of the clones. Although this approach does not eliminate the possibility of off targets completely, the chances are reduced to the level acceptable for our study.

5.4.2. Knockdown activity

The main observation from the screen of the several MsiKD clones is that in some of the cell lines no knockdown activity was observed. For the MsiKD_GATA and MsiKD_MSX none of the tested clones showed a significant knockdown of any of the targeted genes, while in MsiKD_TBX and MsiKD_HEY the knockdown efficiency was variable between the tested clones and, additionally, in MsiKD_TBX only two out of three targeted genes were efficiently knocked down in the clone selected for further testing (clone 4). Although it is impossible without further experiments to say exactly why the knockdowns failed, there are a few possibilities that could explain it.

Firstly, the shRNA sequences selected for the knockdown could have been ineffective. Selection of the optimal shRNA for the experiment is the most important part of starting the design of an inducible knockdown cell line (Bertero et al., 2018). In our experiments, we have searched for previously published shRNA sequences shown to induce a high level of knockdown and yet

several of these failed to cause a desired effect. Another approach to selecting a good shRNA could be to identify three to four sequences targeting each gene and test them on an easily transfected cell line expressing the gene of interest. Although this approach requires some extra work, considering the substantial effort of creating an engineered hiPSC line it could be justified. However, we were limited by time to perform these experiments for a pre-screen study.

Secondly, it is possible that there was not enough of the shRNA secreted to induce significant decrease in the mRNA level of the targeted gene. The Bertero protocol recommends a tetracycline dose of 1µg/ml, which may need optimization for different cell lines. A dose response experiment for each clone could help identify TET concentration that allows a high level of knockdown without causing cell toxicity. Unfortunately, due to time restrictions we were unable to conduct such experiments.

Lastly, although a 'safe harbour' locus was selected for the transgene insertion there have been recent reports suggesting AAVS1 does not fulfil the criteria to be called one. Ordova et al., has first described silencing of the transgene targeted to the locus in undifferentiated hiPSC due to methylation. They also observed a varied expression of a transgene during differentiation of hiPSC to hepatocytes, but not other cell lineages (Ordovas et al., 2018). Klatt and colleagues observed transgene silencing during hiPSC differentiation towards myeloid lineages caused by methylation of the transgene. In this study, the silencing was dependent on the type of promoter used for the transgene (Klatt et al., 2020). Finally, Bhagwan et al., showed a varied expression of a transgene in undifferentiated hiPSC, from complete absence of transgene expression (0% of cells positive for the transgene) to a very high one (96.4%). This variation was independent of the zygosity status of the clone. They also demonstrated transgene silencing upon differentiation to mesodermal lineages, cardiomyocytes and haematopoietic cells (Bhagwan et al., 2020). However, despite the reported issues each group has managed to identify a clone in which silencing did not occur. Therefore, in case of targeting to the AAVS1 locus extensive screening of clones may be necessary to identify the best candidate

for experimental procedures. This necessity makes AAVS1 locus suboptimal for insertion of transgenes and a better candidate should be considered in the future.

In the light of the issues with AAVS1 locus possibly being silenced, the question also arises about the usefulness of the scramble control in our system. As the scramble shRNA does not target any genes there is no easy way for us to identify a clone in which the shRNA is efficiently secreted. If the scramble shRNA is not secreted, the RNAi pathways will not be activated in the cell and the main purpose of the scramble control is not fulfilled. There are a couple of methods that could be applied to check if the control shRNA is secreted, most approachable one would be using Taqman probes for the shRNA detection in a similar way to how expression of microRNAs is analysed using qPCR. This approach requires a purchase of several new reagents for the extraction of short RNA and Taqman qPCR master mix, as well as design of a custom probe for the detection of our scramble shRNA, all of which presents a substantial extra cost. Possibly, an shRNA targeting a ubiquitous protein, such as e.g.: GAPDH, would be a better control as the shRNA activity can be easily checked by qPCR. The MsiKD_SCR cell line can still be used to assess the effects of tetracycline on the cell's physiology, however such check could also be done on a non-targeted cells to gain the same information, therefore, for this purpose only, it would not be worth the considerable effort of creating the extra cell line.

In the case of MsiKD_GATA and MsiKD_MSX cell lines, none of the tested clones showed any degree of knockdown. This could be caused by either complete silencing of the transgene in all the selected clones or a selection of ineffective shRNAs. It is difficult to identify the exact cause without further experiments.

In the tested MsiKD_HEY cell lines the results were more varied. Two clones showed significant knockdown activity although the levels varied between them, the third clone showed no significant knockdown. It is therefore more likely that the reason for clone 9 failure to knockdown HEY genes upon TET addition was due to transgene silencing as the shRNAs are shown to be effective by good knockdown results in the other two clones. The variability between the

knockdown levels between the two functioning cell lines could also be caused by the presence of random integration of the transgene somewhere in the genome. We did not identify any MsiKD_HEY cell line with no extra plasmid copies and therefore it is possible that the varying levels of knockdown could be caused by different levels of shRNAs secretion between the cell lines. Another source of knockdown level variability could be the zygosity status of the cell line. In case of MsiKD_HEY clones we failed to conduct Locus screening and cannot tell which tested cell lines were HM or HT. We could expect that HM cell lines would induce a higher level of knockdown with twice as much of each shRNA being produced. Also, as our selection was not strictly clonal the cell lines identified by PCR as HT clones could actually be a mixture of HM, HT and WT cells. This would also lead to inconsistent knockdown results. To verify if the cell line is HT or HM a single cell seeding on 96 well plate followed by clone expansion and screening should be conducted.

Similar situation is observed in the MsiKD_TBX tested cell line. Clone 41 showed no knockdown activity and clones 4 and 8 showed variable levels of knockdown. Additionally, TBX2 shRNA turned out to be ineffective and failed to induce TBX2 mRNA levels decrease. We can assume this was not due to transgene silencing as the other two shRNAs, against TBX3 and TBX20, were secreted and achieved a good level of gene knockdown.

The qPCR screen was only an initial effort at validation of the inducible knockdown cell lines. Knockdown success should always be confirmed at the protein level by Western blot, which allows rapid, semi-quantitative assessment of protein knockdown. Techniques such as flow cytometry and immunocytochemistry can provide further insight into the homogeneity of the cell population. Additionally, pluripotency status should be checked for each cell line by testing the expression of pluripotency genes and differentiation to cell lineages of the three germ layers. Karyotyping should be performed to ensure that the process of genetic modification has not induced chromosomal changes. However, due to time limits caused by the COVID pandemic those experiments could not be performed in time.

5.4.3. TBX3 and TBX20 knockdown consequences for hepatic specification

Our RNA seq. screen identified three members of the T-box family of TFs that were upregulated upon BMP4 signalling during hepatoblast specification: TBX2, TBX3 and TBX20. All three have been reported to have a role in cardiac development (Sakabe et al., 2012, Singh et al., 2012). TBX2 and TBX3 have also been reported as important in limb (Lopatka and Moon, 2022), lung (Ludtke et al., 2016), inner ear (Kaiser et al., 2021), ureter (Aydoğdu et al., 2018) and mammary development (Jerome-Majewska et al., 2005). Studies on mice have also implicated Tbx3 in the development of the liver. Two reports show that Tbx3 knockdown mice have smaller livers compared to wild type and that hepatoblast fate decision is skewed toward the cholangiocyte linage. Therefore, Tbx3 is suggested to favour hepatic fate by the suppression of cholangiocyte specific genes in mice (Suzuki et al., 2008, Lüdtke et al., 2009). Mukherjee et al., (2021) reported that TBX3 knockout human iPSCs have impaired ability to differentiate to hepatocyte-like cells. The expression of early hepatic TFs in hepatoblasts is not affected but hepatocyte-like cells show decrease in the levels of liver specific genes, suggesting correct specification of hepatoblasts but impaired maturation (Mukherjee et al., 2021). In our experiments, no early hepatic transcription factors were affected by the TBX3/TBX20 knockdown but, already at the stage of hepatoblasts, there was a significant decrease in liver specific genes, suggesting that the process of hepatoblast specification is affected in TBX3/TBX20 knockdown cells. Contrary to the mice studies, we have not observed an increase in cholangiocyte specific genes upon TBX3 knockdown. This could be caused by an incomplete knockdown of TBX3 in our inducible system. Our knockdown efficiency varied between 59% and 87% depending on the day of specification. It is possible that there was still enough of the TBX3 protein produced to supress the expression of cholangiocyte lineage genes. Another explanation could be the differences between species. We notice that Suzuki et al., (2008) report the expression of Tbx2, Tbx3 and Tbx20 mRNA in the foetal liver tissue but only Tbx3 mRNA was detected in the population of cells identified as hepatoblasts. In our model, TBX2 mRNA is expressed during the specification of hepatoblasts. Human TBX2 and TBX3 share a 60% sequence

homology, and both were reported to be transcriptional repressors (Carreira et al., 1998, He et al., 1999). In other developmental context, they have been shown to function in a redundant manner. During mice lung development, Tbx2 deficient lung manifests defects in proliferation and morphogenesis only once Tbx3 expression is naturally downregulated (Lüdtke et al., 2013). Studies on murine lungs lacking both Tbx2 and Tbx3 genes showed that those genes are functionally redundant during branching morphogenesis of the lung (Ludtke et al., 2016). Their redundant function has also been shown in mice mammary (Jerome-Majewska et al., 2005), heart (Singh et al., 2012) inner ear (Kaiser et al., 2021) and limb development (Lopatka and Moon, 2022). Therefore, considering that we detected both TBX3 and TBX2 mRNA expression in our model of liver development, it is possible that the loss of TBX3 is compensated by the presence of TBX2. This could explain why there is no upregulation of cholangiocyte specific genes as reported in mice studies, and it could also be the reason why the TBX3 knockdown has a relatively little effect on the overall specification of the hepatoblasts in our study. Similarly, Mukherjee et al., (2021) observed a disruption of hepatic gene expression and increase in pancreatic markers expression in TBX3 knockout human iPSCs differentiated to hepatocyte-like cells. The authors suggest that TBX3 has a role as a suppressor of genetic drivers of other lineages. In our experiments, although there was a trend of increased PDX1 expression, it has not reached statistical significance, and other markers of pancreatic specification remained unaffected. In this case, the weaker effect of TBX3 knockdown as opposed to a complete elimination via a knockout could be the reason for the data discrepancy. Alternatively, extending the differentiation time of PPC could possibly result in similar effect as the one reported in the study.

The molecular mechanism by which TBX3 affects gene expression patterns in development is not very well understood. In liver development, control of cell cycle inhibitors or suppression of genes involved in hepatoblast delamination or EMT has been suggested (Suzuki et al., 2008, Lüdtke et al., 2009, Lüdtke et al., 2013). In other developmental areas, direct binding and suppression of cell cycle

inhibitors and Wnt signalling inhibitors has been shown (Lüdtke et al., 2013, Aydoğdu et al., 2018). In our study, we noted dysregulation in the expression of EMT regulator, SLUG, indicating that the disruption of this process may be a consequence of TBX3/TBX20 knockdown. We have not had a chance to verify changes in the expression levels of Wnt signalling pathway mediators.

In summary, this chapter described the creation of several inducible knockdown cell lines for the investigation of the role our candidate TFs have in the development of the liver. The adjusted Bertero protocol resulted in an efficient and rapid creation of the cell lines with the use of CRISPR/Cas9 system. The genotyping process allowed identification of correctly engineered clones and we were able to identify ones with suitable knockdown efficiency during LPCs specification. However, COVID pandemic had severely affected the time available for experiments and we were unable to use our model for more robust experiments. Our initial TBX3/TBX20 knockdown studies showed promising effects, although it also revealed issues with TBX2 knockdown efficiency, which we were unable to address due to time limitation. We were also unable to conduct any studies on the MsiKD_MAF and MsiKD_HEY cell lines.

Chapter 6. General discussion

The COVID-19 pandemic had significantly affected the planned work for this thesis. The substantial changes to working practices and restrictions on access to facilities has reduced the effective time in the laboratory to conduct several of the intended experiments. Additionally, one of the requirements for paid time extensions to the PhD time was to reduce the originally planned workload. Therefore, after obtaining our RNA sequencing results and constructing the necessary plasmids for candidate gene knockdowns, we were limited in the number of experiments we could undertake using our inducible knockdown cell lines. We were also unable to create knockoutand overexpression models for the TF selected for further characterization by our pre-screen with the use of knockdown cell lines.

6.1.1. Limitations of 2D hiPSCs culture for modelling of human development

The aim of this project was to apply hiPSCs as a model of early human development and explore the role of BMP4 signalling in the specification of liver progenitor cells (LPCs). The study of human development is limited due to the ethical issue surrounding work on human embryos. Currently, the 14 day limit for maintaining human embryos in culture excludes any study of development form gastrulation onwards (Carlson, 2019). In the past, modelling of human development or disease was mainly based on the use of animal studies. These models allow in vivo studies of development, however the inter-species differences limit their application to understanding human specific processes (Lal et al., 2016). The use of primary cells donated for research also presents many drawbacks due to the scarcity of material and difficulty of in vitro maintenance. Additionally, cells donated form fully developed organs cannot serve as models of development. Likewise, tissue from aborted foetuses is scarce and comes with ethical limitations. Therefore, hiPSCs are an attractive source of cells for the study of early human development. Protocols for differentiation of these cells into almost any cell type of the body have been intensively researched ever since their derivation around 20 years ago. Most of the knowledge of the signals and cues for cell differentiation was based on the

knowledge from animal studies and in *vitro* experimentation. The protocols can now render almost pure populations of specific cell types and the differentiation process allows systematic investigation of molecular changes happening during cell fate transitions (Baxter et al., 2015, Zhao et al., 2022). Especially when the differentiation is undertaken in 2D format, it allows for uniform and synchronized delivery of differentiation signals and a simplified model for the study of molecular changes in a particular cell type as attempted in this project. However, there are also many drawbacks to this approach to the study of the development. The 2D monolayer format of such experiment does not reflect the 3D environment of human embryo. In *vivo*, the cells of anterior foregut endoderm are surrounded by many other cell types with a myriad of secreted signalling molecules, cell-cell contact and ECM-cell interactions directing their behaviour. In our model, we use two signals to direct the specification of FE into LPCs and although to the best current knowledge this activates correct hepatic genes and allows further differentiation towards hepatocytes, it has been established by now that hiPSCs-derived hepatocytes are reflective more of the foetal stage rather than mature adult hepatocytes. The difference in gene expression profile between primary, foetal and hiPSC-derived hepatocytes types is quite pronounced (Godoy et al., 2018). Could this be caused by inadequate specification of LPCs? Considering the complexity of the in *vivo* environment, how likely is it that only two signals activate all the right genes necessary for the specification of the biopotential cells and further hepatocyte differentiation? Are we missing any other, more subtle cues that come from direct interaction with neighbouring cells or the ECM? Wnt signalling pathway has been shown to play a role in the LPC specification (McLin et al., 2007). Neither our protocol nor many other available ones modulate this pathway during the specification of LPCs and still achieve gene expression characteristic of the hepatic stage (Graffmann et al., 2022). However, the modulation of WNT signalling pathways during hepatic specification did result in improved maturation profile of differentiated hepatocytes (Touboul et al., 2016). Therefore, it is possible that inadequate LPC specification is one of the reasons for issues with hiPSCs-derived hepatocytes. Further investigation of cell fate decision might be more

appropriate in the context of a 3D culture with multiple cell types that are normally present in *vivo* development. Studies on creating *ex utero* mice embryo have progressed to the stage where aggregated mouse ESC can be cultured in vitro to an approximate equivalent of E8.5 stage of mice development. By this stage, the synthetic embryos accomplish gastrulation and proceed to forming a beating heart, gut tube, develop a brain and many other organs progenitors (Tarazi et al., 2022). E8.5 is the time where the specification of LPC begins, therefore it is feasible that in near future the research will progress further and allow the study of liver formation within the complex, multicellular environment. Although species specificity will continue to be an issue with this model, if ethical boundaries on the use of human embryo ever change, the knowledge, methods and devices gained on the mouse studies will lay the groundwork for translating such methods to human embryo research. 3D culture does have some drawbacks. They are low throughput, more labour intensive and microscopic analysis can be difficult. Therefore, the use of 2D cell culture can still be appropriate for simpler investigations of molecular functions of single genes.

6.1.2. Limitations of Matrigel use for hiPSCs culture and differentiation

Although Matrigel has been used for cell culture applications for almost 40 years, issues surrounding its use are known since at least 1992 when Vukicevic et al., (1992) identified multiple active growth factors within the matrix (Vukicevic et al., 1992). Matrigel is a highly complex structure derived from a Englebreth-Holm Swarm mouse tumour. It is rich in ECM proteins such as laminins, collagens, enactin and heparan sulphate proteoglycans and growth factors such as FGFs and TGFs. 14 000 different peptides and 2000 proteins have been identified in this substrate. In hPSCs culture, Matrigel replaced the use of feeder layer made of mouse embryonic fibroblasts, which significantly simplified the culture and maintenance of this cell type. However, the complexity of the matrix and the batch to batch variability, not only in protein composition but also in matrix stiffness, calls into question the suitability of this matrix for use in hPSCs culture (Aisenbrey and Murphy, 2020). In the light of the issues with the

substrate used for this project, the results from our RNA sequencing relating to changes in cell adhesion and ECM-receptor interacting genes have to be interpreted with caution. Although, considering the migration of LPC cells from the liver bud it is reasonable to expect changes in the cell adhesion and locomotion genes as detected in our results, would the same genes be activated if the cells were on a matrix more reflective of the natural ECM in the developing embryo? Currently, many alternatives to Matrigel are available. Natural proteins and peptides such as vitronectin or laminin E8 fragments have been shown to support hPSCs maintenance and expansion (Braam et al., 2008, Miyazaki et al., 2012). Additionally, intense research on synthetic substrates has also identified many good alternatives to Matrigel. Hydrophilic, bioinert polymers such as PEG (poly(ethylene glycol)) can be modified with various functional groups to adjust the properties of resulting substrate. Stiffness, biochemical properties and ECM characteristics can be adjusted with synthetic scaffolds (Aisenbrey and Murphy, 2020). Synthetic substrates modified with natural peptides have been successfully applied to hPSC culture and expansion, as well as for differentiation toward hepatocytes where they showed to improve hepatocyte maturity (Yamazoe et al., 2013, Lambshead et al., 2018).

6.1.3. Feasibility of an inducible system in hiPSCs differentiation for functional screen of candidate genes

Our problems with transfecting foregut monolayer prompted the decision to genetically modify R-PAT cells at the pluripotent stage. The tetracycline-controlled inducible knock down system allows temporal control of shRNA expression therefore avoiding gene levels reduction at an inappropriate stage in cases where a gene has a dual role in the development, like, for example, SOX2 (Zhang, 2014). Knockdown is easily activated by the addition of tetracycline at any point of differentiation without the need for transfection (Bertero et al., 2018). This method also optimized creation of knockdown plasmid against multiple genes, useful for the study of redundant gene function. Although the protocol we used was well optimized and allowed for rapid and highly efficient plasmid construction, the whole process of creating MsiKD cell line was quite

labour intensive, especially when handling more than two modified cell lines at a time. Issues with high rate of plasmid backbone insertion and silencing of the transgene resulted in very few clones that were appropriate for further study. Therefore, this approach may not be the most suitable for screening of a high number of candidate genes.

In hindsight, the optimization of FG layer dissociation and passaging could have been given more effort. Derivation of human foregut stem cells (FSC) that have the ability to self-renew and differentiate to hepatic and pancreatic lineages has been described (Hannan et al., 2013a). FSC can be dissociated and seeded at 30-40 x 10^3 per cm^2 creating conditions more conductive to efficient transfection: actively dividing cells and confluency at between 70-80%. Optimized transfection of the FSC would allow for knockdown studies using siRNAs. This approach would come with some advantages. It is easier to achieve high knockdown efficiency with pools of 3-4 siRNAs directed against the gene of interest and exchange of ineffective shRNAs is much simpler. Without the need of creating a cell line for each candidate gene or gene families, there would be more time to perform knock down studies on a higher number of genes of interest and better analysis of knockdown effects. Furthermore, knockout and overexpression studies on the most interesting candidate gene selected after the RNAi screen, would also be possible with a ready FSC transfection method.

6.1.4. Summary of the main findings

- Transfecting foregut endoderm cells with lipid reagents is inefficient and can be replaced with electroporation provided robust protocol for FG dissociation and re-seeding is established.
- Puromycin toxicity testing on cells that have not been electroporated can result in inaccurate assessment of the puromycin concentration needed to eliminate cells without puromycin resistance gene.
- During differentiation of human iPSCs towards LPCs blocking of the BMP4 pathway at the transition between FE and APC completely prevents hepatic gene expression, in line with effects observed in mice studies.

- BMP4 induces changes in the gene expression pattern during LPC specification from foregut endoderm.
- Gene associated with cell adhesion become downregulated upon BMP4 signalling, and genes associated with hepatocyte function become upregulated early in the LPC specification process.
- Several TFs previously not reported to have a role in hepatic development become upregulated by BMP4 signalling during LPC specification. Additionally, some of those TFs have associated lncRNAs also upregulated during the process.
- Knockdown of TBX3 and TBX20 disrupts LPC specification as shown by a decrease in hepatocyte specific genes and changes in the expression of EMT regulators.

6.1.5. Future work

The COVID-19 related adjustment to project plan allowed for only preliminary studies of molecular functions of our candidate transcription factors.

In first instance, I would like to address any deficiencies in the knockdown studies of the candidate TFs:

- Karyotyping, pluripotency check and sequencing confirmation of the MsiKD clones that were used in preliminary screens or would be used in further experiments;
- Increase the number of biological repeats and assess a wider selection of hepatic and biliary markers by qPCR
- Validation of the knockdowns at the protein level
- Thorough examination of the expression levels and pattern of all candidate TFs and other members of the gene families in our model at transcript level and, where possible, at protein level

Next, the repat of ChIP sequencing experiment, following a check and optimization of the antibody would complement and strengthen the RNA sequencing results and any conclusions we can draw on the regulation of LPC specification by BMP4 signalling.

Provided the preliminary results of TBX3 and TBX20 knockdown are confirmed, identification of an efficient shRNA against TBX2 and creation of triple knockdown cell lines would also be undertaken. Additionally, differentiation of the MsiKD_TBX cell line to hepatocyte stage would allow the assessment of the consequences of TBX genes knockdown on hepatocyte maturation and functions (albumin secretion, urea production, cytochrome P450 activity or lipid handling tests). Following that, single and double knockout cell lines would be created to assess the redundancy of the two factors. Cell lines with inducible transgenes containing TBX2 or TBX3 mRNA would also be created to examine phenotype changes resulting from overexpression of these proteins. Finally, double knockout mouse studies could be done to assess TBX family genes in *vivo* functio

Chapter 7. Bibliography

ABRAHAMS, A., PARKER, M. I. & PRINCE, S. 2009. The T-box transcription factor Tbx2: Its role in development and possible implication in cancer. *IUBMB Life*, NA-NA.

ADLI, M. 2018. The CRISPR tool kit for genome editing and beyond. *Nature Communications,* 9.

AHLER, E., SULLIVAN, W. J., CASS, A., BRAAS, D., YORK, A. G., BENSINGER, S. J., GRAEBER, T. G. & CHRISTOFK, H. R. 2013. Doxycycline Alters Metabolism and Proliferation of Human Cell Lines. *PLoS ONE,* 8, e64561.

AISENBREY, E. A. & MURPHY, W. L. 2020. Synthetic alternatives to Matrigel. *Nat Rev Mater,* 5, 539-551.

AKAI, Y., OITATE, T., KOIKE, T. & SHIOJIRI, N. 2014. Impaired hepatocyte maturation, abnormal expression of biliary transcription factors and liver fibrosis in C/EBPalpha(Cebpa)-knockout mice. *Histol Histopathol,* 29, 107-25.

ALVES-BEZERRA, M. & COHEN, D. E. 2017. Triglyceride Metabolism in the Liver. *Comprehensive Physiology*, 1-22.

AMPUJA, M. & KALLIONIEMI, A. 2018. Transcription factors-Intricate players of the bone morphogenetic protein signaling pathway. *Genes Chromosomes Cancer,* 57, 3-11.

ARDISASMITA, A. I., SCHENE, I. F., JOORE, I. P., KOK, G., HENDRIKS, D., ARTEGIANI, B., MOKRY, M., NIEUWENHUIS, E. E. S. & FUCHS, S. A. 2022. A comprehensive transcriptomic comparison of hepatocyte model systems improves selection of models for experimental use. *Communications Biology,* 5.

ASRANI, S. K., DEVARBHAVI, H., EATON, J. & KAMATH, P. S. 2019. Burden of liver diseases in the world. *J Hepatol,* 70, 151-171.

AYDOĞDU, N., RUDAT, C., TROWE, M.-O., KAISER, M., LÜDTKE, T. H., TAKETO, M. M., CHRISTOFFELS, V. M., MOON, A. & KISPERT, A. 2018. TBX2 and TBX3 act downstream of canonical WNT signaling in patterning and differentiation of the mouse ureteric mesenchyme. *Development,* 145, dev171827.

BAAS, R., SIJM, A., VAN TEEFFELEN, H. A. A. M., VAN ES, R., VOS, H. R. & MARC TIMMERS, H. T. 2016. Quantitative Proteomics of the SMAD (Suppressor of Mothers against Decapentaplegic) Transcription Factor Family Identifies Importin 5 as a Bone Morphogenic Protein Receptor SMAD-specific Importin. *Journal of Biological Chemistry,* 291, 24121-24132.

BAKER, B. M. & CHEN, C. S. 2012. Deconstructing the third dimension – how 3D culture microenvironments alter cellular cues. *Journal of Cell Science,* 125, 3015-3024.

BANERJEE, P., SURENDRAN, H., BHARTI, K., MORISHITA, K., VARSHNEY, A. & PAL, R. 2018. Long Noncoding RNA RP11-380D23.2 Drives Distal-Proximal Patterning of the Lung by Regulating PITX2 Expression. *Stem Cells,* 36, 218-229.

BAUMGART, A. K. & BEYER, M. 2017. Genetic engineering as a tool for the generation of mouse models to understand disease phenotypes and gene function. *Curr Opin Biotechnol,* 48**,** 228-233.

BAXTER, M., WITHEY, S., HARRISON, S., SEGERITZ, C.-P., ZHANG, F., ATKINSON-DELL, R., ROWE, C., GERRARD, D. T., SISON-YOUNG, R., JENKINS, R., HENRY, J., BERRY, A. A., MOHAMET, L., BEST, M., FENWICK, S. W., MALIK, H., KITTERINGHAM, N. R., GOLDRING, C. E., PIPER HANLEY, K., VALLIER, L. & HANLEY, N. A. 2015. Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. *Journal of Hepatology,* 62**,** 581-589.

BECIC, T., KERO, D., VUKOJEVIC, K., MARDESIC, S. & SARAGA-BABIC, M. 2018. Growth factors FGF8 and FGF2 and their receptor FGFR1, transcriptional factors Msx-1 and MSX-2, and apoptotic factors p19 and RIP5 participate in the early human limb development. *Acta Histochem,* 120**,** 205-214.

BENEDICTO, A., ROMAYOR, I. & ARTETA, B. 2017. Role of liver ICAM-1 in metastasis. *Oncology Letters,* 14**,** 3883-3892.

BERTERO, A., PAWLOWSKI, M., ORTMANN, D., SNIJDERS, K., YIANGOU, L., CARDOSO DE BRITO, M., BROWN, S., BERNARD, W. G., COOPER, J. D., GIACOMELLI, E., GAMBARDELLA, L., HANNAN, N. R. F., IYER, D., SAMPAZIOTIS, F., SERRANO, F., ZONNEVELD, M. C. F., SINHA, S., KOTTER, M. & VALLIER, L. 2016. Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro studies of human development using hPSCs. *Development,* 143**,** 4405-4418.

BERTERO, A., YIANGOU, L., BROWN, S., ORTMANN, D., PAWLOWSKI, M. & VALLIER, L. 2018. Conditional Manipulation of Gene Function in Human Cells with Optimized Inducible shRNA. *Current Protocols in Stem Cell Biology,* 44.

BHAGWAN, J. R., COLLINS, E., MOSQUEIRA, D., BAKAR, M., JOHNSON, B. B., THOMPSON, A., SMITH, J. G. W. & DENNING, C. 2020. Variable expression and silencing of CRISPR-Cas9 targeted transgenes identifies the AAVS1 locus as not an entirely safe harbour. *F1000Research,* 8**,** 1911.

BIER, E. & DE ROBERTIS, E. M. 2015. EMBRYO DEVELOPMENT. BMP gradients: A paradigm for morphogen-mediated developmental patterning. *Science,* 348**,** aaa5838.

BITINAITE, J., WAH, D. A., AGGARWAL, A. K. & SCHILDKRAUT, I. 1998. FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences,* 95**,** 10570-10575.

BLAKELEY, P., FOGARTY, N. M. E., DEL VALLE, I., WAMAITHA, S. E., HU, T. X., ELDER, K., SNELL, P., CHRISTIE, L., ROBSON, P. & NIAKAN, K. K. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development,* 142**,** 3151-3165.

BORT, R., SIGNORE, M., TREMBLAY, K., MARTINEZ BARBERA, J. P. & ZARET, K. S. 2006. Hex homeobox gene controls the transition of the endoderm to a pseudostratified, cell emergent epithelium for liver bud development. *Dev Biol,* 290**,** 44-56.

BOSWELL, B. A. & MUSIL, L. S. 2015. Synergistic interaction between the fibroblast growth factor and bone morphogenetic protein signaling pathways in lens cells. *Molecular Biology of the Cell,* 26**,** 2561-2572.

BRAAM, S. R., ZEINSTRA, L., LITJENS, S., WARD-VAN OOSTWAARD, D., VAN DEN BRINK, S., VAN LAAKE, L., LEBRIN, F., KATS, P., HOCHSTENBACH, R., PASSIER, R., SONNENBERG, A. & MUMMERY, C. L. 2008. Recombinant vitronectin is a functionally defined substrate that supports human embryonic stem cell self-renewal via alphavbeta5 integrin. *Stem Cells,* 26**,** 2257-65.

BRAGDON, B., MOSEYCHUK, O., SALDANHA, S., KING, D., JULIAN, J. & NOHE, A. 2011. Bone morphogenetic proteins: a critical review. *Cell Signal,* 23**,** 609-20.

BRAZIL, D. P., CHURCH, R. H., SURAE, S., GODSON, C. & MARTIN, F. 2015. BMP signalling: agony and antagony in the family. *Trends Cell Biol,* 25**,** 249-64.

CAI, J., ZHAO, Y., LIU, Y., YE, F., SONG, Z., QIN, H., MENG, S., CHEN, Y., ZHOU, R., SONG, X., GUO, Y., DING, M. & DENG, H. 2007. Directed differentiation of human embryonic stem cells into functional hepatic cells. *Hepatology,* 45**,** 1229-1239.

CAREY, M. F., PETERSON, C. L. & SMALE, S. T. 2009. Chromatin immunoprecipitation (ChIP). *Cold Spring Harb Protoc,* 2009**,** pdb prot5279.

CARLSON, B. M. K. P. N. 2019. *Human embryology and developmental biology,* Mo, Elsevier.

CARREIRA, S., DEXTER, T. J., YAVUZER, U., EASTY, D. J. & GODING, C. R. 1998. Brachyury-Related Transcription Factor Tbx2 and Repression of the Melanocyte-Specific TRP-1 Promoter. *Molecular and Cellular Biology,* 18**,** 5099-5108.

CEREGHINI, S. 1996. Liver-enriched transcription factors and hepatocyte differentiation. *FASEB J,* 10**,** 267-82.

CHATTERJEE, N. & WALKER, G. C. 2017. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis,* 58**,** 235-263.

CHEN, Y.-H., ISHII, M., SUCOV, H. M. & MAXSON, R. E. 2008. Msx1 and Msx2are required for endothelial-mesenchymal transformation of the atrioventricular cushions and patterning of the atrioventricular myocardium. *BMC Developmental Biology,* 8**,** 75.

CHEN, Y., JURGENS, K., HOLLEMANN, T., CLAUSSEN, M., RAMADORI, G. & PIELER, T. 2003. Cell-autonomous and signal-dependent expression of liver and intestine marker genes in pluripotent precursor cells from Xenopus embryos. *Mech Dev,* 120**,** 277-88.

CHO, C. H. H., HANNAN, N. R. F., DOCHERTY, F. M., DOCHERTY, H. M., JOÅO LIMA, M., TROTTER, M. W. B., DOCHERTY, K. & VALLIER, L. 2012. Inhibition of activin/nodal signalling is necessary for pancreatic differentiation of human pluripotent stem cells. *Diabetologia,* 55**,** 3284-3295.

CHO, S. W., KIM, S., KIM, Y., KWEON, J., KIM, H. S., BAE, S. & KIM, J.-S. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research,* 24**,** 132-141.

CHOI, J., LEE, S., MALLARD, W., CLEMENT, K., TAGLIAZUCCHI, G. M., LIM, H., CHOI, I. Y., FERRARI, F., TSANKOV, A. M., POP, R., LEE, G., RINN, J. L., MEISSNER, A., PARK, P. J. & HOCHEDLINGER, K. 2015. A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nature Biotechnology,* 33**,** 1173-1181.

CHONG, J. J. H., YANG, X., DON, C. W., MINAMI, E., LIU, Y.-W., WEYERS, J. J., MAHONEY, W. M., VAN BIBER, B., COOK, S. M., PALPANT, N. J., GANTZ, J. A., FUGATE, J. A., MUSKHELI, V., GOUGH, G. M., VOGEL, K. W., ASTLEY, C. A., HOTCHKISS, C. E., BALDESSARI, A., PABON, L., REINECKE, H., GILL, E. A., NELSON, V., KIEM, H.-P., LAFLAMME, M. A. & MURRY, C. E. 2014. Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature,* 510**,** 273-277.

CHONG, Z. X., YEAP, S. K. & HO, W. Y. 2021. Transfection types, methods and strategies: a technical review. *PeerJ,* 9**,** e11165.

COFFINIER, C., GRESH, L., FIETTE, L., TRONCHE, F. O., SCHÜTZ, G. N., BABINET, C., PONTOGLIO, M., YANIV, M. & BARRA, J. 2002. Bile system morphogenesis defects and liver dysfunction upon targeted deletion of HNF1β. *Development,* 129**,** 1829-1838.

CUI, K. W., ENGEL, L., DUNDES, C. E., NGUYEN, T. C., LOH, K. M. & DUNN, A. R. 2020. Spatially controlled stem cell differentiation via morphogen gradients: A comparison of static and dynamic microfluidic platforms. *J Vac Sci Technol A,* 38**,** 033205.

D'AIUTO, L., PRASAD, K. M., UPTON, C. H., VIGGIANO, L., MILOSEVIC, J., RAIMONDI, G., MCCLAIN, L., CHOWDARI, K., TISCHFIELD, J., SHELDON, M., MOORE, J. C., YOLKEN, R. H., KINCHINGTON, P. R. & NIMGAONKAR, V. L. 2015. Persistent infection by HSV-1 is associated with changes in functional architecture of iPSC-derived neurons and brain activation patterns underlying working memory performance. *Schizophr Bull,* 41**,** 123-32.

DALBY, B., CATES, S., HARRIS, A., OHKI, E. C., TILKINS, M. L., PRICE, P. J. & CICCARONE, V. C. 2004. Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. *Methods,* 33**,** 95-103.

DANESHVAR, K., JOSHUA, KIM, B.-M., ZHOU, C., SAMUEL, JILLIAN, ABUALTEEN, A., TAN, B., ALLA, MARCHO, C., KIMBERLY, MAGER, J., MICHAEL & ALAN 2016. DIGIT Is a Conserved Long Noncoding RNA that Regulates GSC Expression to Control Definitive Endoderm Differentiation of Embryonic Stem Cells. *Cell Reports,* 17**,** 353-365.

DASTON, G. P., MAHONY, C., THOMAS, R. S. & VINKEN, M. 2022. Assessing Safety Without Animal Testing: The Road Ahead. *Toxicol Sci,* 187**,** 214-218.

DE LOS ANGELES, A., FERRARI, F., XI, R., FUJIWARA, Y., BENVENISTY, N., DENG, H., HOCHEDLINGER, K., JAENISCH, R., LEE, S., LEITCH, H. G., LENSCH, M. W., LUJAN, E., PEI, D., ROSSANT, J., WERNIG, M., PARK, P. J. & DALEY, G. Q. 2015. Hallmarks of pluripotency. *Nature,* 525**,** 469-478.

DELAFOREST, A., DI FURIO, F., JING, R., LUDWIG-KUBINSKI, A., TWAROSKI, K., URICK, A., PULAKANTI, K., RAO, S. & DUNCAN, S. 2018. HNF4A Regulates the Formation of Hepatic Progenitor Cells from Human iPSC-Derived

Endoderm by Facilitating Efficient Recruitment of RNA Pol II. *Genes,* 10**,** 21.

DELAFOREST, A., NAGAOKA, M., SI-TAYEB, K., NOTO, F. K., KONOPKA, G., BATTLE, M. A. & DUNCAN, S. A. 2011. HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development,* 138**,** 4143-4153.

DHAWAN, A., CHAIJITRARUCH, N., FITZPATRICK, E., BANSAL, S., FILIPPI, C., LEHEC, S. C., HEATON, N. D., KANE, P., VERMA, A., HUGHES, R. D. & MITRY, R. R. 2020. Alginate microencapsulated human hepatocytes for the treatment of acute liver failure in children. *J Hepatol,* 72**,** 877-884.

DIJKSTERHUIS, J. P., BALJINNYAM, B., STANGER, K., SERCAN, H. O., JI, Y., ANDRES, O., RUBIN, J. S., HANNOUSH, R. N. & SCHULTE, G. 2015. Systematic Mapping of WNT-FZD Protein Interactions Reveals Functional Selectivity by Distinct WNT-FZD Pairs. *Journal of Biological Chemistry,* 290**,** 6789-6798.

DJEBALI, S., DAVIS, C. A., MERKEL, A., DOBIN, A., LASSMANN, T., MORTAZAVI, A., TANZER, A., LAGARDE, J., LIN, W., SCHLESINGER, F., XUE, C., MARINOV, G. K., KHATUN, J., WILLIAMS, B. A., ZALESKI, C., ROZOWSKY, J., RÖDER, M., KOKOCINSKI, F., ABDELHAMID, R. F., ALIOTO, T., ANTOSHECHKIN, I., BAER, M. T., BAR, N. S., BATUT, P., BELL, K., BELL, I., CHAKRABORTTY, S., CHEN, X., CHRAST, J., CURADO, J., DERRIEN, T., DRENKOW, J., DUMAIS, E., DUMAIS, J., DUTTAGUPTA, R., FALCONNET, E., FASTUCA, M., FEJES-TOTH, K., FERREIRA, P., FOISSAC, S., FULLWOOD, M. J., GAO, H., GONZALEZ, D., GORDON, A., GUNAWARDENA, H., HOWALD, C., JHA, S., JOHNSON, R., KAPRANOV, P., KING, B., KINGSWOOD, C., LUO, O. J., PARK, E., PERSAUD, K., PREALL, J. B., RIBECA, P., RISK, B., ROBYR, D., SAMMETH, M., SCHAFFER, L., SEE, L.-H., SHAHAB, A., SKANCKE, J., SUZUKI, A. M., TAKAHASHI, H., TILGNER, H., TROUT, D., WALTERS, N., WANG, H., WROBEL, J., YU, Y., RUAN, X., HAYASHIZAKI, Y., HARROW, J., GERSTEIN, M., HUBBARD, T., REYMOND, A., ANTONARAKIS, S. E., HANNON, G., GIDDINGS, M. C., RUAN, Y., WOLD, B., CARNINCI, P., GUIGÓ, R. & GINGERAS, T. R. 2012. Landscape of transcription in human cells. *Nature,* 489**,** 101-108.

DONG, L., DONG, Q., CHEN, Y., LI, Y., ZHANG, B., ZHOU, F., LYU, X., CHEN, G. G., LAI, P., KUNG, H.-F. & HE, M.-L. 2018. Novel HDAC5-interacting motifs of Tbx3 are essential for the suppression of E-cadherin expression and for the promotion of metastasis in hepatocellular carcinoma. *Signal Transduction and Targeted Therapy,* 3.

DRAKE, R. L. V., WAYNE; MITCHELL, ADAM W. M. 2010. *Gray's anatomy for students [electronic resource],* Canada, Churchhill Livingstone Elsevier.

DUVAL, N., DAUBAS, P., BOURCIER DE CARBON, C., ST CLOMENT, C., TINEVEZ, J.-Y., LOPES, M., RIBES, V. & ROBERT, B. 2014. Msx1 and Msx2 act as essential activators of Atoh1 expression in the murine spinal cord. *Development,* 141**,** 1726-1736.

EGELHOFER, T. A., MINODA, A., KLUGMAN, S., LEE, K., KOLASINSKA-ZWIERZ, P., ALEKSEYENKO, A. A., CHEUNG, M.-S., DAY, D. S., GADEL, S., GORCHAKOV, A. A., GU, T., KHARCHENKO, P. V., KUAN, S., LATORRE, I., LINDER-BASSO,

D., LUU, Y., NGO, Q., PERRY, M., RECHTSTEINER, A., RIDDLE, N. C., SCHWARTZ, Y. B., SHANOWER, G. A., VIELLE, A., AHRINGER, J., ELGIN, S. C. R., KURODA, M. I., PIRROTTA, V., REN, B., STROME, S., PARK, P. J., KARPEN, G. H., HAWKINS, R. D. & LIEB, J. D. 2011. An assessment of histone-modification antibody quality. *Nature Structural & Molecular Biology,* 18**,** 91-93.

FISHER, J. B., PULAKANTI, K., RAO, S. & DUNCAN, S. A. 2017. GATA6 is essential for endoderm formation from human pluripotent stem cells. *Biology Open,* 6**,** 1084-1095.

FU, Y., FODEN, J. A., KHAYTER, C., MAEDER, M. L., REYON, D., JOUNG, J. K. & SANDER, J. D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology,* 31**,** 822-826.

FUREY, T. S. 2012. ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics,* 13**,** 840-852.

FUS-KUJAWA, A., PRUS, P., BAJDAK-RUSINEK, K., TEPER, P., GAWRON, K., KOWALCZUK, A. & SIERON, A. L. 2021. An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front Bioeng Biotechnol,* 9**,** 701031.

GAJ, T., GERSBACH, C. A. & BARBAS, C. F. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology,* 31**,** 397-405.

GAO, N., WHITE, P. & KAESTNER, K. H. 2009. Establishment of Intestinal Identity and Epithelial-Mesenchymal Signaling by Cdx2. *Developmental Cell,* 16**,** 588-599.

GERRARD, L., RODGERS, L. & CUI, W. 2005. Differentiation of Human Embryonic Stem Cells to Neural Lineages in Adherent Culture by Blocking Bone Morphogenetic Protein Signaling. *STEM CELLS,* 23**,** 1234-1241.

GODOY, P., SCHMIDT-HECK, W., HELLWIG, B., NELL, P., FEUERBORN, D., RAHNENFÜHRER, J., KATTLER, K., WALTER, J., BLÜTHGEN, N. & HENGSTLER, J. G. 2018. Assessment of stem cell differentiation based on genome-wide expression profiles. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 373**,** 20170221.

GORDILLO, M., EVANS, T. & GOUON-EVANS, V. 2015. Orchestrating liver development. *Development,* 142**,** 2094-2108.

GORINA, R., LYCK, R., VESTWEBER, D. & ENGELHARDT, B. 2014. beta2 integrin-mediated crawling on endothelial ICAM-1 and ICAM-2 is a prerequisite for transcellular neutrophil diapedesis across the inflamed blood-brain barrier. *J Immunol,* 192**,** 324-37.

GOSSEN, M. & BUJARD, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences,* 89**,** 5547-5551.

GOSSEN, M., FREUNDLIEB, S., BENDER, G., MULLER, G., HILLEN, W. & BUJARD, H. 1995. Transcriptional activation by tetracyclines in mammalian cells. *Science,* 268**,** 1766-9.

GOV.UK. 2021. *Liver disease profiles: November 2021 update* [Online]. Available: https://www.gov.uk/government/statistics/liver-disease-profiles-november-2021-update/liver-disease-profiles-november-2021-update [Accessed 6 February 2023].

GRAFFMANN, N., SCHERER, B. & ADJAYE, J. 2022. In vitro differentiation of pluripotent stem cells into hepatocyte like cells - Basic principles and current progress. *Stem Cell Res,* 61**,** 102763.

HAN, Y., GAO, S., MUEGGE, K., ZHANG, W. & ZHOU, B. 2015. Advanced Applications of RNA Sequencing and Challenges. *Bioinformatics and Biology Insights,* 9s1**,** BBI.S28991.

HANNAN, N. R. F., FORDHAM, R. P., SYED, Y. A., MOIGNARD, V., BERRY, A., BAUTISTA, R., HANLEY, N. A., JENSEN, K. B. & VALLIER, L. 2013a. Generation of Multipotent Foregut Stem Cells from Human Pluripotent Stem Cells. *Stem Cell Reports,* 1**,** 293-306.

HANNAN, N. R. F., SAMPAZIOTIS, F., SEGERITZ, C.-P., HANLEY, N. A. & VALLIER, L. 2015. Generation of Distal Airway Epithelium from Multipotent Human Foregut Stem Cells. *Stem Cells and Development,* 24**,** 1680-1690.

HANNAN, N. R. F., SEGERITZ, C.-P., TOUBOUL, T. & VALLIER, L. 2013b. Production of hepatocyte-like cells from human pluripotent stem cells. *Nature Protocols,* 8**,** 430-437.

HAY, D. C., FLETCHER, J., PAYNE, C., TERRACE, J. D., GALLAGHER, R. C. J., SNOEYS, J., BLACK, J. R., WOJTACHA, D., SAMUEL, K., HANNOUN, Z., PRYDE, A., FILIPPI, C., CURRIE, I. S., FORBES, S. J., ROSS, J. A., NEWSOME, P. N. & IREDALE, J. P. 2008. Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *Proceedings of the National Academy of Sciences,* 105**,** 12301-12306.

HE, M.-L., WEN, L., CAMPBELL, C. E., WU, J. Y. & RAO, Y. 1999. Transcription repression by Xenopus ET and its human ortholog TBX3, a gene involved in ulnar-mammary syndrome. *Proceedings of the National Academy of Sciences,* 96**,** 10212-10217.

HEMMATI-BRIVANLOU, A. & MELTON, D. A. 1994. Inhibition of activin receptor signaling promotes neuralization in Xenopus. *Cell,* 77**,** 273-81.

HIRANO, H., JONATHAN, HORII, T., OMAR, KIMURA, M., PATRICK, NAKANE, T., ISHITANI, R., HATADA, I., ZHANG, F., NISHIMASU, H. & NUREKI, O. 2016. Structure and Engineering of Francisella novicida Cas9. *Cell,* 164**,** 950-961.

HOUSSAINT, E. 1980. Differentiation of the mouse hepatic primordium. I. An analysis of tissue interactions in hepatocyte differentiation. *Cell Differ,* 9**,** 269-79.

HUNTER, M. P., WILSON, C. M., JIANG, X., CONG, R., VASAVADA, H., KAESTNER, K. H. & BOGUE, C. W. 2007. The homeobox gene Hhex is essential for proper hepatoblast differentiation and bile duct morphogenesis. *Dev Biol,* 308**,** 355-67.

IMAIZUMI, K. & OKANO, H. 2021. Modeling neurodevelopment in a dish with pluripotent stem cells. *Development, Growth & Differentiation,* 63**,** 18-25.

ISHIDA, W., HAMAMOTO, T., KUSANAGI, K., YAGI, K., KAWABATA, M., TAKEHARA, K., SAMPATH, T. K., KATO, M. & MIYAZONO, K. 2000. Smad6 Is a Smad1/5-induced Smad Inhibitor. *Journal of Biological Chemistry,* 275**,** 6075-6079.

JEROME-MAJEWSKA, L. A., JENKINS, G. P., ERNSTOFF, E., ZINDY, F., SHERR, C. J. & PAPAIOANNOU, V. E. 2005. Tbx3, the ulnar-mammary syndrome gene, and Tbx2 interact in mammary gland development through a p19 sup Arf/p53-independent pathway. *Developmental Dynamics,* 234**,** 922-933.

JIANG, W., LIU, Y., LIU, R., ZHANG, K. & ZHANG, Y. 2015. The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. *Cell Rep,* 11**,** 137-48.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science,* 337**,** 816-821.

JUNG, J., ZHENG, M., GOLDFARB, M. & ZARET, K. S. 1999. Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science,* 284**,** 1998-2003.

JUNG, M., CORDES, S., ZOU, J., YU, S. J., GUITART, X., HONG, S. G., DANG, V., KANG, E., DONAIRES, F. S., HASSAN, S. A., ALBITAR, M., HSU, A. P., HOLLAND, S. M., HICKSTEIN, D. D., TOWNSLEY, D., DUNBAR, C. E. & WINKLER, T. 2018. GATA2 deficiency and human hematopoietic development modeled using induced pluripotent stem cells. *Blood Adv,* 2**,** 3553-3565.

KAISER, M., WOJAHN, I., RUDAT, C., LÜDTKE, T. H., CHRISTOFFELS, V. M., MOON, A., KISPERT, A. & TROWE, M.-O. 2021. Regulation of otocyst patterning by Tbx2 and Tbx3 is required for inner ear morphogenesis in the mouse. *Development,* 148**,** dev.195651.

KALLUNKI, BARISIC, JÄÄTTELÄ & LIU 2019. How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells,* 8**,** 796.

KARAULANOV, E., KNÖCHEL, W. & NIEHRS, C. 2004. Transcriptional regulation of BMP4 synexpression in transgenic Xenopus. *The EMBO Journal,* 23**,** 844-856.

KATAGIRI, T. & WATABE, T. 2016. Bone Morphogenetic Proteins. *Cold Spring Harbor Perspectives in Biology,* 8**,** a021899.

KAUFMANN, M., SCHUFFENHAUER, A., FRUH, I., KLEIN, J., THIEMEYER, A., RIGO, P., GOMEZ-MANCILLA, B., HEIDINGER-MILLOT, V., BOUWMEESTER, T., SCHOPFER, U., MUELLER, M., FODOR, B. D. & COBOS-CORREA, A. 2015. High-Throughput Screening Using iPSC-Derived Neuronal Progenitors to Identify Compounds Counteracting Epigenetic Gene Silencing in Fragile X Syndrome. *SLAS Discovery,* 20**,** 1101-1111.

KHAN, S. H. 2019. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Molecular Therapy - Nucleic Acids,* 16**,** 326-334.

KHOSHDEL RAD, N., AGHDAMI, N. & MOGHADASALI, R. 2020. Cellular and Molecular Mechanisms of Kidney Development: From the Embryo to the Kidney Organoid. *Front Cell Dev Biol,* 8**,** 183.

KIM, H. & KIM, J.-S. 2014. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics,* 15**,** 321-334.

KIM, J. Y., NAM, Y., RIM, Y. A. & JU, J. H. 2022. Review of the Current Trends in Clinical Trials Involving Induced Pluripotent Stem Cells. *Stem Cell Reviews and Reports,* 18**,** 142-154.

KIM, S.-I., OCEGUERA-YANEZ, F., SAKURAI, C., NAKAGAWA, M., YAMANAKA, S. & WOLTJEN, K. 2015. Inducible Transgene Expression in Human iPS Cells Using Versatile All-in-One piggyBac Transposons. Springer New York.

KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences,* 93**,** 1156-1160.

KLATT, D., CHENG, E., HOFFMANN, D., SANTILLI, G., THRASHER, A. J., BRENDEL, C. & SCHAMBACH, A. 2020. Differential Transgene Silencing of Myeloid-Specific Promoters in the AAVS1 Safe Harbor Locus of Induced Pluripotent Stem Cell-Derived Myeloid Cells. *Human Gene Therapy,* 31**,** 199-210.

KLEINSTIVER, B. P., PREW, M. S., TSAI, S. Q., NGUYEN, N. T., TOPKAR, V. V., ZHENG, Z. & JOUNG, J. K. 2015. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology,* 33**,** 1293-1298.

KNOCK, E. & JULIAN, L. M. 2021. Building on a Solid Foundation: Adding Relevance and Reproducibility to Neurological Modeling Using Human Pluripotent Stem Cells. *Front Cell Neurosci,* 15**,** 767457.

KOYANAGI-AOI, M., OHNUKI, M., TAKAHASHI, K., OKITA, K., NOMA, H., SAWAMURA, Y., TERAMOTO, I., NARITA, M., SATO, Y., ICHISAKA, T., AMANO, N., WATANABE, A., MORIZANE, A., YAMADA, Y., SATO, T., TAKAHASHI, J. & YAMANAKA, S. 2013. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proceedings of the National Academy of Sciences,* 110**,** 20569-20574.

KURIBAYASHI, H., BABA, Y. & WATANABE, S. 2014. BMP signaling participates in late phase differentiation of the retina, partly via upregulation of Hey2. *Dev Neurobiol,* 74**,** 1172-83.

KYLE & LIM, B. 2011. A Precarious Balance: Pluripotency Factors as Lineage Specifiers. *Cell Stem Cell,* 8**,** 363-369.

LAL, S., LI, A. & DOS REMEDIOS, C. 2016. Limitations in Translating Animal Studies to Humans in Cardiovascular Disease. *Journal of Cardiovascular Translational Research,* 9**,** 165-166.

LALLEMAND, Y., NICOLA, M.-A., RAMOS, C., BACH, A., CLOMENT, C. C. S. & ROBERT, B. T. 2005. Analysis of Msx1; Msx2 double mutants reveals multiple roles for Msx genes in limb development. *Development,* 132**,** 3003-3014.

LAMB, T. M., KNECHT, A. K., SMITH, W. C., STACHEL, S. E., ECONOMIDES, A. N., STAHL, N., YANCOPOLOUS, G. D. & HARLAND, R. M. 1993. Neural induction by the secreted polypeptide noggin. *Science,* 262**,** 713-8.

LAMBERT, S. A., JOLMA, A., CAMPITELLI, L. F., DAS, P. K., YIN, Y., ALBU, M., CHEN, X., TAIPALE, J., HUGHES, T. R. & WEIRAUCH, M. T. 2018. The Human Transcription Factors. *Cell,* 172**,** 650-665.

LAMBSHEAD, J. W., MEAGHER, L., GOODWIN, J., LABONNE, T., NG, E., ELEFANTY, A., STANLEY, E., O'BRIEN, C. M. & LASLETT, A. L. 2018. Long-Term Maintenance of Human Pluripotent Stem Cells on cRGDfK-Presenting Synthetic Surfaces. *Scientific Reports,* 8.

LANGE, L., HOFFMANN, D., SCHWARZER, A., HA, T.-C., PHILIPP, F., LENZ, D., MORGAN, M. & SCHAMBACH, A. 2020. Inducible Forward Programming of Human Pluripotent Stem Cells to Hemato-endothelial Progenitor Cells with Hematopoietic Progenitor Potential. *Stem Cell Reports,* 14**,** 122-137.

LE DOUARIN, N. 1968. [Synthesis of glycogen in hepatocytes undergoing differentiation: role of homologous and heterologous mesenchyma]. *Dev Biol,* 17**,** 101-14.

LEE, C. S., FRIEDMAN, J. R., FULMER, J. T. & KAESTNER, K. H. 2005. The initiation of liver development is dependent on Foxa transcription factors. *Nature,* 435**,** 944-947.

LEE, Y.-K., HO, P. W.-L., SCHICK, R., LAU, Y.-M., LAI, W.-H., ZHOU, T., LI, Y., NG, K.-M., HO, S.-L., ESTEBAN, M. A., BINAH, O., TSE, H.-F. & SIU, C.-W. 2014. Modeling of Friedreich ataxia-related iron overloading cardiomyopathy using patient-specific-induced pluripotent stem cells. *Pflügers Archiv - European Journal of Physiology,* 466**,** 1831-1844.

LEE, Y. H., SAUER, B. & GONZALEZ, F. J. 1998. Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse. *Mol Cell Biol,* 18**,** 3059-68.

LEVINE, F. & OGUNWOBI, O. O. 2021. Targeting PVT1 Exon 9 Re-Expresses Claudin 4 Protein and Inhibits Migration by Claudin—Low Triple Negative Breast Cancer Cells. *Cancers,* 13**,** 1046.

LI, J., NING, G. & DUNCAN, S. A. 2000. Mammalian hepatocyte differentiation requires the transcription factor HNF-4alpha. *Genes Dev,* 14**,** 464-74.

LI, Y., DENG, Y., ZHAO, Y., ZHANG, W., ZHANG, S., ZHANG, L., WANG, B., XU, Y. & CHEN, S. 2022. Immunoglobulin superfamily 9 (IGSF9) is trans-activated by p53, inhibits breast cancer metastasis via FAK. *Oncogene,* 41**,** 4658-4672.

LOKMANE, L., HAUMAITRE, C. C., GARCIA-VILLALBA, P., ANSELME, I., SCHNEIDER-MAUNOURY, S. & CEREGHINI, S. 2008. Crucial role of vHNF1 in vertebrate hepatic specification. *Development,* 135**,** 2777-2786.

LOPATKA, A. & MOON, A. M. 2022. Complex functional redundancy of Tbx2 and Tbx3 in mouse limb development. *Dev Dyn*.

LU, T., YANG, M., HUANG, D.-B., WEI, H., OZER, G. H., GHOSH, G. & STARK, G. R. 2013. Role of lysine methylation of NF-κB in differential gene regulation. *Proceedings of the National Academy of Sciences,* 110**,** 13510-13515.

LUDTKE, T. H., RUDAT, C., WOJAHN, I., WEISS, A. C., KLEPPA, M. J., KURZ, J., FARIN, H. F., MOON, A., CHRISTOFFELS, V. M. & KISPERT, A. 2016. Tbx2 and Tbx3 Act Downstream of Shh to Maintain Canonical Wnt Signaling

during Branching Morphogenesis of the Murine Lung. *Dev Cell,* 39**,** 239-253.

LÜDTKE, T. H. W., CHRISTOFFELS, V. M., PETRY, M. & KISPERT, A. 2009. Tbx3 promotes liver bud expansion during mouse development by suppression of cholangiocyte differentiation. *Hepatology,* 49**,** 969-978.

LÜDTKE, T. H. W., FARIN, H. F., RUDAT, C., SCHUSTER-GOSSLER, K., PETRY, M., BARNETT, P., CHRISTOFFELS, V. M. & KISPERT, A. 2013. Tbx2 Controls Lung Growth by Direct Repression of the Cell Cycle Inhibitor Genes Cdkn1a and Cdkn1b. *PLoS Genetics,* 9**,** e1003189.

LV, G. & FAN, J. 2020. Silencing ICAM-1 reduces the adhesion of vascular endothelial cells in mice with immunologic contact urticaria. *Gene,* 760**,** 144965.

MAQSOOD, M. I., MATIN, M. M., BAHRAMI, A. R. & GHASROLDASHT, M. M. 2013. Immortality of cell lines: challenges and advantages of establishment. *Cell Biol Int,* 37**,** 1038-45.

MARGAGLIOTTI, S., CLOTMAN, F., PIERREUX, C. E., BEAUDRY, J. B., JACQUEMIN, P., ROUSSEAU, G. G. & LEMAIGRE, F. P. 2007. The Onecut transcription factors HNF-6/OC-1 and OC-2 regulate early liver expansion by controlling hepatoblast migration. *Dev Biol,* 311**,** 579-89.

MARTINEZ BARBERA, J. P., CLEMENTS, M., THOMAS, P., RODRIGUEZ, T., MELOY, D., KIOUSSIS, D. & BEDDINGTON, R. S. 2000. The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development,* 127**,** 2433-2445.

MASUI, S., NAKATAKE, Y., TOYOOKA, Y., SHIMOSATO, D., YAGI, R., TAKAHASHI, K., OKOCHI, H., OKUDA, A., MATOBA, R., SHAROV, A. A., KO, M. S. H. & NIWA, H. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology,* 9**,** 625-635.

MATSUMOTO, K., YOSHITOMI, H., ROSSANT, J. & ZARET, K. S. 2001. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science,* 294**,** 559-63.

MCLIN, V. R. A., RANKIN, S. A. & ZORN, A. M. 2007. Repression of Wnt/β-catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development,* 134**,** 2207-2217.

MCMAHON, A. P. 2016. Development of the Mammalian Kidney. Elsevier.

MITRY, R. R., HUGHES, R. D. & DHAWAN, A. 2002. Progress in human hepatocytes: isolation, culture & cryopreservation. *Semin Cell Dev Biol,* 13**,** 463-7.

MIYAZAKI, T., FUTAKI, S., SUEMORI, H., TANIGUCHI, Y., YAMADA, M., KAWASAKI, M., HAYASHI, M., KUMAGAI, H., NAKATSUJI, N., SEKIGUCHI, K. & KAWASE, E. 2012. Laminin E8 fragments support efficient adhesion and expansion of dissociated human pluripotent stem cells. *Nature Communications,* 3**,** 1236.

MORERA, C., KIM, J., PAREDES-REDONDO, A., NOBLES, M., RYBIN, D., MOCCIA, R., KOWALA, A., MENG, J., GARREN, S., LIU, P., MORGAN, J. E., MUNTONI, F., CHRISTOFOROU, N., OWENS, J., TINKER, A. & LIN, Y.-Y. 2022. CRISPR-mediated correction of skeletal muscle Ca2+ handling in a novel DMD

patient-derived pluripotent stem cell model. *Neuromuscular Disorders,* 32**,** 908-922.

MUKHERJEE, S., FRENCH, D. L. & GADUE, P. 2021. Loss of TBX3 enhances pancreatic progenitor generation from human pluripotent stem cells. *Stem Cell Reports,* 16**,** 2617-2627.

MUNDADE, R., OZER, H. G., WEI, H., PRABHU, L. & LU, T. 2014. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle,* 13**,** 2847-2852.

NAVA, S., WESTGREN, M., JAKSCH, M., TIBELL, A., BROOME, U., ERICZON, B. G. & SUMITRAN-HOLGERSSON, S. 2005. Characterization of cells in the developing human liver. *Differentiation,* 73**,** 249-60.

NELSEN, S. M. & CHRISTIAN, J. L. 2009. Site-specific Cleavage of BMP4 by Furin, PC6, and PC7. *Journal of Biological Chemistry,* 284**,** 27157-27166.

NICKEL, J. & MUELLER, T. D. 2019. Specification of BMP Signaling. *Cells,* 8**,** 1579.

OCEGUERA-YANEZ, F., AVILA-ROBINSON, A. & WOLTJEN, K. 2022. Differentiation of pluripotent stem cells for modeling human skin development and potential applications. *Front Cell Dev Biol,* 10**,** 1030339.

ODOM, D. T., DOWELL, R. D., JACOBSEN, E. S., GORDON, W., DANFORD, T. W., MACISAAC, K. D., ROLFE, P. A., CONBOY, C. M., GIFFORD, D. K. & FRAENKEL, E. 2007. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genetics,* 39**,** 730-732.

ONICHTCHOUK, D., CHEN, Y.-G., DOSCH, R., GAWANTKA, V., DELIUS, H., MASSAGUE´, J. & NIEHRS, C. 1999. Silencing of TGF-β signalling by the pseudoreceptor BAMBI. *Nature,* 401**,** 480-485.

ORDOVAS, L., BOON, R., PISTONI, M., CHEN, Y., WOLFS, E., GUO, W., SAMBATHKUMAR, R., BOBIS-WOZOWICZ, S., HELSEN, N., VANHOVE, J., BERCKMANS, P., CAI, Q., VANUYTSEL, K., EGGERMONT, K., VANSLEMBROUCK, V., SCHMIDT, B. Z., RAITANO, S., VAN DEN BOSCH, L., NAHMIAS, Y., CATHOMEN, T., STRUYS, T. & VERFAILLIE, C. M. 2018. Efficient Recombinase-Mediated Cassette Exchange in hPSCs to Study the Hepatocyte Lineage Reveals AAVS1 Locus-Mediated Transgene Inhibition. *Stem Cell Reports,* 10**,** 673.

OVALLE, W. K. N., PATRICK C. 2021. *Netter's Essential Histology,* Philadelphia, Elsevier.

PARK, I.-H., ARORA, N., HUO, H., MAHERALI, N., AHFELDT, T., SHIMAMURA, A., LENSCH, M. W., COWAN, C., HOCHEDLINGER, K. & DALEY, G. Q. 2008. Disease-Specific Induced Pluripotent Stem Cells. *Cell,* 134**,** 877-886.

PARVIZ, F., MATULLO, C., GARRISON, W. D., SAVATSKI, L., ADAMSON, J. W., NING, G., KAESTNER, K. H., ROSSI, J. M., ZARET, K. S. & DUNCAN, S. A. 2003. Hepatocyte nuclear factor 4α controls the development of a hepatic epithelium and liver morphogenesis. *Nature Genetics,* 34**,** 292-296.

PAULSEN, M., LEGEWIE, S., EILS, R., KARAULANOV, E. & NIEHRS, C. 2011. Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signaling range and canalizes

development. *Proceedings of the National Academy of Sciences,* 108**,** 10202-10207.

PENNY, G. D., KAY, G. F., SHEARDOWN, S. A., RASTAN, S. & BROCKDORFF, N. 1996. Requirement for Xist in X chromosome inactivation. *Nature,* 379**,** 131-7.

RAMIREZ-CALDERON, G., COLOMBO, G., HERNANDEZ-BAUTISTA, C. A., ASTRO, V. & ADAMO, A. 2022. Heart in a Dish: From Traditional 2D Differentiation Protocols to Cardiac Organoids. *Front Cell Dev Biol,* 10**,** 855966.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

RICHTER, A., VALDIMARSDOTTIR, L., HRAFNKELSDOTTIR, H. E., RUNARSSON, J. F., OMARSDOTTIR, A. R., WARD-VAN OOSTWAARD, D., MUMMERY, C. & VALDIMARSDOTTIR, G. 2014. BMP4 promotes EMT and mesodermal commitment in human embryonic stem cells via SLUG and MSX2. *Stem Cells,* 32**,** 636-48.

ROSSI, J. M., DUNN, N. R., HOGAN, B. L. M. & ZARET, K. S. 2001. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes & Development,* 15**,** 1998-2009.

SADELAIN, M., PAPAPETROU, E. P. & BUSHMAN, F. D. 2012. Safe harbours for the integration of new DNA in the human genome. *Nature Reviews Cancer,* 12**,** 51-58.

SAKABE, N. J., ANEAS, I., SHEN, T., SHOKRI, L., PARK, S.-Y., BULYK, M. L., EVANS, S. M. & NOBREGA, M. A. 2012. Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function. *Human Molecular Genetics,* 21**,** 2194-2204.

SAMPAZIOTIS, F., CARDOSO DE BRITO, M., MADRIGAL, P., BERTERO, A., SAEB-PARSY, K., SOARES, F. A. C., SCHRUMPF, E., MELUM, E., KARLSEN, T. H., BRADLEY, J. A., GELSON, W. T. H., DAVIES, S., BAKER, A., KASER, A., ALEXANDER, G. J., HANNAN, N. R. F. & VALLIER, L. 2015. Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nature Biotechnology,* 33**,** 845-852.

SANCHEZ-DUFFHUES, G., WILLIAMS, E., GOUMANS, M. J., HELDIN, C. H. & TEN DIJKE, P. 2020. Bone morphogenetic protein receptors: Structure, function and targeting by selective small molecule kinase inhibitors. *Bone,* 138**,** 115472.

SCHLIERMANN, A. & NICKEL, J. 2018. Unraveling the Connection between Fibroblast Growth Factor and Bone Morphogenetic Protein Signaling. *International Journal of Molecular Sciences,* 19**,** 3220.

SERLS, A. E., DOHERTY, S., PARVATIYAR, P., WELLS, J. M. & DEUTSCH, G. H. 2005. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development,* 132**,** 35-47.

SETH, A., YE, J., YU, N., GUEZ, F., BEDFORD, D. C., NEALE, G. A., CORDI, S., BRINDLE, P. K., LEMAIGRE, F. P., KAESTNER, K. H. & SOSA-PINEDA, B.

2014. Prox1 ablation in hepatic progenitors causes defective hepatocyte specification and increases biliary cell commitment. *Development,* 141**,** 538-547.

SHARFF, K. A., SONG, W.-X., LUO, X., TANG, N., LUO, J., CHEN, J., BI, Y., HE, B.-C., HUANG, J., LI, X., JIANG, W., ZHU, G.-H., SU, Y., HE, Y., SHEN, J., WANG, Y., CHEN, L., ZUO, G.-W., LIU, B., PAN, X., REID, R. R., LUU, H. H., HAYDON, R. C. & HE, T.-C. 2009. Hey1 Basic Helix-Loop-Helix Protein Plays an Important Role in Mediating BMP9-induced Osteogenic Differentiation of Mesenchymal Progenitor Cells. *Journal of Biological Chemistry,* 284**,** 649-659.

SHARMA, A., MARCEAU, C., HAMAGUCHI, R., BURRIDGE, P. W., RAJARAJAN, K., CHURKO, J. M., WU, H., SALLAM, K. I., MATSA, E., STURZU, A. C., CHE, Y., EBERT, A., DIECKE, S., LIANG, P., RED-HORSE, K., CARETTE, J. E., WU, S. M. & WU, J. C. 2014. Human Induced Pluripotent Stem Cell–Derived Cardiomyocytes as an In Vitro Model for Coxsackievirus B3–Induced Myocarditis and Antiviral Drug Screening Platform. *Circulation Research,* 115**,** 556-566.

SHEN, B., ZHANG, W., ZHANG, J., ZHOU, J., WANG, J., CHEN, L., WANG, L., HODGKINS, A., IYER, V., HUANG, X. & SKARNES, W. C. 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods,* 11**,** 399-402.

SHERWOOD, R. I., CHEN, T.-Y. A. & MELTON, D. A. 2009. Transcriptional dynamics of endodermal organ formation. *Developmental Dynamics,* 238**,** 29-42.

SHERWOOD, R. I., MAEHR, R., MAZZONI, E. O. & MELTON, D. A. 2011. Wnt signaling specifies and patterns intestinal endoderm. *Mech Dev,* 128**,** 387-400.

SHIMOBABA, S., TAGA, S., AKIZUKI, R., HICHINO, A., ENDO, S., MATSUNAGA, T., WATANABE, R., YAMAGUCHI, M., YAMAZAKI, Y., SUGATANI, J. & IKARI, A. 2016. Claudin-18 inhibits cell proliferation and motility mediated by inhibition of phosphorylation of PDK1 and Akt in human lung adenocarcinoma A549 cells. *Biochim Biophys Acta,* 1863**,** 1170-8.

SHIN, D., SHIN, C. H., TUCKER, J., OBER, E. A., RENTZSCH, F., POSS, K. D., HAMMERSCHMIDT, M., MULLINS, M. C. & STAINIER, D. Y. R. 2007. Bmp and Fgf signaling are essential for liver specification in zebrafish. *Development,* 134**,** 2041-2050.

SI-TAYEB, K., LEMAIGRE, F. P. & DUNCAN, S. A. 2010a. Organogenesis and Development of the Liver. *Developmental Cell,* 18**,** 175-189.

SI-TAYEB, K., NOTO, F. K., NAGAOKA, M., LI, J., BATTLE, M. A., DURIS, C., NORTH, P. E., DALTON, S. & DUNCAN, S. A. 2010b. Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology,* 51**,** 297-305.

SINGH, R., HOOGAARS, W. M., BARNETT, P., GRIESKAMP, T., RANA, M. S., BUERMANS, H., FARIN, H. F., PETRY, M., HEALLEN, T., MARTIN, J. F., MOORMAN, A. F. M., 'T HOEN, P. A. C., KISPERT, A. & CHRISTOFFELS, V. M. 2012. Tbx2 and Tbx3 induce atrioventricular myocardial development

and endocardial cushion formation. *Cellular and Molecular Life Sciences,* 69**,** 1377-1389.

SOLOTKE, M. T., DHRUVA, S. S., DOWNING, N. S., SHAH, N. D. & ROSS, J. S. 2018. New and incremental FDA black box warnings from 2008 to 2015. *Expert Opinion on Drug Safety,* 17**,** 117-123.

SOSA-PINEDA, B., WIGLE, J. T. & OLIVER, G. 2000. Hepatocyte migration during liver development requires Prox1. *Nature Genetics,* 25**,** 254-255.

SUZUKI, A., SEKIYA, S., BÜSCHER, D., IZPISÚA BELMONTE, J. C. & TANIGUCHI, H. 2008. Tbx3 controls the fate of hepatic progenitor cells in liver development by suppressing p19ARF expression. *Development,* 135**,** 1589-1595.

SZCZEPEK, M., BRONDANI, V., BÜCHEL, J., SERRANO, L., SEGAL, D. J. & CATHOMEN, T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnology,* 25**,** 786-793.

SZKOLNICKA, D., FARNWORTH, S. L., LUCENDO-VILLARIN, B., STORCK, C., ZHOU, W., IREDALE, J. P., FLINT, O. & HAY, D. C. 2014. Accurate Prediction of Drug-Induced Liver Injury Using Stem Cell-Derived Populations. *Stem Cells Translational Medicine,* 3**,** 141-148.

SZKOLNICKA, D. & HAY, D. C. 2016. Concise Review: Advances in Generating Hepatocytes from Pluripotent Stem Cells for Translational Medicine. *Stem Cells,* 34**,** 1421-6.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell,* 131**,** 861-872.

TAKEBE, T., SEKINE, K., KIMURA, M., YOSHIZAWA, E., AYANO, S., KOIDO, M., FUNAYAMA, S., NAKANISHI, N., HISAI, T., KOBAYASHI, T., KASAI, T., KITADA, R., MORI, A., AYABE, H., EJIRI, Y., AMIMOTO, N., YAMAZAKI, Y., OGAWA, S., ISHIKAWA, M., KIYOTA, Y., SATO, Y., NOZAWA, K., OKAMOTO, S., UENO, Y. & TANIGUCHI, H. 2017. Massive and Reproducible Production of Liver Buds Entirely from Human Pluripotent Stem Cells. *Cell Reports,* 21**,** 2661-2670.

TARAZI, S., AGUILERA-CASTREJON, A., JOUBRAN, C., GHANEM, N., ASHOUOKHI, S., RONCATO, F., WILDSCHUTZ, E., HADDAD, M., OLDAK, B., GOMEZ-CESAR, E., LIVNAT, N., VIUKOV, S., LOKSHTANOV, D., NAVEH-TASSA, S., ROSE, M., HANNA, S., RAANAN, C., BRENNER, O., KEDMI, M., KEREN-SHAUL, H., LAPIDOT, T., MAZA, I., NOVERSHTERN, N. & HANNA, J. H. 2022. Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. *Cell,* 185**,** 3290-3306.e25.

TERRY, C., MITRY, R. R., LEHEC, S. C., MUIESAN, P., RELA, M., HEATON, N. D., HUGHES, R. D. & DHAWAN, A. 2005. The effects of cryopreservation on human hepatocytes obtained from different sources of liver tissue. *Cell Transplant,* 14**,** 585-94.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science,* 282**,** 1145-7.

TORTORA, G. J. 2011. Principles of anatomy and physiology / Gerard J. Tortora, Bryan Derrickson.

TOUBOUL, T., CHEN, S., TO, C. C., MORA-CASTILLA, S., SABATINI, K., TUKEY, R. H. & LAURENT, L. C. 2016. Stage-specific regulation of the WNT/β-catenin pathway enhances differentiation of hESCs into hepatocytes. *Journal of Hepatology,* 64**,** 1315-1326.

TOUBOUL, T., HANNAN, N. R. F., CORBINEAU, S., MARTINEZ, A., MARTINET, C., BRANCHEREAU, S., MAINOT, S., STRICK-MARCHAND, H., PEDERSEN, R., DI SANTO, J., WEBER, A. & VALLIER, L. 2010. Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology,* 51**,** 1754-1765.

TREMBLAY, K. D. & ZARET, K. S. 2005. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Dev Biol,* 280**,** 87-99.

TREMBLAY, M., SANCHEZ-FERRAS, O. & BOUCHARD, M. 2018. GATA transcription factors in development and disease. *Development,* 145**,** dev164384.

TWAROSKI, K., MALLANNA, S. K., JING, R., DIFURIO, F., URICK, A. & DUNCAN, S. A. 2015. FGF2 mediates hepatic progenitor cell formation during human pluripotent stem cell differentiation by inducing the WNT antagonist NKD1. *Genes & Development,* 29**,** 2463-2474.

VAINIO, S., KARAVANOVA, I., JOWETT, A. & THESLEFF, I. 1993. Identification of BMP-4 as a signal mediating secondary induction between epithelial and mesenchymal tissues during early tooth development. *Cell,* 75**,** 45-58.

VEGA, S. L., ARVIND, V., MISHRA, P., KOHN, J., SANJEEVA MURTHY, N. & MOGHE, P. V. 2018. Substrate micropatterns produced by polymer demixing regulate focal adhesions, actin anisotropy, and lineage differentiation of stem cells. *Acta Biomaterialia,* 76**,** 21-28.

VUKICEVIC, S., KLEINMAN, H. K., LUYTEN, F. P., ROBERTS, A. B., ROCHE, N. S. & REDDI, A. H. 1992. Identification of multiple active growth factors in basement membrane Matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. *Exp Cell Res,* 202**,** 1-8.

WALTERS, L. 2004. Human Embryonic Stem Cell Research: An Intercultural Perspective. *Kennedy Institute of Ethics Journal,* 14**,** 3-38.

WANG, Y., YU, M., HAO, K., LEI, W., TANG, M. & HU, S. 2022. Cardiomyocyte Maturation–the Road is not Obstructed. *Stem Cell Reviews and Reports,* 18**,** 2966-2981.

WANG, Z., ORON, E., NELSON, B., RAZIS, S. & IVANOVA, N. 2012. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell,* 10**,** 440-454.

WATT, A. J., ZHAO, R., LI, J. & DUNCAN, S. A. 2007. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Developmental Biology,* 7**,** 37.

WEAVER, R. J. & VALENTIN, J. P. 2019. Today's Challenges to De-Risk and Predict Drug Safety in Human "Mind-the-Gap". *Toxicol Sci,* 167**,** 307-321.

WEBB, P. G., SPILLMAN, M. A. & BAUMGARTNER, H. K. 2013. Claudins play a role in normal and tumor cell motility. *BMC Cell Biology,* 14**,** 19.

WEBER, D., WIESE, C. & GESSLER, M. 2014. Hey bHLH transcription factors. *Curr Top Dev Biol,* 110**,** 285-315.

WHITFIELD, T. W., WANG, J., COLLINS, P. J., PARTRIDGE, E. C., ALDRED, S., TRINKLEIN, N. D., MYERS, R. M. & WENG, Z. 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome Biology,* 13**,** R50.

WHITLEY, S. K., HORNE, W. T. & KOLLS, J. K. 2016. Research Techniques Made Simple: Methodology and Clinical Applications of RNA Sequencing. *Journal of Investigative Dermatology,* 136**,** e77-e82.

WONG, A. P., BEAR, C. E., CHIN, S., PASCERI, P., THOMPSON, T. O., HUAN, L.-J., RATJEN, F., ELLIS, J. & ROSSANT, J. 2012. Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nature Biotechnology,* 30**,** 876-882.

XIA, Y., CARPENTIER, A., CHENG, X., BLOCK, P. D., ZHAO, Y., ZHANG, Z., PROTZER, U. & LIANG, T. J. 2017. Human stem cell-derived hepatocytes as a model for hepatitis B virus infection, spreading and virus-host interactions. *Journal of Hepatology,* 66**,** 494-503.

XIE, J., ZHANG, D., ZHOU, C., YUAN, Q., YE, L. & ZHOU, X. 2018. Substrate elasticity regulates adipose-derived stromal cell differentiation towards osteogenesis and adipogenesis through beta-catenin transduction. *Acta Biomater,* 79**,** 83-95.

YAMANAKA, S. 2020. Pluripotent Stem Cell-Based Cell Therapy—Promise and Challenges. *Cell Stem Cell,* 27**,** 523-531.

YAMAZOE, T., SHIRAKI, N., TOYODA, M., KIYOKAWA, N., OKITA, H., MIYAGAWA, Y., AKUTSU, H., UMEZAWA, A., SASAKI, Y., KUME, K. & KUME, S. 2013. A synthetic nanofibrillar matrix promotes in vitro hepatic differentiation of embryonic stem cells and induced pluripotent stem cells. *Journal of Cell Science,* 126**,** 5391-5399.

YANAI, M., TATSUMI, N., HASUNUMA, N., KATSU, K., ENDO, F. & YOKOUCHI, Y. 2008. FGF signaling segregates biliary cell-lineage from chick hepatoblasts cooperatively with BMP4 and ECM components in vitro. *Developmental Dynamics,* 237**,** 1268-1283.

YANG, S., TUO, C., IU, E. & PLOTNIKOV, S. V. 2021. Control of Cell Geometry through Infrared Laser Assisted Micropatterning. *J Vis Exp*.

YOSHIDA, T., TAKAYAMA, K., KONDOH, M., SAKURAI, F., TANI, H., SAKAMOTO, N., MATSUURA, Y., MIZUGUCHI, H. & YAGI, K. 2011. Use of human hepatocyte-like cells derived from induced pluripotent stem cells as a model for hepatocytes in hepatitis C virus infection. *Biochem Biophys Res Commun,* 416**,** 119-24.

ZARET, K. S. 2016. From Endoderm to Liver Bud: Paradigms of Cell Type Specification and Tissue Morphogenesis. *Curr Top Dev Biol,* 117**,** 647-69.

ZENG, J., YI, D., SUN, W., LIU, Y., CHANG, J., ZHU, L., ZHANG, Y., PAN, X., DONG, Y., ZHOU, Y., LAI, M., BIAN, G., ZHOU, Q., LIU, J., CHEN, B. & MA, F. 2021. Overexpression of HOXA9 upregulates NF-κB signaling to promote

human hematopoiesis and alter the hematopoietic differentiation potentials. *Cell Regeneration,* 10.

ZHANG, C. & GUO, Z. M. 2015. Multiple functions of Maf in the regulation of cellular development and differentiation. *Diabetes/Metabolism Research and Reviews,* 31**,** 773-778.

ZHANG, S. 2014. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World Journal of Stem Cells,* 6**,** 305.

ZHANG, W., YATSKIEVYCH, T. A., BAKER, R. K. & ANTIN, P. B. 2004. Regulation of Hex gene expression and initial stages of avian hepatogenesis by Bmp and Fgf signaling. *Dev Biol,* 268**,** 312-26.

ZHANG, X., ZHANG, T., LIU, B., ZHANG, Y., JI, Z. & WANG, X. 2022. Effects of Biomimetic Micropatterned Surfaces on the Adhesion and Morphology of Cervical Cancer Cells. *ACS Omega,* 7**,** 19913-19919.

ZHANG, Y. & QUE, J. 2020. BMP Signaling in Development, Stem Cells, and Diseases of the Gastrointestinal Tract. *Annu Rev Physiol,* 82**,** 251-273.

ZHAO, H., YANG, T., MADAKASHIRA, B. P., THIELS, C. A., BECHTLE, C. A., GARCIA, C. M., ZHANG, H., YU, K., ORNITZ, D. M., BEEBE, D. C. & ROBINSON, M. L. 2008. Fibroblast growth factor receptor signaling is essential for lens fiber cell differentiation. *Dev Biol,* 318**,** 276-88.

ZHAO, R., WATT, A. J., LI, J., LUEBKE-WHEELER, J., MORRISEY, E. E. & DUNCAN, S. A. 2005. GATA6 Is Essential for Embryonic Development of the Liver but Dispensable for Early Heart Formation. *Molecular and Cellular Biology,* 25**,** 2622-2631.

ZHAO, Y., LI, X., SONG, G., LI, Q., YAN, H. & CUI, Z. 2022. Comparative Transcriptome Analysis Provides Novel Molecular Events for the Differentiation and Maturation of Hepatocytes during the Liver Development of Zebrafish. *Biomedicines,* 10**,** 2264.

ZHOU, S. W., WANG, J., CHEN, S. Y., REN, K. F., WANG, Y. X. & JI, J. 2022. The substrate stiffness at physiological range significantly modulates vascular cell behavior. *Colloids Surf B Biointerfaces,* 214**,** 112483.

ZORN, A. M. 2008. Liver development. *StemBook.* Cambridge (MA).