Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering, and Natural Sciences

of the

Ruprecht-Karls-University

Heidelberg, Germany

Presented by

Bryce Lim, MSci

Born in: Singapore, 1996

Oral examination: 8th September, 2023

Active repression of alternative cell fates
safeguards hepatocyte identity

Referees: Prof. Dr Ingrid Lohmann
Prof. Dr Duncan Odom

Abstract

Cell identity plasticity is a normal and essential feature of cells during development, but is also a hallmark of diseases such as cancer. The mechanisms involved in suppressing inappropriate cell fates are poorly understood. Gene repression in terminally-differentiated cells is typically thought to be mediated by passive epigenetic silencing, such as through DNA methylation and repressive histone modifications.

However, recent studies suggest that active repression by cell type-specific transcription factors may play an essential role in maintaining cell identity. MYT1L, a neuron-specific transcriptional repressor, has been shown to maintain neuronal identity by the inhibition of non-neuronal lineages. The observation that deficiencies in MYT1L may lead to diseases such as neurodevelopmental disorders and cancer demonstrates its important role in cell identity maintenance by inhibition of cellular plasticity.

In this thesis, I investigate whether active, lifelong transcriptional repression of alternate lineages – or *safeguard repression* – is a general mechanism for cell identity stabilisation. Using computational prediction and direct cell reprogramming, I demonstrate that PROX1 safeguards hepatocyte identity by directly suppressing alternate lineage programs, hence demonstrating safeguard repression in a second cell type and germ layer. With single-cell transcriptomics analysis, I find that PROX1 robustly silences donor and alternative cell identities.

I also show that PROX1 directly binds to promoters of key mesodermal transcription factors, including *Prrx1* and *Pparg*, hence decreasing their chromatin accessibility and expression. In addition, repression of PROX1 target genes during hepatocyte reprogramming mimicked effects of *Prox1* overexpression. On the other hand, *Prox1* deletion or target gene activation permitted inappropriate gene expression.

Finally, in a hepatocellular carcinoma mouse model, *Prox1* overexpression prevented tumour initiation, reducing overall tumour load and extending survival substantially. *Prox1* depletion in the same model induced cholangiocarcinoma-like morphology and gene expression, demonstrating the role of PROX1 in cell identity maintenance.

These results show that PROX1 is a repressive safeguard of hepatocyte identity. They also support a model whereby continuous silencing of alternate lineage gene expression programs by safeguard repressors prevents cell fate plasticity, thereby maintaining cell identity.

Zusammenfassung

Die Plastizität der Zellidentität ist ein normales und lebensnotwendiges Merkmal von Zellen in der Entwicklung, aber auch ein Kennzeichen von Krankheiten wie beispielsweise Krebs. Die Mechanismen, die an der Unterdrückung ungewollter Plastizität beteiligt sind unzureichend bekannt. Man geht davon aus, dass die Unterdrückung von Genen in terminal differenzierten Zellen durch passives epigenetisches Stummschalten wie DNA-Methylierung und repressive Histonmodifikationen erfolgt.

Neuere Studien deuten jedoch darauf hin, dass die aktive Unterdrückung durch zelltypspezifische Transkriptionsfaktoren für die Aufrechterhaltung der Zellidentität wichtig sein könnte. MYT1L, ein neuronenspezifischer Transkriptionsrepressor, erhält die neuronale Identität durch die Hemmung nicht-neuronaler Gene aufrecht. Die Beobachtung, dass ein Mangel an MYT1L zu Krankheiten wie neurologischen Entwicklungsstörungen und Krebs führen kann, zeigt seine wichtige Rolle bei der Aufrechterhaltung der Zellidentität durch Hemmung der zellulären Plastizität.

In dieser Studie untersuche ich, ob die aktive, lebenslange transkriptionelle Unterdrückung alternierender Zellidentitäten – oder *Safeguard-Repression* – ein allgemeiner Mechanismus zur Stabilisierung der Zellidentität ist. Mithilfe von computergestützten Vorhersagen und direkter Zellreprogrammierung zeige ich, dass PROX1 die Identität von Hepatozyten schützt, indem es alternative Genprogramme direkt unterdrückt und somit die Safeguard-Repression in einem zweiten Zelltyp und Keimblatt demonstriert. Anhand von Einzelzell-Transkriptomanalysen konnte ich feststellen, dass PROX1 die Spender- und alternativen Zellidentitäten robust unterdrückt.

PROX1 bindet direkt an Promotoren wichtiger mesodermaler Transkriptionsfaktoren, einschließlich *Prrx1* und *Pparg*, und verringert so deren Chromatin-Zugänglichkeit und Expression. Die Unterdrückung von PROX1-Zielgenen während der Reprogrammierung von Hepatozyten ahmte die Effekte der Überexpression von *Prox1* nach. Andererseits ermöglichte die Deletion von *Prox1* oder die Aktivierung von Zielgenen eine unerwünschte Genexpression.

Im Mausmodell für hepatozelluläre Karzinome verhinderte die Überexpression von *Prox1* die Tumorentstehung, reduzierte die Gesamttumorlast und verlängerte die Überlebenszeit erheblich. Der Verlust von *Prox1* führte im gleichen Modell zu einer Morphologie und Genexpression, die dem Cholangiokarzinom ähnelt, was die Rolle von PROX1 bei der Erhaltung der Zellidentität belegt.

Diese Ergebnisse zeigen, dass PROX1 ein Safeguard-Repressor der Hepatozytenidentität ist. Meine Daten unterstützen ein Modell, bei dem die kontinuierliche Unterdrückung von Genexpressionsprogrammen für alternative Zellidentitäten durch terminale Repressoren die Plastizität des Zellschicksals verhindert und dadurch die Zellidentität aufrechterhält.

Contributions

The work presented in this thesis was performed in collaboration with others. Outside contributions to experiments are listed here:

- **Dr Moritz Mall** and **Dr Judith Zaugg** planned, implemented, and guided this project.
- **Dr Moritz Mall**, **Sarah Grieder**, and **Kanela Kafetzopoulou** performed initial validation experiments on hepatocyte candidates.
- **Dr Juan Adrian-Segarra** performed the majority of early reprogramming experiments, including cloning of overexpression constructs. Later experiments were performed in close collaboration with him.
- **Dr Juan Adrian-Segarra** planned and established the single-cell RNA-seq, ATAC-seq, and CUT&RUN protocols. Experiments using these techniques were performed in close collaboration with him.
- **Dr Taishi Kanamaru**, **Eric Poisel**, and **Brenda Besemer** provided support in performing reprogramming experiments.
- **Kai Volz**, **Lio Böse**, and **Luise Butthof**, with supervision from **Prof. Dr Darjus Tscharhaganeh**, performed mouse hydrodynamic tail vein experiments.
- **Danijela Heide**, **Jenny Hetzer**, and **Tim Machauer** performed histological mounting, cutting, and staining of mouse livers, and **Florian Müller** provided technical support for liver experiments, all with supervision from **Prof. Dr Mathias Heikenwälder**.
- **Suchira Gallage**, **Jose Efren Barragan Avila**, and **Hendrik Wiethoff**, with supervision from **Prof. Dr Mathias Heikenwälder** and **Prof. Dr Darjus Tscharhaganeh**, performed analysis of mouse liver histological samples.
- **Sarah Schnabellehner**, with supervision from **Dr Taija Mäkinen**, provided *Prox1*$^{fl/fl}$ mouse embryonic fibroblasts.

- **Aryan Kamal**, with supervision from **Dr Judith Zaugg**, performed bioinformatics analysis of single-cell RNA-seq data, gene regulatory network analysis of bulk RNA-seq data, and integrative analysis of multi-omics datasets.
- **Dr Ignacio Ibarra**, with supervision from **Dr Judith Zaugg**, performed the *in silico* screen for terminal repressor candidates.
- **Dr Christian Arnold** provided pipelines and advice for the analysis of bulk RNA-seq data.
- Sequencing experiments were performed with support from the **DKFZ core facilities**, specifically the scOpenLab (**Dr Jan-Philipp Mallm**), Sequencing Open Lab (**Nina Glaser**), and Genomics and Proteomics Core Facility (**Dr Angela Schulz**)
- High-throughput microscopy was performed with support from the Nikon Imaging Centre, University of Heidelberg (**Dr Christian Ackermann**).

Acknowledgements

My PhD has been a rewarding journey, and throughout it I have enjoyed the company and support of many. Firstly, I would like to thank my supervisor Moritz Mall for his guidance and support throughout my time at the DKFZ.

To the members of the Mall group, I express my gratitude for creating a supportive and dynamic research environment. I am lucky to have had you as colleagues these four years, and luckier still to be able to call you friends. And so, thank you to Juan, Taishi, Manu, Betti, Elisa, Jana, Jule, both Lauras, Sarah, Tim, Maria, Elias, Bhuvi, Janina, Katharina, Marleen, Mona, Felix, Cedar, Eric, and Brenda.

To Juan and Taishi in particular, I thank you for the tremendous support and guidance. Without you both, I would not be the scientist I am. To all my collaborators and all those who have worked on this project with me, thank you for your contributions and advice. Thank you also to my thesis advisory committee – Judith, Duncan, and Ingrid – for providing invaluable guidance throughout my time as a graduate student.

My gratitude also goes to all the students and postdocs throughout the DKFZ, particularly in our small lunch group, who always managed to lift spirits. To my friends from beyond the lab, especially my athletics training partners, thank you. Thank you to friends who have given their support from afar, even through a pandemic. Thank you in particular to Ran, for being a constant source of humour and support.

I am incredibly grateful to Luisa (and Inu) for her kindness and patience.

Finally, I am especially thankful to my parents and to my sister for their unwavering support, despite being many oceans away. I would not be the person I am today without you.

Heidelberg, May 2023

Contents

Abstract v

Zusammenfassung vii

Contributions ix

Acknowledgements xi

Contents xiii

List of Figures xvii

List of Tables xix

List of Acronyms xxi

1 Introduction 1

1.1 Cell identity and differentiation . . . 1

1.1.1 Embryonic development and germ layer formation . . . 2

1.1.2 Regulation of cell identity . . . 3

1.2 Transcriptional regulation . . . 4

1.2.1 Transcription . . . 4

1.2.2 Chromatin and nucleosome organisation . . . 4

1.2.3 Histone modifications . . . 6

1.2.4 Transcription factors . . . 7

1.3 Cell identity regulation by transcription factors . . . 8

1.3.1 Master regulators . . . 8

1.3.2 Pioneer factors . . . 9

1.3.3 Terminal selectors . . . 10

1.3.4 Cellular reprogramming . . . 11

1.3.5 Transcriptional regulation and disease . . . 13

1.3.6 Safeguard repression . . . 14

1.4 Liver physiology . . . 16

1.4.1 Liver development . . . 16
1.4.2 Liver cancer . . . 17
1.5 Aims of study . . . 20

2 Results **21**
2.1 Screen for safeguard repressors . . . 21
2.1.1 Candidate safeguard repressors in 18 cell types . . . 22
2.1.2 Hepatocyte safeguard repressor candidates . . . 24
2.2 Validation of hepatocyte safeguard repressor candidates . . . 26
2.2.1 *Prox1* overexpression promotes hepatocyte reprogramming . . . 26
2.2.2 Endogenous PROX1 is necessary for hepatocyte reprogramming . . . 29
2.3 PROX1 suppresses multiple non-hepatocyte cell types . . . 32
2.3.1 PROX1 promotes hepatocyte reprogramming . . . 32
2.3.2 PROX1 activity inversely correlates with alternative cell types . . . 34
2.3.3 PROX1 inhibits neuron and myocyte reprogramming . . . 35
2.4 PROX1 target gene repression induces hepatocyte identity . . . 39
2.4.1 PROX1 target gene repression enhances hepatocyte reprogramming . . . 39
2.4.2 Repressor fusion protein mimics full-length PROX1 . . . 41
2.4.3 Repressor fusion protein impairs alternative fate reprogramming . . . 42
2.5 PROX1 represses alternative fate inducers . . . 45
2.5.1 PROX1 decreases chromatin accessibility . . . 45
2.5.2 PROX1 represses transcription factors driving alternative fates . . . 48
2.5.3 PROX1 directly represses fibroblast and adipocyte fate inducers . . . 49
2.6 PROX1 influences liver cancer initiation and development . . . 52
2.6.1 *PROX1* expression levels segregate liver cancer patients . . . 52
2.6.2 PROX1 prevents HCC initiation . . . 53
2.6.3 PROX1 inhibits HCC transdifferentiation . . . 54

3 Discussion **57**
3.1 Safeguard repression . . . 57
3.1.1 Computationally predicting safeguard repressors . . . 58
3.2 PROX1 as a hepatocyte safeguard repressor . . . 59
3.2.1 Validation of PROX1-driven alternative fate repression . . . 59
3.2.2 Cell type specificity of safeguard repressors . . . 60
3.2.3 PROX1 as a hepatocyte tumour suppressor . . . 61
3.3 Conclusion . . . 63
3.3.1 Significance . . . 63
3.3.2 Limitations . . . 64
3.3.3 Future perspectives . . . 65

4 Methods **67**
4.1 *In vitro* experiments . . . 67
4.1.1 Primary cell line derivation . . . 67
4.1.2 Virus production . . . 68
4.1.3 Direct reprogramming from MEFs . . . 68
4.2 Molecular biology . . . 70
4.2.1 SDS-PAGE and Western blot . . . 70
4.2.2 Quantitative real-time polymerase chain reaction . . . 70
4.2.3 Immunocytochemistry . . . 71
4.2.4 Fluorescence microscopy and quantification . . . 72
4.2.5 Enzyme-linked immunosorbent assay . . . 73
4.3 *In Silico* Screen . . . 74
4.3.1 Data sources . . . 74
4.3.2 Data analysis . . . 74
4.4 Genomics . . . 75
4.4.1 Bulk RNA-seq . . . 75
4.4.2 CUT&RUN . . . 75
4.4.3 ATAC-seq . . . 77
4.4.4 Single-cell RNA-seq . . . 78
4.5 *In vivo* experiments . . . 81
4.5.1 Hydrodynamic tail vein injection . . . 81
4.6 Plasmids . . . 82
4.7 Antibodies and enzymes . . . 85
4.8 Reagents . . . 87
4.9 Kits and consumables . . . 90
4.10 Devices and Software . . . 91
4.11 Statistics . . . 93

References **95**

A *In silico* screen supplemental data **115**
A.1 Screen datasets and neuronal screen results . . . 115

B Reprogramming supplemental data **117**
B.1 scRNA-seq . . . 117
B.2 CUT&RUN and ATAC-seq . . . 119

C *In vivo* supplemental data **121**
C.1 HDTVI experiments . . . 121

List of Figures

1.1.1 Differentiation involves sequential epigenetic changes . . . 2
1.1.2 Germ layers are an early point of specialisation in development . . . 3

1.2.1 Levels of transcriptional regulation . . . 5

1.3.1 Groundbreaking cellular reprogramming experiments . . . 11
1.3.2 Unlocked cellular plasticity can lead to cancer . . . 13

1.4.1 Structure and cell types of the hepatic lobule and bile duct . . . 17

2.1.1 Proposed features of safeguard repressors . . . 22
2.1.2 Computational screen predicts safeguard repressor candidates . . . 23
2.1.3 Relative TF motif densities and expression in hepatocytes . . . 25

2.2.1 PROX1 promotes acquisition of hepatocyte morphology and function . . . 26
2.2.2 PROX1 increases hepatocyte marker gene expression in hepatocyte reprogramming . . . 27
2.2.3 *Prox1* expression is highly hepatocyte-specific and the PROX1 binding motif is depleted at hepatocyte marker genes . . . 28
2.2.4 *Prox1* knockdown reduces hepatocyte reprogramming efficiency . . . 28
2.2.5 *Prox1* deletion reduces hepatocyte reprogramming efficiency . . . 30
2.2.6 *Prox1* deletion reduces hepatocyte marker gene expression . . . 31

2.3.1 Single-cell RNA-seq of hepatocyte reprogramming time course reveals dynamics of reprogramming enhancement by PROX1 . . . 32
2.3.2 PROX1 enhances hepatocyte identity and suppresses MEF identity . . . 33
2.3.3 CUT&RUN reveals PROX1 binding locations . . . 35
2.3.4 PROX1 activity correlates inversely with many non-hepatocyte cell types . . . 36
2.3.5 WB analysis shows PROX1 suppresses neuronal and myocyte reprogramming . . . 36
2.3.6 PROX1 reduces neuronal and myocyte reprogramming efficiency . . . 37

2.4.1 PROX1-DBD repressor fusion protein promotes hepatocyte reprogramming . . . 40

2.4.2 Global transcriptional effects of PROX1-DBD repressor fusion protein mimic those of full-length PROX1 . . . 41
2.4.3 PROX1-DBD repressor fusion protein inhibits neuronal and myocyte reprogramming . . . 42
2.4.4 PROX1-DBD repressor fusion protein reduces alternative fate induction in hepatocyte reprogramming . . . 43

2.5.1 PROX1 represses alternative cell identities by chromatin closure . . . 46
2.5.2 PROX1 induces gene regulatory network changes via transcriptional repression 48
2.5.3 PROX1 binds and closes the promoter regions of *Prrx1* and *Pparg* . . . 49
2.5.4 *Prrx1* and *Pparg* are directly repressed by PROX1 and their expression impairs reprogramming . . . 50
2.5.5 *Prrx1* and *Pparg* are key PROX1 target genes . . . 51

2.6.1 *PROX1* expression levels correlate with cancer subtype and survival . . . 52
2.6.2 PROX1 prevents HCC tumour initiation and extends survival . . . 53
2.6.3 PROX1 prevents cell fate plasticity in HCC-like tumours . . . 55

A.1.1 *In silico* screen expression dataset and equations . . . 115
A.1.2 *In silico* screen results for neurons . . . 116

B.1.1 scRNA-seq quality control and marker gene expression during hepatocyte reprogramming time course . . . 117
B.1.2 Marker gene expression in scRNA-seq time course for selected alternate cell identities . . . 118

B.2.1 CUT&RUN statistics . . . 119
B.2.2 ATAC-seq quality control and data visualisation . . . 120

C.1.1 PROX1 influences liver tumour fate in mice . . . 121

List of Tables

4.1.1 Sources of cell lines and their culture media . . . 67
4.1.2 Self-made cell culture media . . . 67
4.1.3 Commercial cell culture media . . . 68
4.1.4 Cell culture reagents . . . 69
4.1.5 Cell culture medium additives . . . 69

4.2.1 SDS-PAGE and Western blot reagents . . . 71
4.2.2 qRT-PCR amplification conditions . . . 71
4.2.3 Oligonucleotides for qPCR . . . 72

4.4.1 CUT&RUN buffer compositions . . . 76
4.4.2 Oligonucleotides for ATAC-seq . . . 77

4.6.1 Plasmids for lentiviral transduction . . . 82
4.6.2 Plasmids for shRNA-mediated gene knockdown . . . 83
4.6.3 Plasmids for hydrodynamic tail vein injection . . . 83
4.6.4 Bacterial culture reagents . . . 84

4.7.1 Primary antibodies . . . 85
4.7.2 Secondary antibodies . . . 86
4.7.3 Molecular biology enzymes . . . 86

4.8.1 Commercial molecular biology reagents . . . 87
4.8.2 General molecular biology buffer compositions . . . 87
4.8.3 Organic buffer components . . . 88
4.8.4 Inorganic buffer components . . . 89

4.9.1 Commercial kits . . . 90
4.9.2 Plasticware . . . 90

4.10.1 Devices . . . 91
4.10.2 Software (R and R libraries) . . . 91
4.10.3 Software (command line) . . . 92

4.10.4 Software (GUI) 92

List of Acronyms

aa	amino acid
Afp	Alpha Fetoprotein
Alb	Albumin
ASCL1	Achaete-Scute homolog 1
ATAC-seq	assay for transposase-accessible chromatin with sequencing
bHLH	basic helix-loop-helix
BHLHA15	Basic Helix-Loop-Helix Family Member A15
bp	base-pair
BSA	bovine serum albumin
CCA	cholangiocarcinoma
CCS	cosmic calf serum
Cdh1	Cadherin 1
c-MYC	Cellular Myelo-Cytomatosis Proto-Oncogene
CREB	cyclic adenosine monophosphate (cAMP) response element-binding factors
CUT&RUN	cleavage under targets and release using nuclease
DAPI	4′,6-diamidino-2-phenylindole
DBD	DNA-binding domain
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E2F8	E2F Transcription Factor 8
EBF2	Early B-cell Factor Transcription Factor 2
ELISA	enzyme-linked immunosorbent assay
ELK3	Erythroblast Transformation Specific Transcription Factor ELK3
EnR	Engrailed Repressor
EP300	E1A-associated Protein p300
ESC	embryonic stem cell
FBS	fetal bovine serum
FGF	fibroblast growth factor
FOXA1	Forkhead Box A1
Foxa3	Forkhead Box A3
Gata4	GATA Binding Protein 4

GFP	green fluorescent protein
GO	gene ontology
GRN	gene regulatory network
GSEA	gene set enrichment analysis
HAT	histone acetyl transferase
HBSS	Hanks' buffered saline solution
HCC	hepatocellular carcinoma
HDAC	histone deacetylase
HDTVI	hydrodynamic tail-vein injection
Hnf1a	Hepatocyte Nuclear Factor 1-Alpha
Hnf4a	Hepatocyte Nuclear Factor 4-Alpha
HNF4α	Hepatocyte Nuclear Factor 1-Alpha
HNF1β	Hepatocyte Nuclear Factor 1-Beta
HNF6	Hepatocyte Nuclear Factor 6
IF	immunofluorescence
IHC	immunohistochemistry
iNs	induced neurons
iPSCs	induced pluripotent stem cells
KLF4	Kruppel-Like Factor 4
KLF15	Kruppel-Like Factor 15
Krt18	Keratin 18
Krt19	Keratin 19
MEFs	mouse embryonic fibroblasts
MEM	minimum essential medium
MITF	Melanocyte Inducing Transcription Factor
Myc	Myelo-Cytomatosis Proto-Oncogene
MYOD1	Myogenic Differentiation 1
MYT1L	Myelin Transcription Factor 1-Like
NEAA	non-essential amino acids
NICD	NOTCH Intracellular Domain
Nkx2.2	NK2 Homeobox 2
NLS	nuclear localisation signal
NuRD	Nucleosome Remodeling and Deacetylase
NRF2F2	Nuclear Receptor Subfamily 2 Group F Member 2
OCT4	Octamer-binding Transcription Factor 4
OLIG2	Oligodendrocyte Transcription Factor 2
ONECUT2	One Cut Homeobox 2
OSR1	Odd-Skipped-Related 1
Pax6	Paired Box Protein 6
PBS	phosphate-buffered saline
Pen-Strep	Penicillin-Streptomycin

PFA	paraformaldehyde
PEI	polyethyleneimine
PO	poly-L-ornithine
PPARα	Peroxisome Proliferator-Activated Receptor α
PPARγ	Peroxisome Proliferator-Activated Receptor γ
PRC	Polycomb Repressive Complex
Prdm5	PR Domain 5
PROX1	Prospero Homeobox Protein 1
PRRX1	Paired Related Homeobox 1
PTF1a	Pancreas Associated Transcription Factor 1a
qRT-PCR	quantitative real time polymerase chain reaction
RNA-Seq	RNA-sequencing
REST	Repressor Element 1-Silencing Transcription Factor
RNA	ribonucleic acid
RPKM	Reads Per Kilobase per Million mapped reads
RT	room temperature
scRNA-seq	single cell RNA-sequencing
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SOX2	Sex Determining Region Y-box 2
SOX10	Sex Determining Region Y-box 10
SREBP1	Sterol Regulatory Element-Binding Protein 1
Tbx3	T-Box Transcription Factor 3
TBX5	T-Box Transcription Factor 5
TF	transcription factor
TGF-β	Transforming Growth Factor Beta
TJP1	Tight Junction Protein 1
Trf	Transferrin
Trp53	Transformation-Related Protein 53
t-SNE	t-distributed Stochastic Neighbour Embedding
TSS	transcriptional start site
Ttr	Transthyretin
TUBB3	Tubulin Beta 3 Class III
UMAP	Uniform Manifold Approximation and Projection
VP64	Viral Protein 64
WB	Western blot
Wnt	Wingless and Int-1
WT	wild-type
ZF	zinc finger
ZNF771	Zinc Finger Protein 771
ZNF778	Zinc Finger Protein 778

Chapter 1

Introduction

This thesis discusses the role of transcriptional repression as a key mechanism in establishing and maintaining cell identity. It focuses on PROX1, a transcription factor that acts as a repressive safeguard of hepatocyte identity through its suppression of alternative cell identities. This chapter introduces general concepts necessary to understand the contents of this thesis, such as cell identity and transcription factor-based regulation of gene expression. It also outlines current knowledge about repressive transcription factors, and presents an overview of liver physiology and cancer. It concludes with a summary of the findings presented later in this thesis.

1.1 Cell identity and differentiation

Multicellular organisms comprise a broad range of cell types, each with diverse phenotypes and different functions. Cellular specialisation, or differentiation, is the process by which unspecialised cells, such as stem cells, alter their state, typically resulting in their conversion to a more specialised cell type with specific functions in the body. Conrad Waddington's concept of the epigenetic landscape is a metaphor for this process (Waddington, 1957). This illustration depicts marbles rolling down a hill and settling in valleys which correspond to specific cell types (Figure 1.1.1).

This process is essential for the proper development and functioning of multicellular organisms. However, nearly every cell in an organism contains essentially the same genetic information. Indeed, a single zygote, through cell division and differentiation, is the source of all complex tissues and systems in multicellular organisms. Cellular specialisation therefore involves cells undergoing changes in gene expression, resulting in the synthesis of different types of proteins that enable cells to perform specific functions (Britten & Davidson, 1969; Mohn & Schübeler, 2009). For example, a stem cell may differentiate into a myocyte, which can contract and produce movement, or into a neuron, which has specific ion channels and synapses that enable transmission of electrical signals.

Differentiation from adult stem cells also plays an important role in tissue repair and cellular turnover in adulthood. Terminal differentiation is a form of differentiation that results in a specialised cell incapable of re-entering the cell cycle in normal circumstances, except in special circumstances such as injury. Such cells serve their specific functions until they die, without undergoing cell division. The establishment and maintenance of cellular identity therefore plays important roles both in development and throughout life.

1.1.1 Embryonic development and germ layer formation

Embryonic development involves several stages of differentiation. A preimplantation zygote undergoes an initial lineage segregation to form an embryo that can implant (De Paepe et al., 2014). An intermediate step in embryogenesis is the formation of the primary embryonic germ layers: the ectoderm, mesoderm, and endoderm (Kiecker et al., 2016). These germ layers eventually give rise to the various tissues and organs in the body. For example, the ectoderm gives rise to the skin and nervous system, the mesoderm gives rise to the muscles and bones, and the endoderm gives rise to the digestive and respiratory systems (Loebel et al., 2003) (Figure 1.1.2).

After the initial formation of the germ layers, cells undergo lineage commitment, whereby they differentiate into progressively more specialised cell types (Loebel et al., 2003). For example, myocytes arise from mesodermal progenitor cells, neurons from ectodermal progenitor cells, and hepatocytes from endodermal progenitors. Once a cell is committed to a particular lineage, it proceeds through further rounds of differentiation to fully mature.

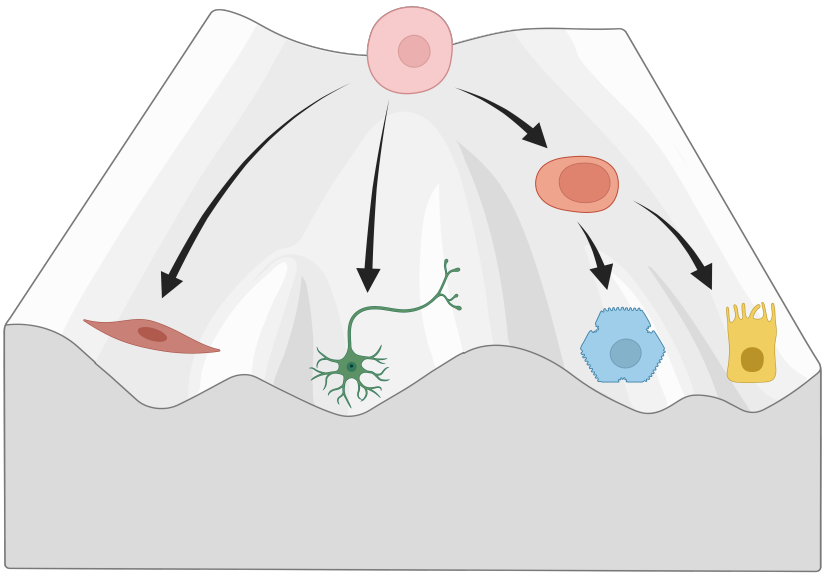

Figure 1.1.1. Differentiation involves sequential epigenetic changes.
Waddington's epigenetic landscape is a metaphor for the changes to the epigenetic state of a cell as it differentiates. The top of the hill depicts pluripotency and more permissive epigenetic states. As cells differentiate, they descend toward more specialised and restricted cellular identities with limited potential for dedifferentiation or transdifferentiation. Modified from Waddington, 1957.

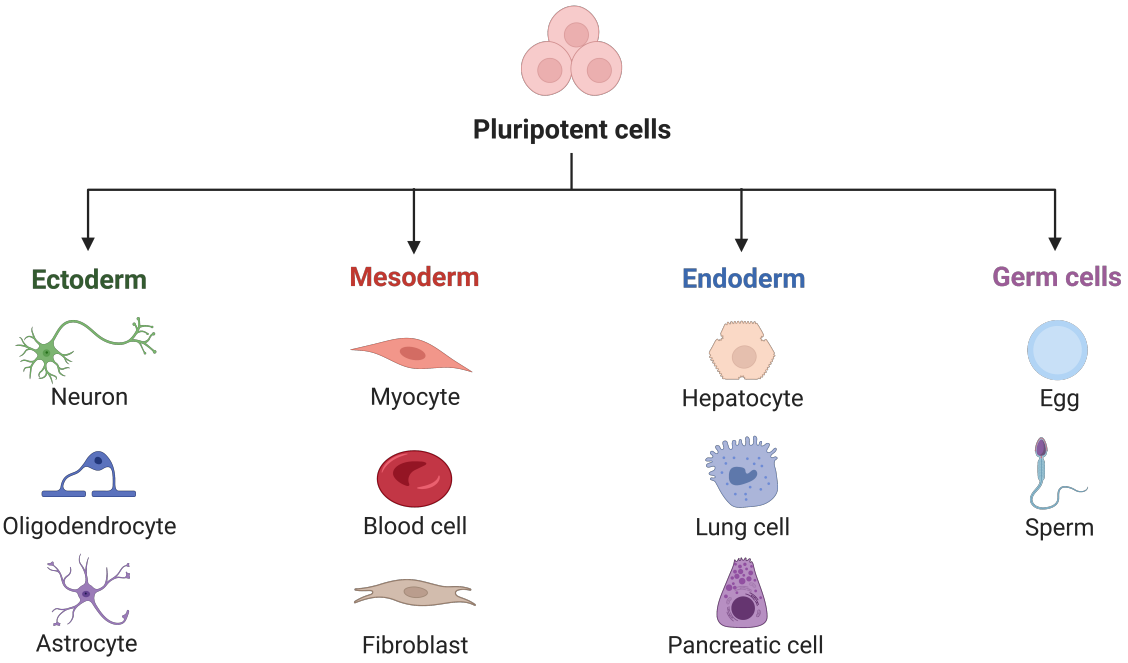

Figure 1.1.2. Germ layers are an early point of specialisation in development.
Differentiation in embryonic development involves the formation of three germ layers, from which all the somatic cells in the body arise, as well as germ cells. Adapted from "Pluripotent Stem Cell Differentiation", BioRender.

1.1.2 Regulation of cell identity

Differentiation is regulated by a complex interplay of genetic and environmental factors. In the early stages of development, cells receive signals from their surroundings that guide them towards a particular fate. Initially, these signals are typically external cues. For instance, signalling molecules such as Wnt, FGF, and TGF-β influence a multitude of cellular characteristics (Loebel et al., 2003). The types and dosage of signals received by each cell is dependent on cellular location. Migration of cells therefore plays a large role in determining the fate of a cell.

Once a cell has committed to a particular path of differentiation, it becomes less responsive to external signals and more dependent on its own intrinsic program of gene expression. This program is controlled by transcription factors, which are proteins that bind to specific deoxyribonucleic acid (DNA) sequences and regulate the expression of target genes (Drouin, 2014).

1.2 Transcriptional regulation

Cell type specification occurs due to differential gene expression or activity between different cell types. This is controlled by mechanisms at every level of the process of transcription, in which the genetic code is converted into ribonucleic acid (RNA). Some RNA undergoes translation, whereby messenger RNA is converted into protein. These mechanisms include structural or chemical modifications, as well as transcriptional, post-transcriptional, translational, and post-translational regulation. Since genes are generally agreed to be any genomic sequence encoding for either RNA or protein (Gerstein et al., 2007), and since proteins are virtually always produced using RNA as an intermediate, all genes must necessarily involve RNA production but not necessarily protein production. As such, transcriptional regulation plays a key role in controlling the expression of genes. This involves control at the level of chromatin and nucleosome organisation, histone modifications, transcription factors, and cis-regulatory elements within the sequence of DNA itself (Figure 1.2.1) (Cramer, 2019). Control of gene expression by transcription factors, and how this influences cell identity, is the primary focus of this thesis.

1.2.1 Transcription

Eukaryotic transcription involves production of RNA from the genomic DNA template by RNA polymerases. In eukaryotes, there are three known DNA-dependent RNA polymerases (Roeder & Rutter, 1969). RNA polymerase I transcribes genes encoding for ribosomal RNA (Paule & White, 2000). RNA polymerase II transcribes messenger and micro RNA (Lee et al., 2004; McCracken et al., 1997). RNA polymerase III transcribes transfer RNA and some small RNA precursors (Dieci et al., 2007; Paule & White, 2000). Transcription begins at the transcriptional start site (TSS). Transcription initiation involves recruitment of RNA polymerase and other components of the pre-initiation complex to the TSS by DNA cis-regulatory elements called promoters (Haberle & Stark, 2018).

While less than 1.5% of the human genome is composed of protein-coding sequences (ENCODE Project Consortium, 2012; Lander et al., 2001), the remainder of eukaryotic genomes is generally rich in regulatory elements. Some of these are gene promoters, and act locally. However, many genes are also regulated by distal cis-regulatory elements called enhancers (Banerji et al., 1981; Bulger & Groudine, 2011; Sagai et al., 2005). The variety of regulatory sequences acting in cis is increased by the fact that many promoters and enhancers are cell type-specific (Lenhard et al., 2012; Lettice et al., 2003; Pennacchio et al., 2007).

1.2.2 Chromatin and nucleosome organisation

Chromatin is the higher-order complex of DNA and proteins that comprise the chromosomes of eukaryotic cells (Kornberg, 1977). This facilitates efficient packaging of DNA, which in hu-

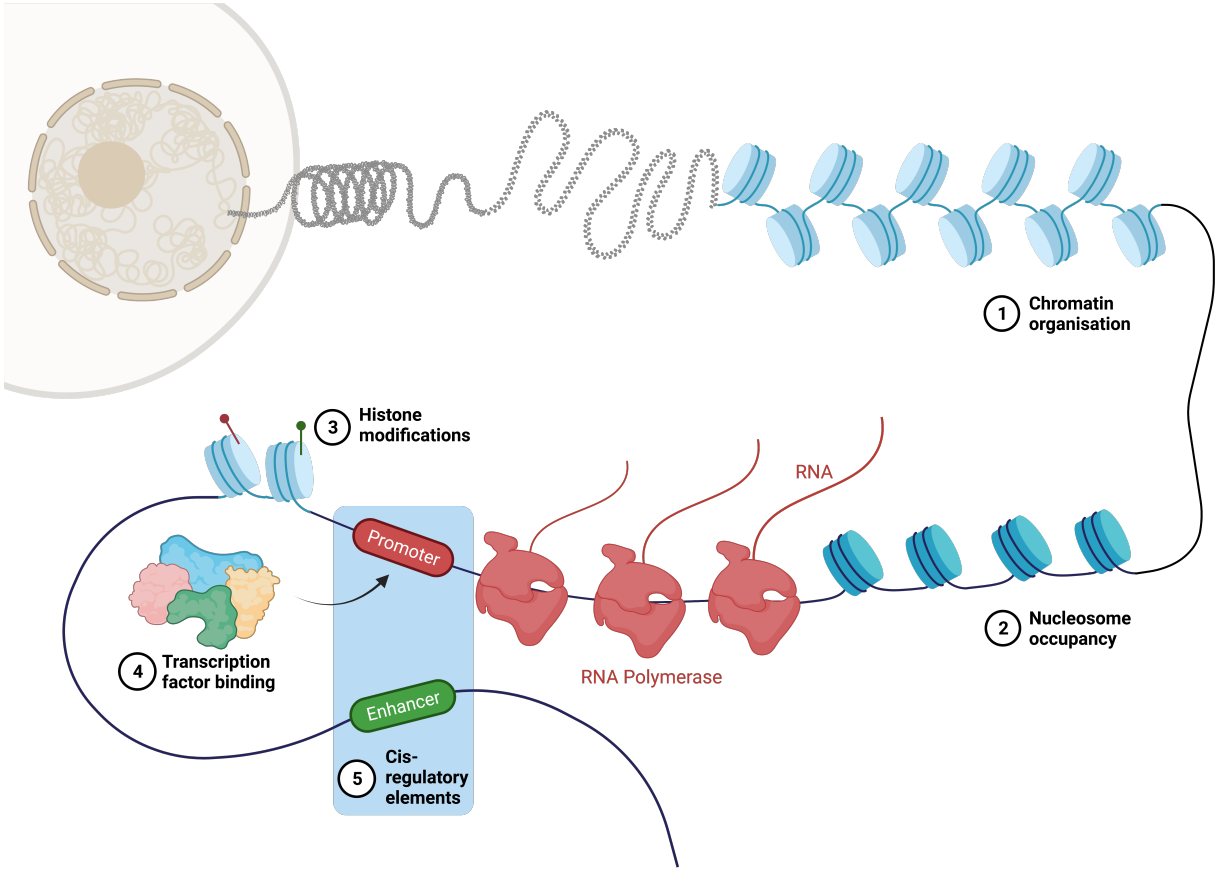

Figure 1.2.1. Levels of transcriptional regulation.
The transcription of genes from DNA is regulated at multiple levels. These include (1) the global organisation of chromatin into compact or accessible conformations, (2) the local occupancy of nucleosomes, (3) modifications to histone tails, or to DNA itself, which can induce larger-scale changes in nucleosome or chromatin accessibility, (4) the presence of transcription factors that can alter histone modifications or the recruitment of transcriptional machinery, and (5) cis-regulatory elements in the DNA sequence itself that influence the binding of transcription factors and other regulatory proteins. Adapted from "Regulation of Transcription in Eukaryotic Cells", BioRender.

mans would be 2 metres long if laid end-on-end, into the nucleus, which usually measures 5-20 micrometres in diameter (de Wit & de Laat, 2012; Lammerding, 2011; Mirny, 2011). The organisation of chromatin influences its activity, and can be seen as a spectrum between a relaxed or compact state (Bannister & Kouzarides, 2011). Relaxed chromatin mostly contains actively-transcribed genes and is referred to as euchromatin. Compact chromatin, which often includes centromeres and telomeres, primarily comprises inactive genes and is usually called heterochromatin.

The compaction of chromatin plays a role in transcriptional regulation because transcription initiation requires RNA polymerase to have access to the promoter region upstream of a gene (Radman-Livaja & Rando, 2010). Chromatin organisation therefore influences the level of access of RNA polymerase and other regulatory proteins to the promoter (Fuda et al., 2009). Chromatin organisation is typically discussed in a broad sense, since it affects longer tracts of chromatin.

The formation of superhelices by supercoiling of chromatin serves as a further high-level regulatory mechanism by which transcription is controlled (Talbert et al., 2019).

Nucleosome occupancy describes the smaller-scale organisation of DNA around nucleosomes, which are the packaging molecules bound by DNA which comprise histone proteins. Nucleosomes comprise 147 base pairs of DNA coiled around eight histone molecules: two each of H2A, H2B, H3, and H4 (where H stands for histone) (Luger et al., 1997). These histone proteins contain a globular core and an unstructured amino-terminal tail, which all contribute to packaging of DNA (Dorigo et al., 2003; Luger & Richmond, 1998; Zheng et al., 2005). A higher density of nucleosomes in specific regions typically leads to decreased transcriptional initiation due again to inhibition of binding of transcriptional machinery (Knezetic & Luse, 1986). Unwinding of chromatin by removal or movement of nucleosomes away from specific regions is a key mechanism in regulating the activation of genes, while the converse is true for the repression of genes (Lorch et al., 1987) As such, nucleosomes generally serve to repress gene transcription in the absence of activating nucleosome or chromatin remodelling complexes (Lorch & Kornberg, 2017).

1.2.3 Histone modifications

The nucleosome occupancy and chromatin organisation observed within a cell is the cumulative result of many smaller molecular changes made to DNA and its scaffold. One major subset of these molecular alterations are modifications to histones. The most commonly-studied modifications are generally the most biochemically tractable ones, and include methylation and acetylation.

For example, acetylation of histones generally leads to formation of actively-transcribable euchromatin (Allfrey et al., 1964; Luger & Richmond, 1998). Acetylation, which correlates with transcriptional activation, is typically performed by histone acetyl transferases (HATs), while removal of acetyl modifications can be performed by histone deacetylases (HDACs) (Kuo & Allis, 1998). Additionally, a broad range of histone methylations have been discovered, primarily on arginine and lysine residues. Methylation may be induced by histone methyltransferases which are present in many chromatin remodelling complexes, which include Polycomb Repressive Complexes (PRCs) (Shao et al., 1999). These have been shown to be extremely important for either gene activation and repression (Barski et al., 2007), and play an important role in development and cell differentiation (Benevolenskaya, 2007; Creyghton et al., 2010).

However, there are at least 80 known modifications that can be covalently added to histone tails on the H3 and H4 subunits alone, involving many different chemical groups and residues (Bannister & Kouzarides, 2011). More global analyses of the histone code have also revealed just how important it is in regulating eukaryotic transcription (ENCODE Project Consortium, 2012; Koch et al., 2007; Rosenfeld et al., 2009).

1.2.4 Transcription factors

Transcription factors are proteins that contain DNA-binding domains which can recognise specific DNA sequences in promoter or enhancer regions, allowing their effector domains to regulate the expression of their target genes. The specificity of transcription factor binding allows them to target specific genes for regulation, which is essential for ensuring that cells can properly control the expression of the genes that they need in order to function properly. The human genome encodes around 1,600 transcription factors. Each cell type may express several hundred of these, which act in concert to control cell fate induction and maintenance (Lambert et al., 2018; Vaquerizas et al., 2009).

Transcription factor effector domains can recruit co-activator or co-repressor proteins, which induce gene expression activation or repression. The mechanisms for this are varied, and include regulation of chromatin and polymerase recruitment (Vos, 2021). Structural changes to chromatin can be induced by histone modifications as discussed in Section 1.2.3. These modifications can either make the DNA more or less accessible to transcription factors and other protein machinery involved in gene expression, which can either activate or repress the expression of the gene (Li, 2002).

1.3 Cell identity regulation by transcription factors

Cell identity induction requires the transcriptional activation of genes specific to the target cell type, as well as the repression of alternative fate genes. Transcription factors are essential for facilitating these events (Mannervik et al., 1999). While many transcription factors are expressed and influence the function of a given cell type, expression of one or a small number of transcription factors has been demonstrated to be sufficient for the induction of specific cell types during development (Mannervik et al., 1999). Such factors are also known as master regulators. Some transcription factors also have more specific features that help them regulate cell identity. For example, some transcription factors may act as pioneer factors with the capacity to remodel closed chromatin, induce a gene expression network, and thus specify cell identity (Zaret, 2020). Alternatively, transcription factors may act as terminal selectors which specify and maintain terminal cell identity (Hobert, 2016).

Transcription factors regulate several target genes and downstream gene expression programs, also known as gene regulatory networks (GRNs). GRNs are complex systems of interactions between genes and the elements that regulate the expression of those genes (Karlebach & Shamir, 2008). The application of computational analysis to large-scale biological data has made the study of gene expression networks more relevant in recent times. Understanding the GRN of a given cell allows for greater knowledge of the important elements regulating gene expression in that cell, since the GRN is what ultimately determines the identity and fate of a cell. Since differentiation is guided by the coordinated expression of specific genes, it is clear that the GRN within a specific cell would be critical in regulating the identity of that cell.

GRN stability is also crucial for the maintenance of cell identity. Identifying transcription factors with key roles in the establishment and maintenance of cell fate-relevant GRNs, and understanding what mechanisms facilitate these roles, can help elucidate how cell identity is regulated.

1.3.1 Master regulators

Master regulators drive the expression of a network of genes that collectively induce the characteristics of a specific cell type (Chan & Kyba, 2013). Such factors are defined as being at the apex of a GRN, making them important for cell identity establishment (Ohno, 1979). Throughout development, many cells will adopt a terminal cell identity. The transcriptional landscape required to facilitate acquisition of that cell identity is often induced by master regulators (Baker, 2001; Hobert, 2016; Mattick et al., 2010). Expression of master regulators is typically induced by other transcription factors which themselves are only temporarily expressed. After activation, master regulators can often sustain their own expression through auto-regulation or positive feedback.

Myogenic Differentiation 1 (MYOD1), a basic helix-loop-helix transcription factor, is perhaps the most famous example of a master regulator. MYOD1 drives the skeletal myocyte differen-

tiation program from mesoderm (Chan & Kyba, 2013; Weintraub et al., 1989). After MYOD1 expression is activated in skeletal muscle, it maintains its own expression (Zingg et al., 1994). This continued expression is important for maintaining the expression of downstream genes necessary for myocyte function (Fong & Tapscott, 2013). The example of MYOD1 demonstrates that a single master regulator can drive the establishment and maintenance of a gene expression program to regulate cell identity in development and throughout life.

1.3.2 Pioneer factors

Some transcription factors may be able to bind genomic regions even when they are located in closed chromatin. Once there, they can recruit additional factors to open the condensed chromatin, permitting other non-pioneer factors to bind to the now-accessible DNA. These factors can then act together to induce expression of previously-repressed genes. Transcription factors with the ability to bind in closed chromatin are described as pioneer factors (Balsalobre & Drouin, 2022). One example is Achaete-Scute homolog 1 (ASCL1), which can activate neuron-specific gene expression to induce direct reprogramming from mouse embryonic fibroblasts (MEFs) to neuron-like induced neurons (iNs) (Chanda et al., 2014; Wapinski et al., 2013). Another is MYOD1 which, as previously mentioned, can cause differentiated non-muscle cells to transdifferentiate to myocytes (Davis et al., 1987). Forkhead Box A1 (FOXA1) has also been described as a pioneer factor, and is a key inducer of endodermal identity. It has been shown that FOXA1 opens chromatin, and that one mechanism for this activity is its ability to displace the linker histone H1 (Iwafuchi-Doi et al., 2016). However, a more recent study has challenged the binary nature of the pioneer factor hypothesis, which proposes that pioneer factors open chromatin for non-pioneer factors. By overexpression of both FOXA1, a pioneer factor, and Hepatocyte Nuclear Factor 1-Alpha (HNF4α), a non-pioneer factor, the authors showed that both actually possessed pioneer activity (Hansen et al., 2022). The main difference between the two was that FOXA1 needed fewer copies of its DNA binding motif to bind to DNA than HNF4α. This suggests that pioneer activity is less binary than previously described, although the concept is still a useful one.

One feature of pioneer factors is that they may also bind promiscuously to many different genomic loci, and thereby activate genes that induce alternative cell identities (Lee et al., 2020). For instance, ASCL1 can bind to and activate transcription of myocyte-related genes in addition to neuronal genes during reprogramming to neurons from fibroblasts (Mall et al., 2017; Treutlein et al., 2016). Similarly, MYOD1 can activate transcription of neuronal genes during myocyte reprogramming (Lee et al., 2020). As such, repression of genes associated with alternative cell identities is likely to be required in addition to gene expression activation in order for cells to properly differentiate, and for cells to remain differentiated.

1.3.3 Terminal selectors

Some transcription factors involved in cell type specification continue to be expressed even in mature, terminally-differentiated cells, acting as "terminal selectors" that help to maintain cell identity (Hobert, 2008, 2016; Hobert & Kratsios, 2019). The concept of terminal selectors was first described by Oliver Hobert in his studies on the development of the nervous system in *Caenorhabditis elegans* (Hobert, 2008). Hobert identified terminal selectors specifically expressed in neuronal subtypes which are required for the proper development and function of those cell types.

Terminal selectors are specifically expressed in a particular cell type, where they help to determine and maintain the identity of that cell type. They do so by regulating the expression of a set of terminal effector genes that are unique to that cell type (Hobert, 2010). In neurons, these terminal effector genes are responsible for generating the unique features of each subtype, such as neurotransmitter production and receptor expression. Removal of these terminal selector genes leads to a failure to properly differentiate into the correct cell type, causing impaired development and function, or can cause loss of subtype identity in differentiated cells.

Terminal selectors act via combinatorial control mechanisms, whereby multiple terminal selectors act in conjunction to regulate the expression of effector genes. An example of combinatorial control is the regulation of identity of six mechanosensory neurons in *C. elegans* by MEC-3 and UNC-86. These two transcription factors form a heterodimer that directly binds to a conserved motif at genes associated with terminal differentiation of mechanosensory neurons (Hobert, 2008). It is by this combinatorial mechanism that the vast variety of cell types and subtypes can emerge and exist from the limited number of distinct transcriptional regulators. The collective set of transcription factors necessary to delineate a specific cell type has been termed a core regulatory complex, which by addition or removal of individual transcriptional regulators can allow for new cell types to evolve (Arendt et al., 2016). Importantly, Hobert also describes a coupling of identity initiation and maintenance. This means that terminal selectors are involved not only in the initial lineage choice that determines a cell's identity, but also the maintenance of that identity throughout the life of the organism (Hobert, 2016). Once a cell has differentiated into a specific cell type, the expression of terminal selectors therefore helps to maintain the identity of that cell and prevent it from undergoing dedifferentiation or transdifferentiation into a different cellular subtype. This is essential for the proper function of the organism, ensuring that each cell maintains its specialised function and preventing diseases associated with cellular plasticity.

Beyond neurons, terminal selectors have been described in the maintenance of terminal identity of multiple other tissues. One example is *Pax5*, a terminal selector in mature B-cells. Conditional deletion of *Pax5* induces expression of T-cell-specific genes and leads to the generation of aggressive lymphomas (Cobaleda et al., 2007).

The concept of terminal selectors is important to this thesis because it lays the foundation for the safeguard repressor hypothesis presented here. While Hobert's terminal selectors are mostly transcriptional activators, since they drive expression of terminal effector genes, I predict that safeguard repressors may exist that perform a similar role but which act instead through transcriptional repression.

1.3.4 Cellular reprogramming

Cellular reprogramming is the process of converting a cell from one identity to another. It may involve dedifferentiation to a more pluripotent cell type or, in the case of direct reprogramming, can also occur without going through a pluripotent intermediate (Figure 1.3.1).

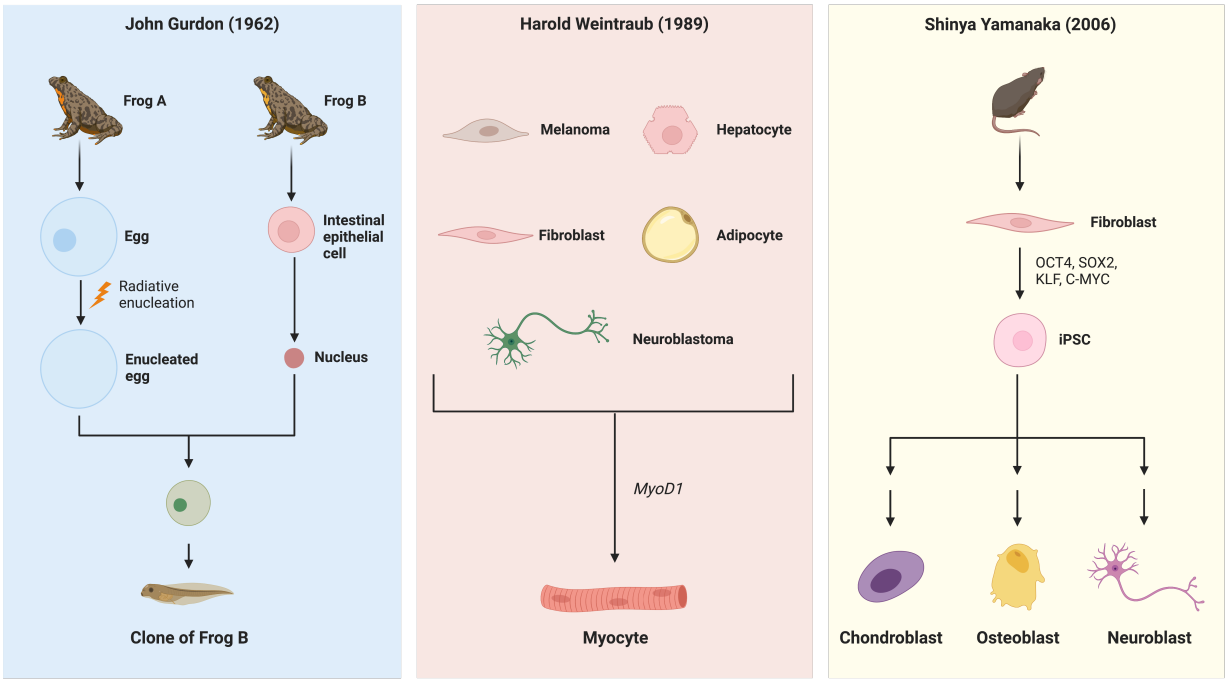

Figure 1.3.1. Groundbreaking cellular reprogramming experiments.
Reprogramming of somatic cells to pluripotency was demonstrated by John Gurdon and Shinya Yamanaka (Gurdon, 1962a; Takahashi & Yamanaka, 2006). Expression of MYOD1 to different somatic cell types can reprogram them into muscle cells (Weintraub et al., 1989). Adapted from "Nobel Prize-Winning Experiments in Physiology or Medicine (2012)", BioRender.

Reprogramming to pluripotency

The reversion of cellular identity, which was previously thought impossible due to the assumed unidirectionality of differentiation, was first demonstrated by John Gurdon (Figure 1.3.1). Gurdon's seminal experiments showed that nuclei from intestinal epithelial cells derived from feeding *Xenopus laevis* tadpoles could be transplanted into an enucleated unfertilised frog egg. The resulting cells could generate fully-formed feeding tadpoles (Gurdon, 1960, 1962a, 1962b). A

follow-up paper four years later showed that seven of these tadpoles had matured into frogs, and some of these were fertile (Gurdon & Uehlinger, 1966). This showed that terminal identities, such as those within the transplanted nuclei, can be reprogrammed such that the resulting cells attained pluripotency and could generate a complete organism.

Shinya Yamanaka later demonstrated that reprogramming can also be achieved with a minimal set of transcription factors (Takahashi & Yamanaka, 2006; Takahashi et al., 2007) (Figure 1.3.1). Based on the assumption that specific factors within the enucleated eggs in Gurdon's experiments drove reprogramming, Yamanaka screened for and identified Octamer-binding Transcription Factor 4 (OCT4), Sex Determining Region Y-box 2 (SOX2), Kruppel-Like Factor 4 (KLF4), and Cellular Myelo-Cytomatosis Proto-Oncogene (c-MYC). A cocktail of these factors was sufficient to reprogram human and mouse fibroblasts into induced pluripotent stem cells (iPSCs) (Takahashi & Yamanaka, 2006; Takahashi et al., 2007). iPSCs mimic embryonic stem cells in gene expression, morphology, and differentiation capacity. These cells can also generate cells from all three germ layers, like embryonic stem cells.

Both Gurdon and Yamanaka demonstrated that factors within the cell, especially transcription factors, could drive reprogramming of terminally-differentiated cells. These data established that terminal differentiation is reversible and manipulable, and that a relatively small number of key factors are vital in determining cell identity (Graf, 2011).

Direct reprogramming

Direct cellular reprogramming, also known as transdifferentiation, is the process by which a differentiated cell is converted into a different cell type without going through an intermediate pluripotent state (Wang et al., 2021). Direct reprogramming involves manipulating the expression of genes within the donor cell to induce changes in its identity and function, allowing for the generation of the target cell type. This technique has limited potential for use in therapeutic applications (Wang et al., 2021), though to a lesser extent than iPSCs. Direct reprogramming also has utility as a tool for studying the mechanisms underlying cell fate stabilisation. It is the latter use which is the primary focus of this thesis.

Direct reprogramming can be achieved by inducing the expression of one to several master regulator transcription factors in a cell, with the aim of activating the expression of genes specific to the target cell type. This technique has been effective in generating various cell types, including cardiomyocytes, neurons, hepatocytes, and pancreatic beta cells, among many others (Wang et al., 2021). The resulting cells have been shown to be functional and in some cases can integrate into existing tissues. One example of successful reprogramming is the use of MYOD1 to induce myocyte reprogramming from MEFs (Davis et al., 1987). Another is the generation of iNs from MEFs by overexpression of ASCL1, which showed that cellular conversions across different germ layers could be induced (Chanda et al., 2014; Wapinski et al., 2013).

1.3.5 Transcriptional regulation and disease

Transcriptional dysregulation is the disruption of normal regulation of gene expression. Recent discoveries that cell identity is manipulable lead naturally to the implication that unintentional dysfunction can also occur, and perhaps lead to disease. A classical example of transcriptional dysregulation in disease is the activation of oncogenes, which promote cell growth and division. Gain of function mutations in these can lead to uncontrolled proliferation of cells and the development of cancer. Similarly, loss of function mutations in tumour suppressor genes, which inhibit cell growth and division, can also contribute to the development of cancer.

Unlocked cellular plasticity

Generalised transcriptional dysregulation can also lead to disease. In most cases, terminal differentiation is associated with a slowing or total stop of proliferation. Unlocking this anti-proliferative characteristic can lead to cellular plasticity that can cause dedifferentiation, transdifferentiation, or blocked differentiation (Hanahan, 2022) (Figure 1.3.2). This potentially explains how terminally-differentiated cells, or cells destined for terminal differentiation, can undergo tumorigenesis.

Dedifferentiation involves the reversal of cell identity from a differentiated state towards a less mature state (Hanahan, 2022) (Figure 1.3.2). Since immature cells retain proliferative and multi-potent characteristics, dedifferentiation can lead to carcinogenesis. One example of this is the loss of Melanocyte Inducing Transcription Factor (MITF), a master regulator of melanocyte identity. The loss of MITF in mature melanocytes causes dedifferentiation towards neural crest progenitor cells which are the developmental precursor to melanocytes (Köhler et al., 2017).

Transdifferentiation is the process by which cells committed to a specific lineage convert to an alternative developmental pathway (Figure 1.3.2). This often involves acquisition of characteristics associated with the alternative fate, but also retention of characteristics of the original identity. The result, which can be transcriptional confusion, or accumulation of traits that are inappropriate for the transdifferentiated cell, can lead to tumorigenesis. In pancreatic acinar cells, a cell of origin for pancreatic ductal adenocarcinoma, both Pancreas Associated Transcription Factor 1a (PTF1a) and Basic Helix-Loop-Helix Family Member A15 (BHLHA15) are responsible for

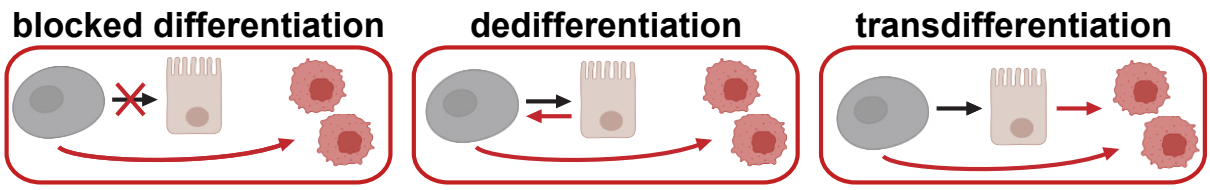

Figure 1.3.2. Unlocked cellular plasticity can lead to cancer.
Cells whose differentiation is blocked may remain undifferentiated and form a cancer stem cell-like population. Alternatively, a fully-differentiated cell might dedifferentiate to a progenitor-like state, or could transdifferentiate into a different cell type with neoplastic traits. Adapted from (Hanahan, 2022) and BioRender.com.

cell fate maintenance (Jiang et al., 2016). Loss of either can leads to transdifferentiation towards cells similar to pancreatic ductal cells, which are more sensitive to oncogenic transformation (Krah et al., 2015; Shi et al., 2013).

Finally, blocked differentiation is the short-circuiting of the normal differentiation process (Figure 1.3.2). This traps progenitors in an immature state, in which they can continue proliferating. Such a situation can also lead to cancer development (Hanahan, 2022). For instance, melanocyte differentiation can be blocked by Sex Determining Region Y-box 10 (SOX10), leading to trapping of neuronal progenitor cells in an immature state that is more likely to lead to melanoma formation (Kaufman et al., 2016).

1.3.6 Safeguard repression

This thesis seeks to answer the question of whether active repression of lineage-inappropriate genes is essential for the induction and maintenance of cell identity. This has not been the focus of much of the existing literature, which describes transcriptional activation as playing an important role. While several transcriptional repressors play crucial roles during development (Schoenherr & Anderson, 1995), many of these are not involved in cell identity maintenance.

One classic example of a cell identity-related repressor is Repressor Element 1-Silencing Transcription Factor (REST), which prevents neuronal identity induction through direct repression of neuronal genes in non-neuronal cells (Schoenherr & Anderson, 1995). However, REST is far from the only example of repression of alternative cell identities. For instance, Paired Box Protein 6 (Pax6) is a core member of the eye determination transcriptional network in *Drosophila melanogaster*. It recruits Polycomb to actively suppress alternative fates, in particular the wing fate, and maintain eye fate (Zhu et al., 2018). In the pancreas, NK2 Homeobox 2 (Nkx2.2) maintains β-cell identity by repression of α-cell identity. Deletion of Nkx2.2 causes β- to α-cell transdifferentiation. However, removal of a key repressed target gene of Nkx2.2 in an Nkx2.2 mutant background causes reverse transdifferentiation back to β-cells (Papizan et al., 2011). These and other similar factors are responsible for maintenance of a specific cellular subtype, with removal of safeguarding leading to loss of that cellular subtype identity.

However, the mechanism by which REST and these other factors act, whereby a single cellular identity is repressed by one transcription factor, is unlikely to be universal. Given the many different cell types that are known, hundreds of repressors would be needed to silence each alternative identities in each cell type. Since fewer than 2,000 transcription factors have been described in humans, it is likely that some factors exist which are capable of repressing a relatively broad range of cell types. Myelin Transcription Factor 1-Like (MYT1L) is one such factor. It is a recently-described cell type-specific "safeguard repressor" that, unlike REST, represses multiple alternative cell identities. MYT1L is almost exclusively expressed in neurons and directly represses non-neuronal genes, enhancing neuronal identity in physiology and in reprogramming (Lee et al., 2020; Mall et al., 2017; Treutlein et al., 2016). Importantly, MYT1l

is expressed throughout life (Cardoso-Moreira et al., 2019). Its depletion in postmitotic primary mouse neurons leads to de-repression of many non-neuronal target genes, and impairs neuronal gene expression and function (Mall et al., 2017). Loss of MYT1L is associated with diseases such as brain cancer and mental disability (Chen et al., 2021; Hu et al., 2013; Melhuish et al., 2018; Weigel et al., 2023). This suggests that safeguard repressor function is required to prevent cell identity loss, and that errors may lead not only to developmental disorders but also to diseases later in life, such as cancer.

Based on the example of MYT1L, I propose that other safeguard repressors also exist. Such factors should exhibit three key characteristics. Firstly, they should be specifically expressed in the safeguarded cell type. Secondly, they should be expressed not only in development but throughout life, since their function is necessary for maintaining the identity of the cell. Thirdly, they should directly transcriptionally repress their target genes. It follows that these target genes would primarily be associated with alternative cell identities, and that their silencing would lead to a stabilisation of the safeguarded cell type.

1.4 Liver physiology

The liver is a vital organ located in the upper right-hand side of the abdominal cavity, just below the diaphragm. It is the largest gland in the body and performs a wide range of important functions, including detoxification of harmful substances, production of bile for digestion, and regulation of glucose, cholesterol, and other metabolites in the blood (Kalra et al., 2023). The liver is made up of several different types of cells arranged into structures called hepatic lobules (Figure 1.4.1). Hepatocytes make up the majority of liver tissue and are responsible for most of its functions. Other cell types found in the liver include Kupffer cells, which are part of the immune system and which help to remove bacteria and other foreign substances from the blood, and cholangiocytes (also known as bile duct cells), which line the bile ducts and help produce and transport bile (Kalra et al., 2023).

The liver receives a dual blood supply. 25% of the total blood supply is oxygenated blood from the hepatic artery, which provides nutrients and oxygen to the liver cells. The remaining 75% is nutrient-rich blood from the portal vein, which carries substances from the intestines and other organs to the liver for processing (Kalra et al., 2023). The liver filters this blood, removing toxins and other harmful substances, before returning it to the rest of the body through the hepatic vein. This filtration is accomplished by the wide-ranging processes that comprise liver metabolism. The wide variety of enzymes in hepatocytes convert substances in the portal vein's blood supply into water-soluble forms that can be excreted in the urine or bile (Kalra et al., 2023). In addition to its role in detoxification, the liver also plays a crucial role in the digestion of food by producing bile, which is stored in the gallbladder and released into the small intestine to help digest fats and absorb fat-soluble vitamins (Kalra et al., 2023). The liver also regulates the levels of glucose in the blood, releasing stored glycogen to maintain normal blood sugar levels between meals and breaking down excess glucose to prevent hyperglycemia. The liver also produces and releases serum proteins such as Albumin, which maintains oncotic pressure in the body and whose secretion can be used as a measure of hepatocyte function (Buyl et al., 2015).

1.4.1 Liver development

The development of the liver involves the differentiation of hepatic (or liver) progenitor cells into hepatocytes and cholangiocytes, the two primary cell types in the liver (Si-Tayeb et al., 2010). The development of the liver begins during the early stages of embryonic development, when the liver bud, a group of cells that will give rise to the liver, forms as part of the foregut endoderm. As the embryo continues to develop, the liver bud grows and undergoes a series of structural changes, ultimately giving rise to the liver (Si-Tayeb et al., 2010).

One key aspect of liver development is the differentiation of hepatic progenitor cells into hepatocytes and cholangiocytes. Hepatic progenitor cells are undifferentiated cells that have the potential to differentiate into either hepatocytes or cholangiocytes, the two primary cell types

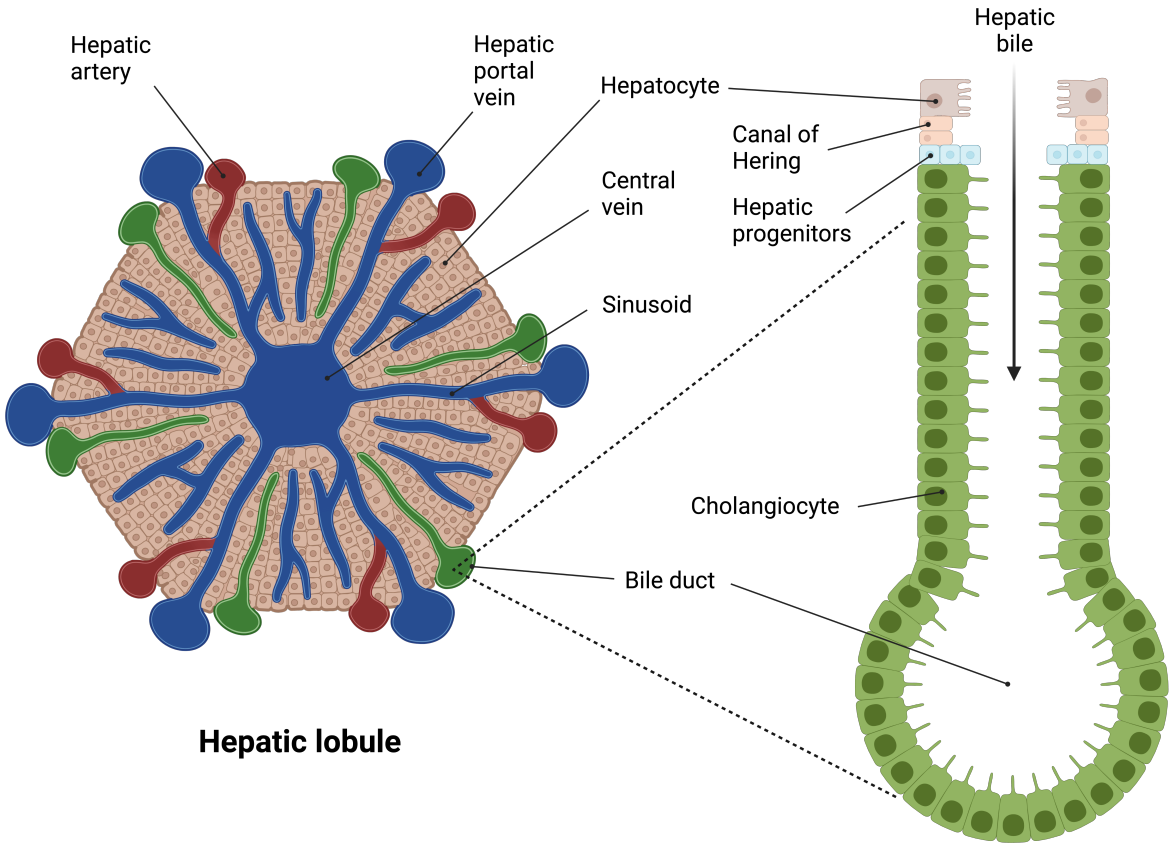

Figure 1.4.1. Structure and cell types of the hepatic lobule and bile duct.
The hepatic lobule is the primary component of liver tissue. It forms a hexagonal structure and contains hepatocytes, bile ducts, and blood vessels arranged radially around a central vein. Blood enters the lobules through capillaries branching from the portal vein, as well as from the hepatic artery. Blood passes through sinusoids and past hepatocytes, which exchange materials with the blood. Bile and processed blood then drain into the central vein. Primary bile produced by hepatocytes passes through bile ductules to the bile duct, during which it is modified by cholangiocytes which line the ducts (shown as inset). Adapted from "Structure of a Hepatic Lobule", BioRender.

in the liver. Hepatocytes are the most abundant cell type in the liver and are responsible for a variety of functions, including the detoxification of harmful substances and the production of bile. Cholangiocytes line the bile ducts and are responsible for the modification of primary bile, which originates from hepatocytes, and the secretion of the modified bile. The differentiation of hepatic progenitor cells into hepatocytes and cholangiocytes is controlled by multiple regulatory genes and signalling pathways including Forkhead box transcription factors and hepatic-specific transcription factors such as HNF4α, Hepatocyte Nuclear Factor 1-Beta (HNF1β), and Hepatocyte Nuclear Factor 6 (HNF6) (Zaret, 2002).

1.4.2 Liver cancer

Primary liver cancer describes any cancer that originates in the liver. The most common forms of primary liver cancer are hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA) (Farazi

& DePinho, 2006; Rizvi & Gores, 2013). HCC is the most common type of primary liver cancer, accounting for approximately 90% of all cases (Llovet et al., 2021). It arises from the hepatocytes, and often develops in people with chronic liver disease, such as cirrhosis or fatty liver disease. Cirrhosis is the strongest HCC risk factor (Galle et al., 2018; Marrero et al., 2018). It is also associated with risk factors such as alcohol abuse, exposure to certain toxins, and viral hepatitis infection (Nkontchou et al., 2013; Perla Oliveira Schulz et al., 2015).

CCA, also known as bile duct cancer, is less common. It is classically described to arise from the bile ducts, which are responsible for carrying bile from the liver to the small intestine (Sirica & Gores, 2014). CCA metastasises earlier and to more distant organs than HCC (Wu et al., 2017). However, it has been shown that it can also derive from hepatocytes, since hepatocytes display a high degree of plasticity (Fan et al., 2012; Mu et al., 2015; Seehawer et al., 2018; Sun et al., 2019). CCA is most commonly observed in patients with chronic inflammatory conditions of the bile ducts, such as primary sclerosing cholangitis, and is also associated with risk factors such as exposure to certain toxins and parasitic infections.

Primary liver cancer is often diagnosed at an advanced stage, as it often does not cause symptoms until it has reached a large size or has metastasised. When symptoms develop, they may include abdominal pain, weight loss, nausea, and jaundice. Treatment options for primary liver cancer depend on the stage and type of cancer, as well as the overall health of the patient (Liu et al., 2015a). In some cases, surgical resection may be possible, but this may not be feasible if the cancer is found at an advanced stage. Other treatment options may include liver transplantation, radiation therapy, and chemotherapy. Further research is still needed to develop better methods for early detection, improved criteria for treatment choice, and more effective treatments.

Plasticity in the liver, especially of hepatocytes, is of importance to multiple diseases, including cancer. For instance, acute liver failure has been shown to be at least partly caused by inappropriate reversion to the foetal liver gene expression program (Hyun et al., 2019). Plasticity of liver progenitor cells, hepatocytes, and cholangiocytes allows all these cell types to contribute to regeneration in the liver (Li et al., 2020). In fact, transdifferentiation between cholangiocytes and hepatocytes, and dedifferentiation from hepatocytes to liver progenitor cells, allows cross-conversion between these three cell types. In some human liver diseases, intermediate hepatobiliary cells have been found. These cells display both hepatocyte and cholangiocyte marker gene expression and possess phenotypic features of both (Li et al., 2020). The mechanisms behind plasticity of liver cells, and how failures can occur and lead to disease, are not well understood. However, some hints include the observation that mature hepatocytes retain a high level of chromatin accessibility, which may allow easier reversion to an immature, proliferative state (Li et al., 2020). This plasticity allows for the impressive regeneration potential of the liver, but can also contribute to tumorigenesis. A similar association of metaplasia with dedifferentiation of mature cells has been shown in intestinal epithelial cells, for example (Radyk et al., 2018). The

role of safeguard repression in maintaining homeostasis and preventing diseases such as cancer in this highly plastic cellular environment will be a major focus of my thesis.

1.5 Aims of study

This thesis aims to generalise the hypothesised concept of safeguard repression as a fate-stabilising mechanism. The three cardinal features of a safeguard repressor are (1) cell type-specific expression, (2) lifelong expression, and (3) increased binding to, and direct repression of, target genes. In this model, a lineage-specific safeguard repressor silences genes from many inappropriate cell identities but not the desired one, to a second cell type. While safeguard repression is exemplified by MYT1L in neurons, it has not been systematically analysed in other cell types. This thesis therefore aims to systematically analyse the existence of safeguard repressor candidates in cell types other than neurons.

To that end, I first present a computational screen designed to identify cell type-specific repressor candidates across 18 cell types. I then test three hepatocyte-specific candidates, identified using the screen, using direct cellular reprogramming from MEFs towards induced hepatocytes. These validation experiments revealed that Prospero Homeobox Protein 1 (PROX1) enhances hepatocyte identity by suppression of alternative cell types, specifically fibroblast and adipocyte identities. I then show that this suppression is due to direct transcriptional repression of master regulator transcription factors that drive the gene expression programs underlying those cell types. Finally, I show that PROX1 prevents hepatocyte fate loss and neoplastic transformation in mice during *in vivo* liver tumorigenesis. Together, these data demonstrate the utility of our safeguard repressor prediction, and provide mechanistic evidence for PROX1 as a novel hepatocyte safeguard. These findings also extend the safeguard repressor hypothesis beyond MYT1L, thus supporting a model in which cell-type-specific safeguard repressors inhibit alternative cell identities to allow development and prevent disease.

Chapter 2

Results

The results presented in this section are part of a manuscript that is being prepared for submission. Portions of the text, figures, and figure legends are adapted from this manuscript. Detailed information regarding contributions can be found in the Contributions section.

2.1 Screen for safeguard repressors

Safeguard repressors are a proposed new class of transcription factor predicted to promote and maintain a specific cell identity by silencing gene expression programs associated with multiple alternative cell types. Such factors would be similar to MYT1L, which promotes neuronal identity in reprogramming and in normal physiology through direct transcriptional repression of non-neuronal cell identities (Mall et al., 2017). The primary hypothesised features of a safeguard repressor are high relative expression in the safeguarded cell type, and enrichment of binding sites in the promoters of alternate cell type marker genes.

To identify safeguard repressor candidates, an *in silico* screen was performed in collaboration with Ignacio Ibarra. This screen was designed to identify transcription factor (TF)s that fulfil the hypothesised features of safeguard repressors. TFs that are specifically expressed in the safeguarded cell type, and whose DNA-binding motifs are enriched in the promoters of non-target cell type marker genes, were considered to be high-confidence safeguard repressor candidates (Figure 2.1.1). Candidate safeguard repressors identified in this way can then be validated experimentally in direct reprogramming experiments to determine if they help promote and maintain the safeguarded cell identity.

18 well-characterised cell types were included in the screen, representing all three germ layers. Within the ectodermal germ layer, I included neurons, astrocytes, epidermal cells, epithelial cells, keratinocytes, and oligodendrocytes. Within the mesodermal germ layer, I included B cells, cardiomyocytes, endothelial cells, fibroblasts, granulocytes, microglia, monocytes, skeletal muscle satellite cells, and T cells. Within the endodermal germ layer, I included bladder cells, pancreatic beta cells, and hepatocytes.

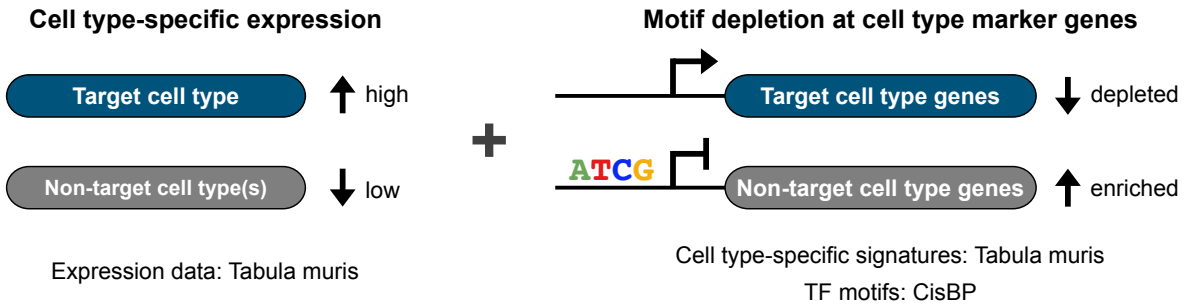

Figure 2.1.1. Proposed features of safeguard repressors.
Transcription factors with high expression in the target cell type, and motif enrichment in promoters of non-target cell type-specific genes, are predicted to be safeguard repressors. Transcription factors fulfilling these criteria have high safeguard repressor scores.

The analysis was based primarily on data from Tabula Muris, a single-cell gene expression atlas of cells derived from multiple tissues of 12 week-old adult mice (Appendix A.1 and Section 4.3.1) (The Tabula Muris Consortium et al., 2018). Using these data, I first calculated cell type-specific expression of all 1,296 TFs detected across the 18 cell types (Section 4.3.2). For this, median expression of each TF in each cell type was calculated, and a Z-score calculated for each TF in each cell type. High-confidence candidates were considered to be those that exhibit high expression (a high Z-score) in one cell type relative to the other 17 cell types.

Next, I generated cell type-specific gene signatures for all 18 cell types (Appendix A.1). I also determined the number of binding motifs of each TF in the promoter region of every gene present in any of these gene signatures (Section 4.3.1). By considering the mean binding motif presence for each TF across each cell type gene signature, I could calculate the average motif enrichment of each TF in promoters of all cell types' marker genes (Section 4.3.2). The DNA binding motifs of high-confidence candidates were hypothesised to be enriched in promoters of genes silenced in the desired cell type, and depleted in promoters of expressed lineage-specific genes.

Finally, the Z-scores for each TF describing expression specificity and DNA binding motif enrichment were normalised to a maximum of 1 and minimum of -1, then summed to calculate a single, combined "safeguard repressor score" for each TF in each cell type (Section 4.3.2). A searchable database of this bioinformatics analysis was designed by Aryan Kamal and is available at `https://apps.embl.de/safeguard`.

2.1.1 Candidate safeguard repressors in 18 cell types

Using this approach, I shortlisted 59 safeguard repressor candidates across the 18 analysed cell types (Figure 2.1.2). I then performed literature searches to classify these 59 candidates based upon previously-described transcriptional activity. Since the loss of safeguard repression is hypothesised to lead to cellular plasticity, which can cause tumorigenesis through dedifferentiation, transdifferentiation, or blocked differentiation (Hanahan, 2022), I also performed a literature

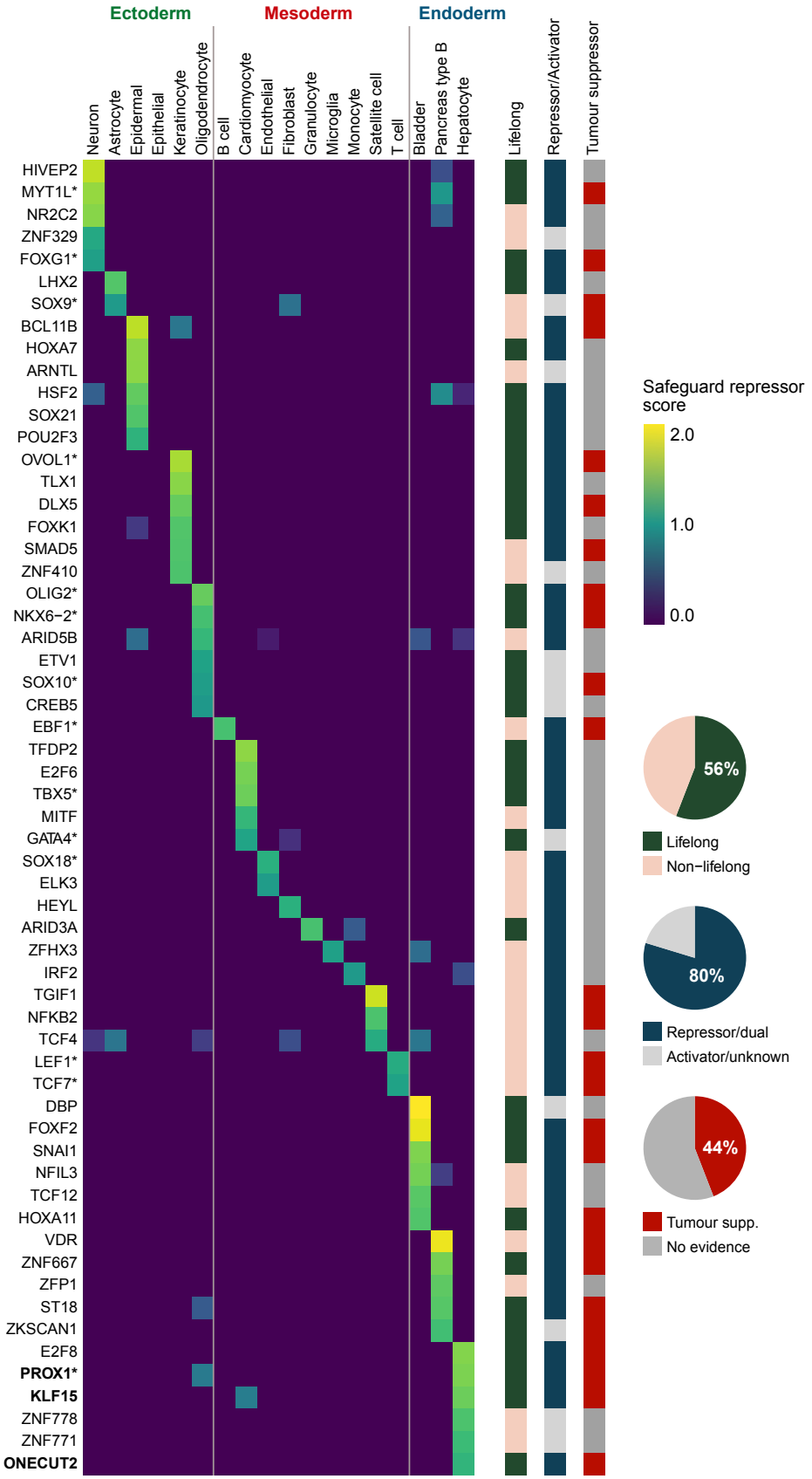

Figure 2.1.2. Computational screen predicts safeguard repressor candidates.
Top six safeguard repressor candidates across eighteen cell types based on safeguard repressor score > 0. Lifelong expression from Tabula Muris Senis (dark green), published repressor/activator activity (dark blue), and published tumour suppressor activity (red) are shown. Bold: tested candidates.

search for tumour suppressor activity of these candidates. 47 of the 59 candidates (80%) have been reported to act either as a repressor or a dual activator/repressor in their respective tissues, based on literature evidence. 26 of the 59 candidates (44%) have been reported to act as a tumour suppressor in their respective tissues, based on literature evidence.

I further classified the candidates based upon lifelong expression using expression data for approximately two year old mice from Tabula Muris Senis (The Tabula Muris Consortium et al., 2020). I considered candidates to be lifelong expressed if their expression levels in aged mice were >50% that of the mean expression levels in 3 month-old mice. By this criterion, 33 of the 59 (56%) candidates were classified as having lifelong expression in the tissue of interest.

Overall, 27 candidates satisfied the initial criteria, allowing us to classify them as lifelong-expressed candidate safeguard repressors in the analysed cell types. Based on further assessment of the literature, 14 of these 27 candidates also promote development of, or reprogramming towards, their respective predicted safeguarded cell identities (Figure 2.1.2). Some notable examples include T-Box Transcription Factor 5 (TBX5) in cardiomyocytes and Oligodendrocyte Transcription Factor 2 (OLIG2) in oligodendrocytes (Waldron et al., 2016; Zhou et al., 2001). This screen also identified MYT1L as a safeguard repressor (Figure 2.1.2), recapitulating previous experimental evidence that MYT1L induces and maintains neuronal identity by actively repressing non-neuronal genes (Lee et al., 2020; Mall et al., 2017).

To extend the safeguard repressor hypothesis, I aimed to identify and test candidate safeguards of a non-neuronal and ideally non-ectodermal cell type. I chose to test hepatocyte safeguard candidates due to the role of cellular plasticity in many liver diseases such as cancer, fibrosis, and fatty liver disease. I also chose to study safeguard repression in hepatocytes since it is an endodermal cell type, and would extend the list of known safeguard repressors to a second germ layer.

2.1.2 Hepatocyte safeguard repressor candidates

The liver is highly specialised yet retains the capacity for massive regeneration upon injury, facilitated by proliferation of hepatocytes. Unintended cell identity loss by hepatocytes plays a crucial role in liver diseases, which can include cirrhosis and fatty liver disease (Dubois et al., 2020). This includes dedifferentiation and transdifferentiation in addition to dysfunctional regulation of the liver's metabolic, detoxification, and systemic homeostasis functions. Loss of hepatocellular differentiation underlie many chronic liver diseases, including hepatocellular carcinoma (Berasain & Avila, 2015).

It has already been shown that specific TFs act to coordinate liver function. These include Kruppel-Like Factor 15 (KLF15), cyclic adenosine monophosphate (cAMP) response element-binding factors (CREB), Sterol Regulatory Element-Binding Protein 1 (SREBP1), and nuclear receptors like Peroxisome Proliferator-Activated Receptor α (PPARα) and Peroxisome Proliferator-Activated Receptor ɣ (PPARɣ), which regulate metabolic functions, and One Cut Homeobox

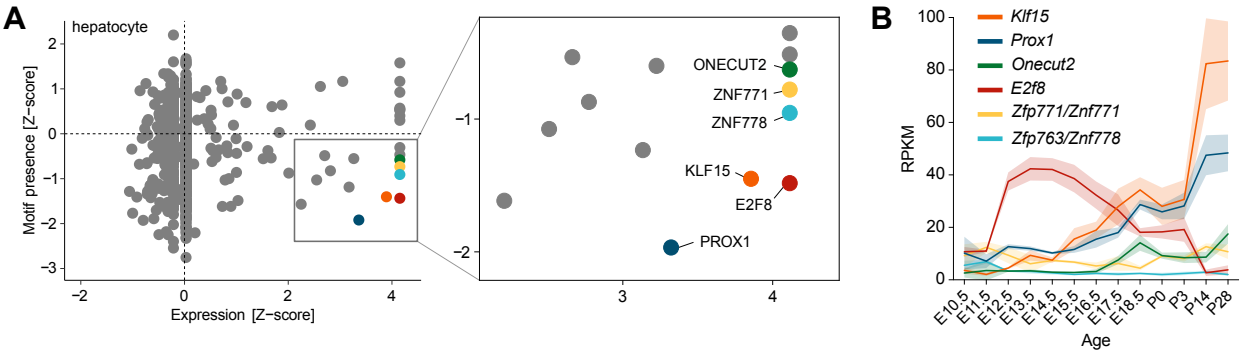

Figure 2.1.3. Relative TF motif densities and expression in hepatocytes.
(A) 1,296 transcription factors displayed based upon their relative expression specificity in hepatocytes (x-axis, displayed as Z-score) and relative motif enrichment in promoters of hepatocyte marker genes relative to non-hepatocyte marker genes (y-axis, displayed as Z-score). Positive x-axis values indicate higher hepatocyte-specific mRNA expression, and negative y-axis values indicate motif depletion in hepatocyte marker gene promoters. Top six hepatocyte safeguard repressor candidates are coloured. Inset shows top six candidates. (B) Mean Reads Per Kilobase per Million mapped reads (RPKM) of top six hepatocyte safeguard repressor candidates across mouse development (Cardoso-Moreira et al., 2019).

2 (ONECUT2) and PROX1, which play a role in hepatoblast migration and maturation (Berasain & Avila, 2015). I therefore decided to examine hepatocyte-specific candidates for their ability to safeguard liver cell identity. The top six hepatocyte safeguard repressor candidates were E2F Transcription Factor 8 (E2F8), PROX1, KLF15, Zinc Finger Protein 778 (ZNF778), Zinc Finger Protein 771 (ZNF771), and ONECUT2 (Figure 2.1.3).

Promisingly, three of these factors are known to play some role in hepatocyte fate maintenance (Berasain & Avila, 2015). Of the top six predicted safeguard repressors, only *Prox1*, *Onecut2*, and *Klf15* are continuously expressed throughout development and aging (Figure 2.1.3) (Cardoso-Moreira et al., 2019; The Tabula Muris Consortium et al., 2020). *E2f8* displayed relatively ubiquitous expression across multiple tissues, while *Znf771*, and *Znf778* did not show lifelong expression in hepatocytes. Given that the latter three factors do not fulfil fundamental assumed criteria for safeguard repressors, I therefore chose to validate only *Prox1*, *Onecut2*, and *Klf15* using direct cellular reprogramming towards hepatocytes.

2.2 Validation of hepatocyte safeguard repressor candidates

To test if any of the candidate hepatocyte safeguard repressors promote hepatocyte identity, I measured the effects of each factor in the context of hepatocyte reprogramming upon overexpression. I began by testing hepatocyte fate acquisition in terms of both mRNA and protein expression of hepatocyte marker genes, as well as morphological and functional changes indicative of altered reprogramming (Section 2.2.1). From these experiments, I found that PROX1 is sufficient to substantially increase hepatocyte reprogramming efficiency. Following this, I conducted knockdown and knockout experiments to show that endogenous *Prox1* expression is necessary for efficient baseline reprogramming (Section 2.2.2).

2.2.1 *Prox1* overexpression promotes hepatocyte reprogramming

Previous work has shown that the overexpression of four liver transcription factors – Forkhead Box A3 (*Foxa3*), GATA Binding Protein 4 (*Gata4*), Hepatocyte Nuclear Factor 1-Alpha (*Hnf1a*), and Hepatocyte Nuclear Factor 4-Alpha (*Hnf4a*) – from a single polycistronic vector (4in1) is

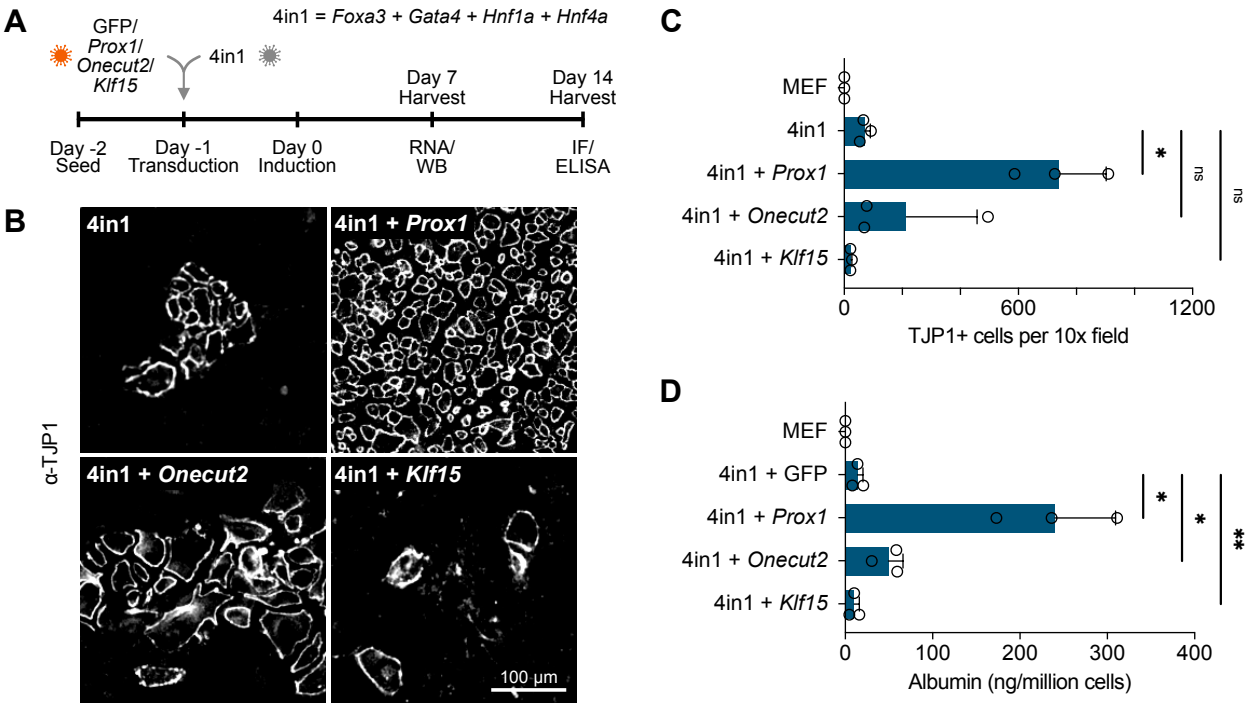

Figure 2.2.1. PROX1 promotes acquisition of hepatocyte morphology and function.
(A) Experimental test of each of three hepatocyte safeguard repressor candidates (or GFP control) by lentiviral overexpression during 4in1-induced hepatocyte reprogramming from MEFs. (B) Representative TJP1 immunofluorescence (IF) images of induced hepatocytes following overexpression of top three hepatocyte repressor candidates or GFP control. (C) Number of TJP1+ cells based on immunofluorescence quantification at day 14 of iHep reprogramming with indicated candidates. (D) Albumin secretion from enzyme-linked immunosorbent assay (ELISA) measurements at day 14 of iHep reprogramming with indicated candidates. Bar graphs show mean values from three biological replicates (indicated as points). Error bars = SD. Dunnett's test, * p-adj < 0.05, ** p-adj < 0.01.

capable of reprogramming MEFs towards functional induced hepatocytes (Song et al., 2016). An essential feature of a safeguard repressor is the promotion of the intended cell identity. As such, I began the *in vitro* test of the computation safeguard repressor screen by assessing the effects of *Prox1*, *Onecut2*, or *Klf15* overexpression during hepatocyte reprogramming (Figure 2.2.1A). The aim was to identify the factor or factors which most strongly increased the efficiency of hepatocyte reprogramming.

Firstly, I investigated cellular morphology and Tight Junction Protein 1 (TJP1) expression by immunofluorescence (IF) (Figure 2.2.1B and C). Cells with small, rounded/hexagonal morphology and complete TJP1-positive membranes were considered to be successfully reprogrammed cells, because this protein marker and morphology mimics that of hepatocytes. Relative to baseline reprogramming, in which only 4in1 was used to induce hepatocyte identity, the overexpression of *Prox1* generated >10-fold more hepatocyte-like cells, a statistically significant difference. By contrast, overexpression of *Onecut2* led to roughly twice as many hepatocyte-like cells on average, but did not generate a consistent and statistically significant improvement over baseline. Finally, overexpression of *Klf15* led instead to a reduction of reprogramming efficiency.

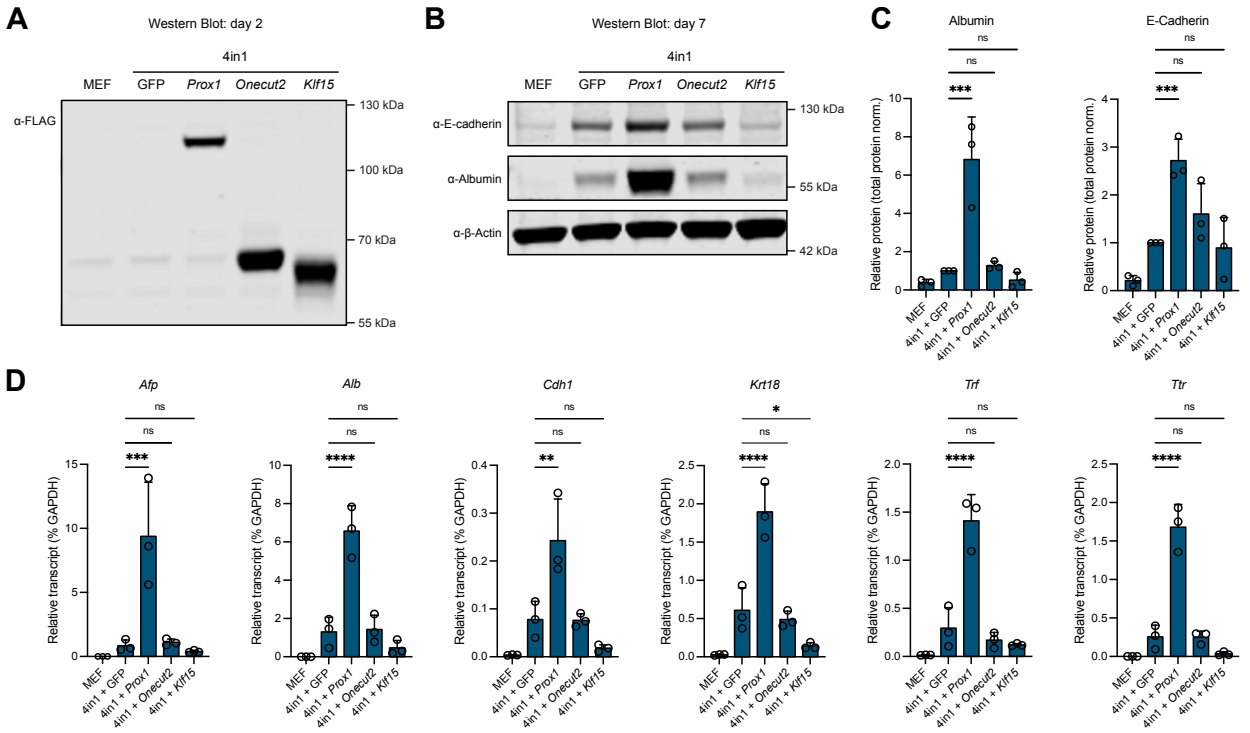

Figure 2.2.2. PROX1 increases hepatocyte marker gene expression in hepatocyte reprogramming. (A) Detection of FLAG-tagged hepatocyte safeguard repressor candidates upon overexpression in MEFs at day 2 of hepatocyte reprogramming using anti-FLAG Western blot. (B) Western blot analysis of Albumin and E-cadherin protein levels at day 7 of hepatocyte reprogramming with indicated hepatocyte safeguard repressor candidates. (C) Quantification of Albumin and E-cadherin protein expression in B, normalised to total protein expression. (D) Gene expression analysis of indicated hepatocyte markers in cells treated as in B using qRT-PCR. Bar graphs show mean values from three biological replicates. Error bars = SD. Dunnett's test, * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.

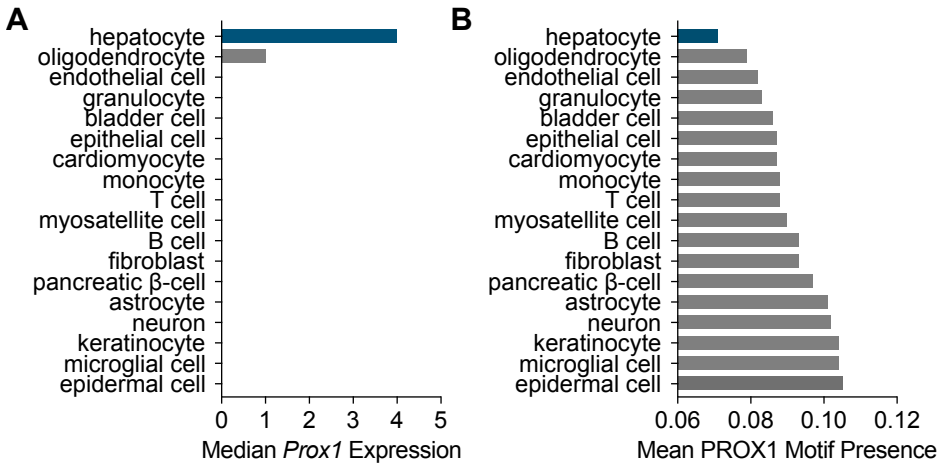

Figure 2.2.3. *Prox1* expression is highly hepatocyte-specific and the PROX1 binding motif is depleted at hepatocyte marker genes.
(A) Median expression levels of *Prox1* in 18 cell types from the Tabula Muris single-cell RNA-seq dataset. (B) Mean motif presence of PROX1 in promoters of marker genes of these 18 cell types.

Secondly, I investigated acquisition of hepatocyte function by measuring Albumin secretion per cell using an enzyme-linked immunosorbent assay (ELISA) (Figure 2.2.1D). Overexpression of *Prox1* led to 17-fold more Albumin secretion per cell relative to baseline, a statistically significant difference. Here, overexpression of *Onecut2* led to approximately double the Albumin secretion relative to baseline, a difference which was statistically significant. Finally, overexpression of *Klf15* led to a reduction in Albumin secretion that was statistically significant.

As a verification of transgene expression, I performed Western blot using anti-FLAG antibody (Figure 2.2.2A). All transgenes were N-terminally FLAG-tagged. Importantly, all candidate factors were expressed at a similar level, so the effects observed could be attributed to specific TF overexpression.

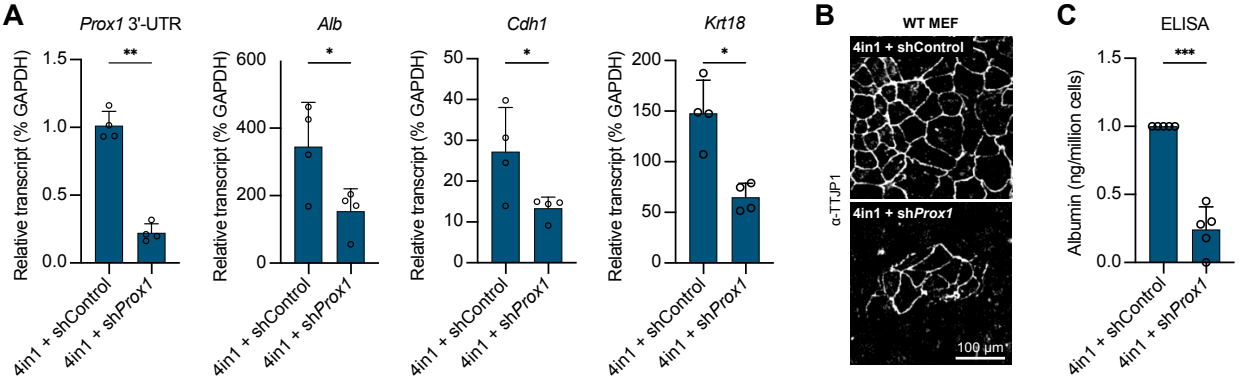

Figure 2.2.4. *Prox1* knockdown reduces hepatocyte reprogramming efficiency.
(A) Expression of *Prox1* and indicated hepatocyte marker genes at day 14 of 4in1-induced hepatocyte reprogramming upon treatment with *Prox1* (sh*Prox1*) or control (shControl) shRNAs based on qRT-PCR. (B) Representative TJP1 IF of cells in A at day 14. Scale bar = 100 μm. (C) Quantification of normalised Albumin secretion of cells in A. Bar graphs show mean values from four (A) or five (C) biological replicates. Error bars = SD. Two-tailed t-test, * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001.

To determine hepatocyte marker gene expression, I measured the protein levels of Albumin and E-cadherin, two hepatocyte-specific genes, by Western blot (Figure 2.2.2B and C). For both these proteins, expression increased significantly and by the greatest amount upon *Prox1* overexpression compared to control. On the other hand, *Onecut2* and *Klf15* overexpression did not induce any significant changes to Albumin or E-cadherin levels. I additionally performed quantitative real time polymerase chain reaction (qRT-PCR) to determine the mRNA expression levels of hepatocyte markers, namely Alpha Fetoprotein (*Afp*), Albumin (*Alb*), Cadherin 1 (*Cdh1*), Keratin 18 (*Krt18*), Transferrin (*Trf*) and Transthyretin (*Ttr*) (Figure 2.2.2D). In all cases, only *Prox1* overexpression led to a significant increase in hepatocyte marker gene expression. *Onecut2* and *Klf15* overexpression either induced no changes, or decreased marker gene expression.

These data all demonstrated a striking promotion of hepatocyte cell fate by PROX1 during reprogramming. This effect was consistent across mRNA, protein, morphological, and functional phenotypes. It was also unique to PROX1, while ONECUT2 and KLF15 showed no effect or sometimes the opposite effect. I also performed a sanity check of the raw screen data, verifying that *Prox1* mRNA expression was highly hepatocyte-specific and that the PROX1 binding motif was depleted in hepatocyte marker gene promoters relative to marker genes of other cell types (Figure 2.2.3). All of this evidence, along with existing literature showing that PROX1 plays a role during the development of the liver (Elsir et al., 2012), encouraged me to proceed with further study of PROX1 as a potential hepatocyte-specific safeguard repressor.

2.2.2 Endogenous PROX1 is necessary for hepatocyte reprogramming

Since exogenous *Prox1* overexpression could promote hepatocyte fate acquisition, I sought to determine whether endogenous Prox1 expression is necessary for efficient hepatocyte reprogramming. I first performed shRNA-mediated *Prox1* knockdown during hepatocyte reprogramming (Figure 2.2.4). Compared to an empty shRNA control, transducing cells with an shRNA pair against *Prox1* led to a significant decrease in *Prox1* transcript levels by qRT-PCR, confirming shRNA-mediated down-regulation. This was accompanied by a significant decrease in hepatocyte marker genes, namely *Alb*, *Cdh1*, and *Krt18*, determined by their transcript levels as measured by qRT-PCR (Figure 2.2.4A). These results show that PROX1 may be necessary for efficient hepatocyte reprogramming.

I next determined reprogramming efficiency using TJP1 immunofluorescence, which showed a significant reduction in liver reprogramming efficiency upon *Prox1* knockdown (Figure 2.2.4B). Finally, Albumin secretion as measured by ELISA was significantly decreased relative to control, demonstrating impaired function of the reprogrammed cells (Figure 2.2.4C).

To substantiate these results, I utilised MEFs from genetically engineered mice in which exon 2 of Prox1 is flanked by loxP sites ($Prox1^{fl/fl}$ MEFs) (Martinez-Corral et al., 2015). Upon exogenous cre expression, cre-lox recombination between the loxP sites induced homozygous *Prox1* deletion (Figure 2.2.5A). Cre-induced conditional *Prox1* deletion caused a significant, 7-fold decrease in

the number of TJP1-positive hepatocyte-like cells compared to Δcre-transduced isogenic controls following 14 days of reprogramming (Figure 2.2.5B to D). *Prox1* deletion also significantly decreased Albumin secretion per cell. Immunofluorescence analysis revealed that the proportion of cells expressing the muscle marker Desmin was significantly greater upon *Prox1* deletion (Figure 2.2.5E). On the other hand, the proportion of cells that were TJP-positive decreased. Gene expression analysis by RNA-seq confirmed that genetic *Prox1* deletion during hepatocyte reprogramming decreased the expression of many liver markers, but led to increased expression of multiple alternative fate markers such as neuronal *Map2* or muscle *Myh9* (Figure 2.2.5F).

This decrease in reprogramming efficiency was also observed in the reduction of hepatocyte marker gene expression as determined by qRT-PCR and Western blot analysis (Figure 2.2.6). Addition of cre led to a significant reduction in *Prox1* mRNA expression relative to Δcre-transduced isogenic controls, confirming deletion. This was accompanied by a significant decrease in *Alb*, *Cdh1*, and *Krt18* transcript levels as measured by qRT-PCR (Figure 2.2.6A). Western blot analysis showed the same pattern, with a decrease in E-cadherin protein levels in cre-treated cells compared to control, accompanied by an increase in Desmin protein levels (Figure 2.2.6B). Overall,

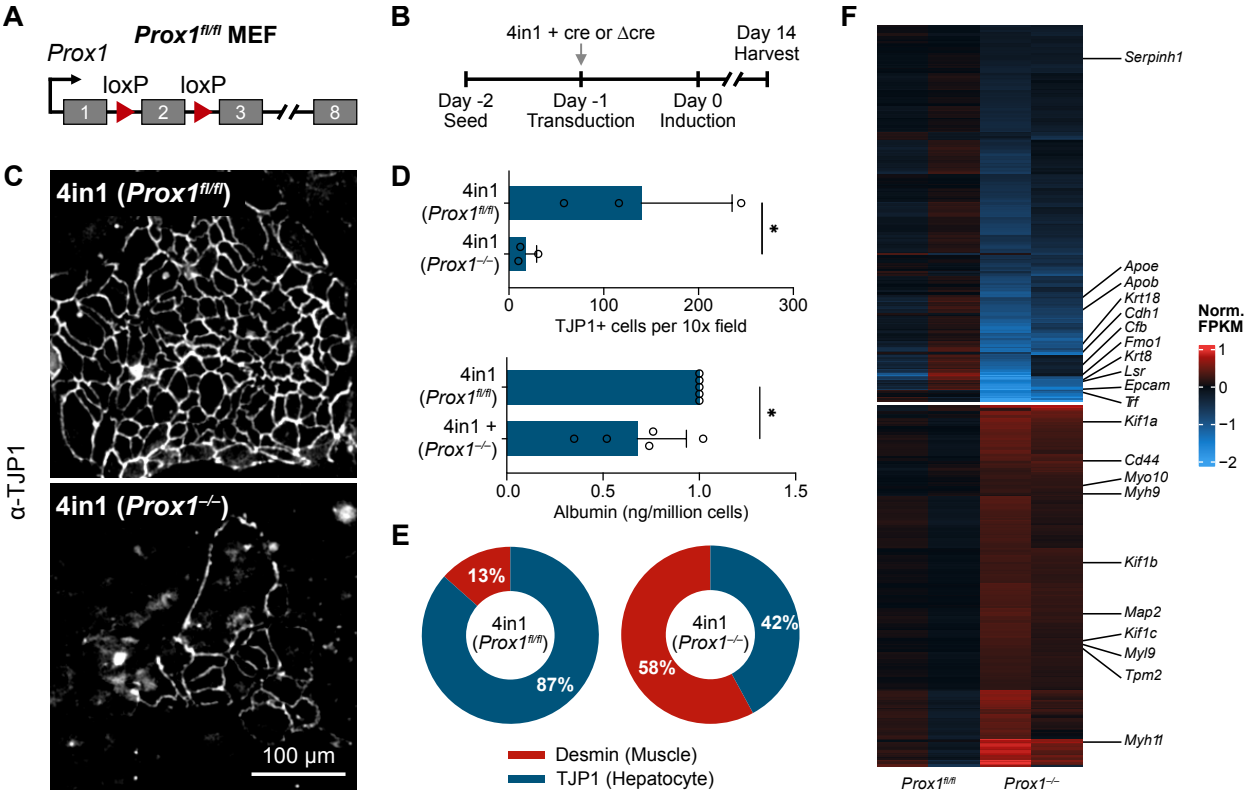

Figure 2.2.5. *Prox1* deletion reduces hepatocyte reprogramming efficiency.
(A) *Prox1*$^{fl/fl}$ MEFs contain *Prox1* with loxP sites flanking exon 2. (B) Schematic of experimental timeline. (C) Immunofluorescence of cells treated as in B show loss of reprogramming efficiency upon *Prox1* deletion, in terms of TJP1-positive cells. (D) Top, quantification of TJP1-positive cells from B; bottom, Albumin secretion by cells from B. (E) Proportion of cells from B that are positive for muscle and hepatocyte markers. (F) Transcriptomics analysis of cells from B shows that *Prox1* deletion leads to up-regulation of non-hepatocyte marker genes and down regulation of hepatocyte marker genes. Bar graphs show mean values from three (IF) or five (ELISA) biological replicates. Error bars = SD. Two-tailed t-test, * p-adj < 0.05.

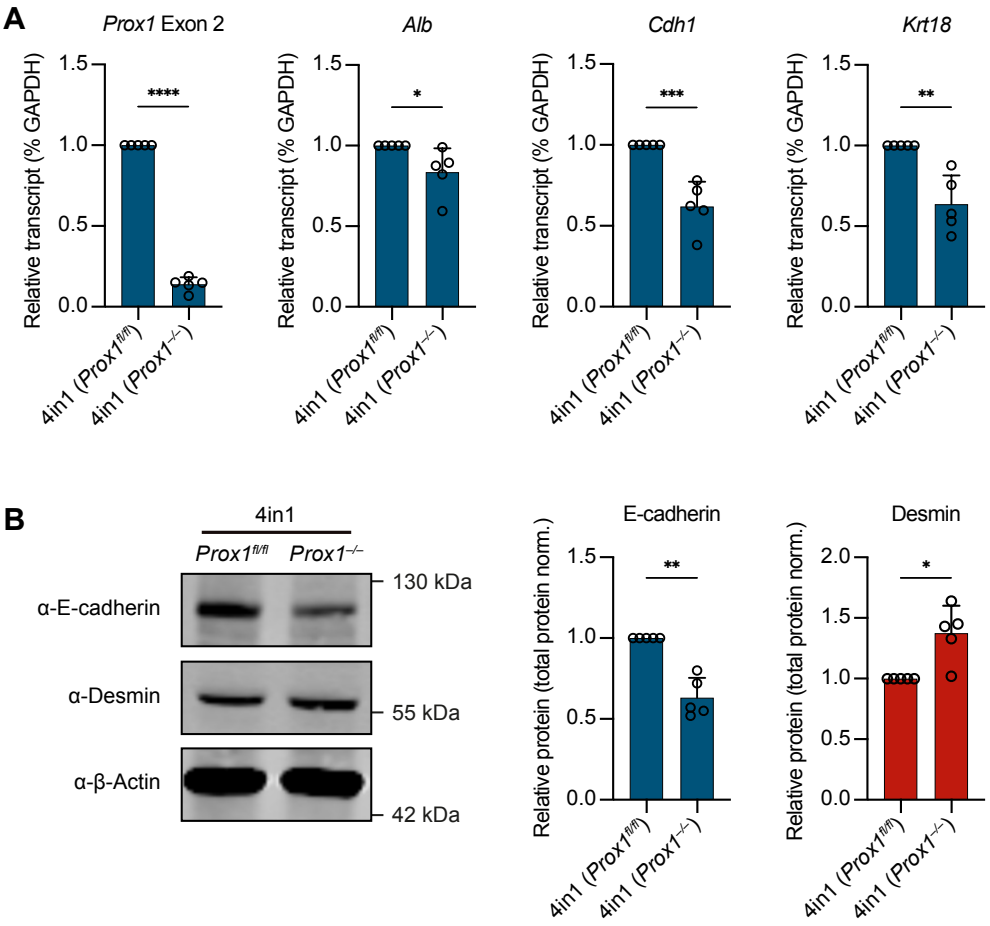

Figure 2.2.6. *Prox1* deletion reduces hepatocyte marker gene expression.
(A) Expression of *Prox1* and indicated hepatocyte marker genes at day 14 day of hepatocyte reprogramming using *Prox1*$^{fl/fl}$ MEFs upon treatment with cre (*Prox1*$^{-/-}$), or Δcre (*Prox1*$^{fl/fl}$) as an isogenic control, based on qRT-PCR. (B) Western blot analysis of E-cadherin (hepatocyte) and Desmin (muscle) protein expression of cells in A, and quantification of this. Bar graphs show mean values from five biological replicates. Error bars = SD. Two-tailed t-test, * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.

these experiments showed that PROX1 is sufficient and necessary for efficient hepatocyte fate induction, likely by repressing alternative fates.

2.3 PROX1 suppresses multiple non-hepatocyte cell types

Here, I investigated whether PROX1 promotes hepatocyte fate acquisition by suppression of alternative cell identities. For this, I performed single cell RNA-sequencing (scRNA-seq) of hepatocyte reprogramming to assess the influence of PROX1 on hepatocyte reprogramming, as well as the gene expression programs repressed and activated by PROX1 (Section 2.3.1). I also investigated PROX1 activity and the gene regulatory networks it influences, with the aim of quantifying the effect of PROX1 on alternative fates (Section 2.3.2). Finally, I found that PROX1 inhibits neuron and myocyte reprogramming (Section 2.3.3).

2.3.1 PROX1 promotes hepatocyte reprogramming

To assess the influence of PROX1 on the dynamics of hepatocyte fate acquisition, I overexpressed *Prox1* in MEFs during hepatocyte reprogramming with 4in1. As a control, I overexpressed green fluorescent protein (GFP). I then analysed the gene expression of the cells using scRNA-seq at

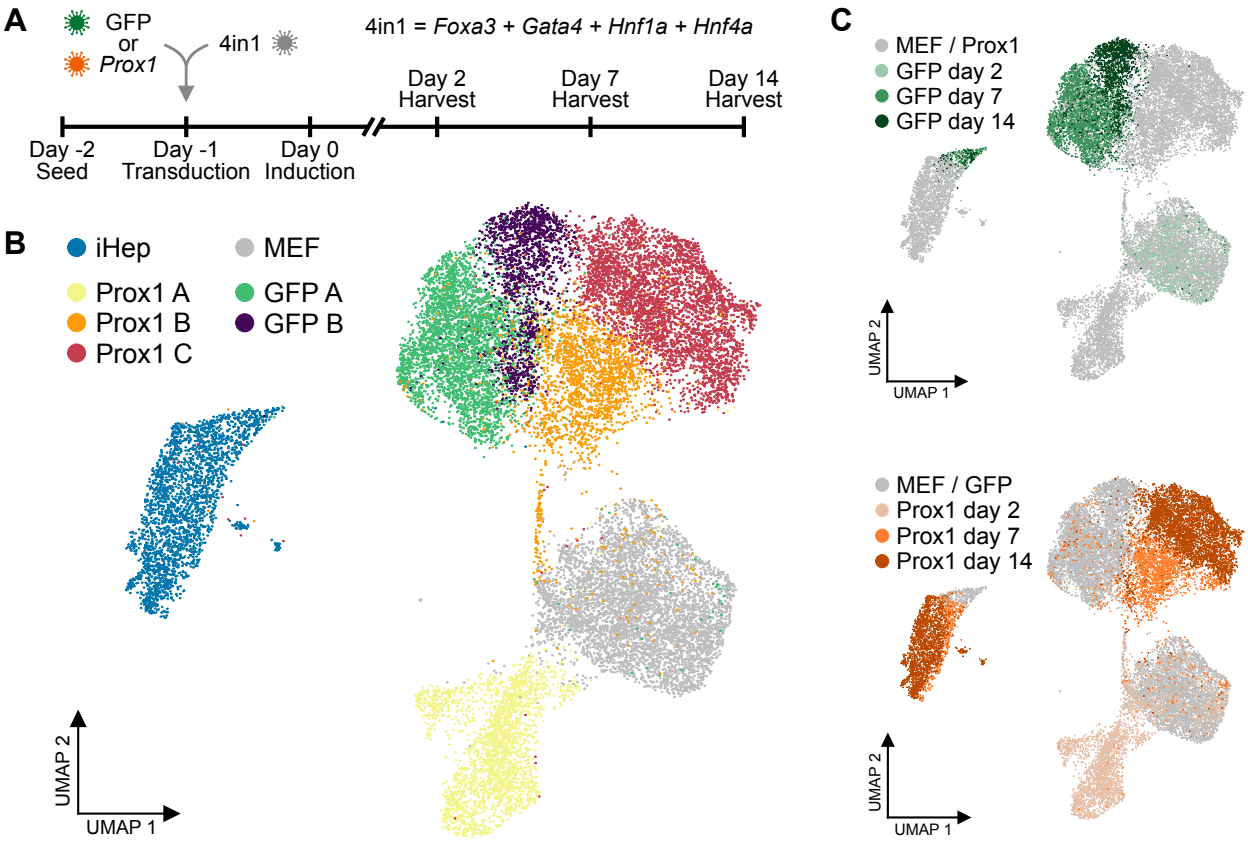

Figure 2.3.1. Single-cell RNA-seq of hepatocyte reprogramming time course reveals dynamics of reprogramming enhancement by PROX1.
(A) Schematic of hepatocyte reprogramming time course with or without *Prox1* overexpression analysed by single cell RNA-seq. (B) Single cell UMAP and clustering of 21,283 cells from two biological replicates as in A. (C) Annotation of experimental treatment and time point. GFP control cells are distinct from cells with *Prox1* overexpression.

three different time points during the reprogramming process – on days 2, 7, and 14. As an additional control, MEFs with no viral transduction, and hence no transgene expression, were also included in the dataset.

After processing of the data and filtering cells based on several quality control metrics (Figure B.1.1 and Section 4.4.4), 21,283 cells from two reprogramming experiments were projected onto two-dimensional space using Uniform Manifold Approximation and Projection (UMAP) and then grouped into seven major clusters by Louvain clustering (Figure 2.3.1A and Section 4.4.4). Cells that were transduced with the *Prox1* overexpression plasmid were clearly distinct from control cells. More specifically, cells with *Prox1* overexpression were generally grouped into four clusters (Prox1 A, Prox1 B, Prox1 C, and iHep), and GFP-transduced cells were grouped into three clusters (GFP A, GFP B, and MEF) (Figure 2.3.1B).

To determine the effect of *Prox1* overexpression on the cells, I analysed the expression of marker genes for different cell types. Cells with *Prox1* overexpression were >7-fold more numerous in the population of successfully reprogrammed hepatocytes (iHep cluster, Figure 2.3.1C) compared to control cells (2,258 Prox1 cells vs 312 control, odds ratio 2.32, p<2.2e-16). *Prox1* overexpressing cells were also present in three clusters (annotated as Prox1 A/B/C). These clus-

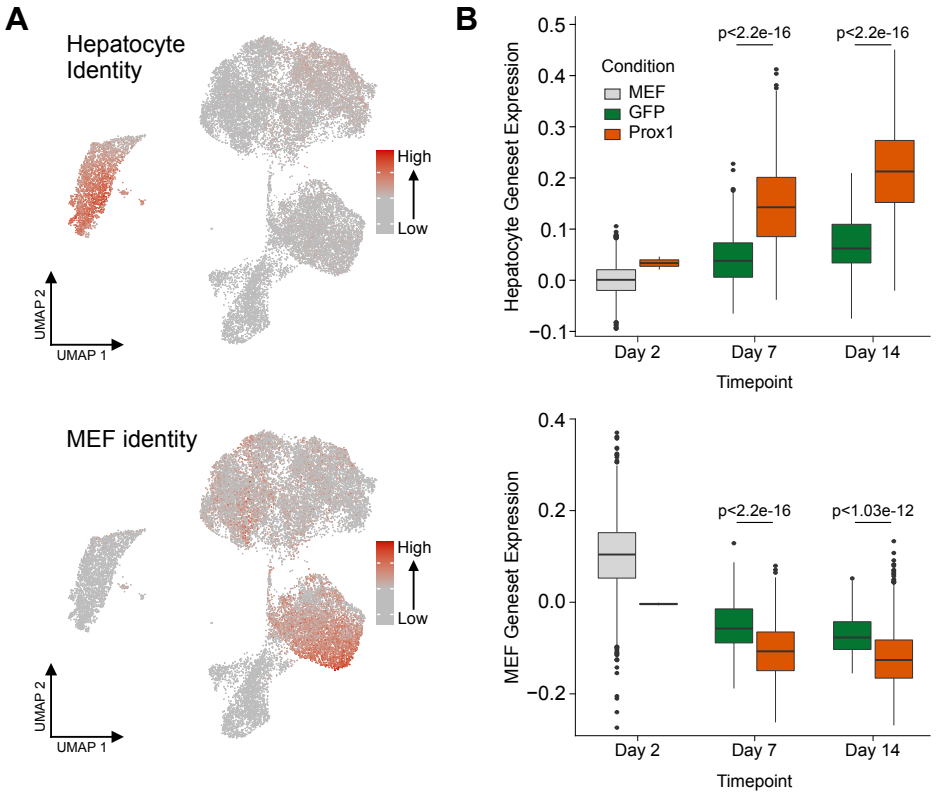

Figure 2.3.2. PROX1 enhances hepatocyte identity and suppresses MEF identity.
(A) Projection of hepatocyte (top) and fibroblast (bottom) identity scores onto all cells. (B) Quantification of hepatocyte (top) and fibroblast (bottom) identity scores, shown as boxplots with p-values (two-tailed t-test) for each timepoint and treatment. All analyses based on 21,283 cells from two biological replicates and following 2, 7, or 14 days of reprogramming as in Figure 2.3.1.

ters were most likely to contain cells at each of the three timepoints which received the *Prox1* virus but no or low levels of 4in1 virus (Figure 2.3.1B).

The MEF cluster strongly expressed MEF genes and contained MEFs without any viral transduction as well as day 2 GFP cells. This cluster therefore reflects the starting cell identity. The other two clusters (GFP A/B) contain mostly cells from the GFP condition, and were annotated as such. The MEF, GFP A, and GFP B clusters therefore represent alternative fate clusters populated with GFP control cells or MEFs.

Importantly, *Prox1* overexpression significantly increased hepatocyte cell identity in single cells at each of the three time points, as measured by the aggregate expression of hepatocyte marker genes (Figure 2.3.2). Cells with *Prox1* overexpression also appeared to down-regulate the donor MEF cell identity more efficiently and to a greater degree across all time points (Figure 2.3.2B), and showed decreased expression of markers for alternative cell identities such as fibroblasts, adipocytes, myocytes, and neurons (Figure B.1.2). Overall, these findings suggest that PROX1 strongly enhances hepatocyte cell fate acquisition and suppresses the gene expression programs of donor and alternative cell types.

2.3.2 PROX1 activity inversely correlates with alternative cell types

I next wanted to determine whether the promotion of hepatocyte identity by PROX1 can be explained by its direct regulation of cell-type-specific gene regulatory networks. To do this, I first needed to identify PROX1 target genes in the context of MEF reprogramming to hepatocytes. For this, I overexpressed FLAG-tagged PROX1 or GFP as a control. Following 2 days of transgene induction, I conducted cleavage under targets and release using nuclease (CUT&RUN) using antibodies against FLAG or IgG as control (Figure 2.3.3). With this technique, I identified 9,574 regions bound by PROX1 which also contained PROX1 DNA-binding motifs. Reassuringly, PROX1 binding was enriched at gene promoters (Figure B.2.1). With these data, it was then possible to identify 48 high-confidence PROX1 targets.

The expression levels of these targets could then be used to infer PROX1 activity in single cells. I then correlated PROX1 activity with cell type-specific gene expression signatures across all single cells (Section 4.4.4). As expected, PROX1 activity showed a significant negative correlation with the fibroblast signature, suggesting that PROX1 can down-regulate the donor cell identity (Figure 2.3.4). I then extended this analysis to other cell types. PROX1 activity appeared to correlate inversely with the expression of markers for the majority of the tested cell identities. The exceptions were the hepatocyte and oligodendrocyte signatures, suggesting that PROX1 represses many identities except for the safeguarded hepatocyte identity and oligodendrocyte fate.

To unravel the effects of 4in1 on the induction or suppression of the different cell identity signatures analysed, I also calculated the activity of 4in1 in single cells. For this, I determined the aggregate activity of the liver inducers FOXA3, GATA4, HNF1A, and HNF4A, based upon curated regulons for each factor (Garcia-Alonso et al., 2019). A key question was whether 4in1

was specifically driving hepatocyte-related genes. Alternatively, I posited that 4in1 might be activating many cell identities promiscuously.

When correlating 4in1 activity with cell type-specific gene expression signatures, I found that 4in1 indeed did not exclusively promote hepatocyte identity. Instead, 4in1 activity also correlated with many alternative cell identities, including astrocytes, epithelial cells, and cholangiocytes (Figure 2.3.4). Notably, these signatures negatively correlated with PROX1 activity. This suggests a role for PROX1 in repressing many non-hepatic gene signatures, countering the promiscuity of gene activation by 4in1 and stabilising hepatocyte identity. Such an activity would reflect a role to that observed for MYT1L, whereby promiscuous gene activation by master regulator transcription factors can be countered by a repressive safeguard that down-regulates alternative cell identities (Lee et al., 2020).

2.3.3 PROX1 inhibits neuron and myocyte reprogramming

The inverse correlations or lack of correlation observed between PROX1 activity and alternative identities imply repression of non-hepatocyte cell types. To validate this inference, I tested the effect of *Prox1* overexpression upon alternative cell types reprogramming. I specifically chose

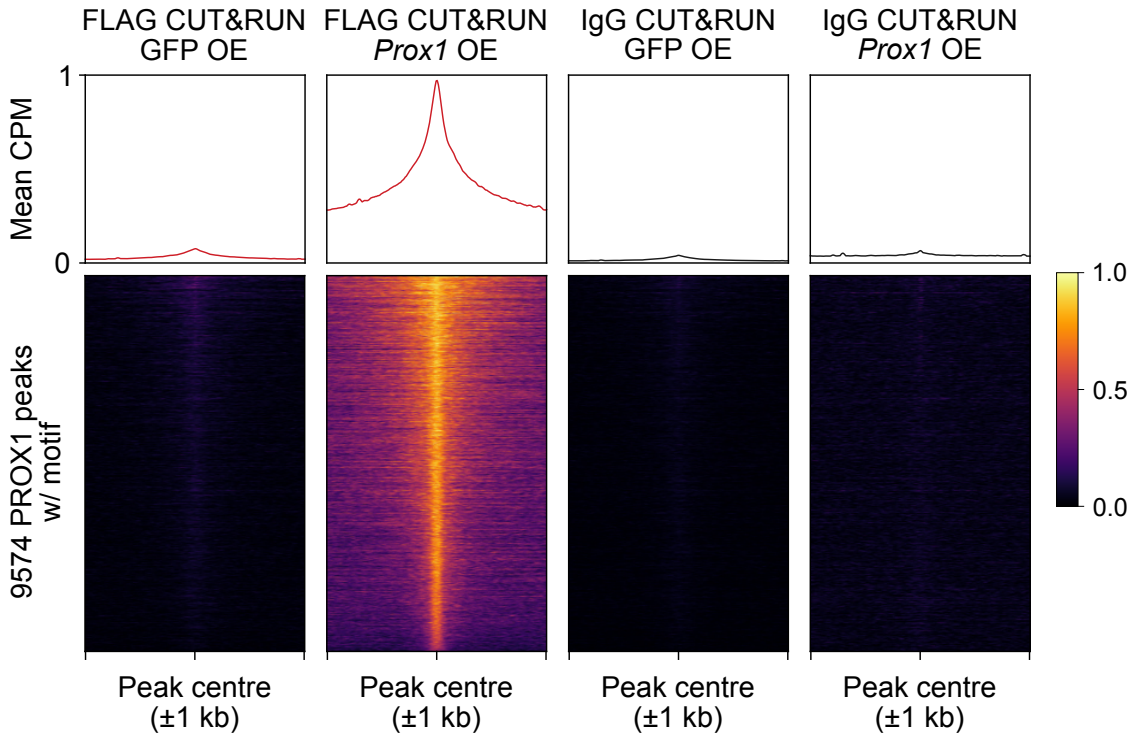

Figure 2.3.3. CUT&RUN reveals PROX1 binding locations.
PROX1 chromatin binding determined by CUT&RUN with FLAG antibody in MEFs following two days of overexpression of FLAG-tagged PROX1 compared to GFP control. IgG controls in MEFs with overexpression of GFP or FLAG-tagged PROX1 are also shown. Line graphs show mean CPM at ± 1kb from the centre of all 9,574 PROX1-bound regions containing a PROX1 binding motif. Heatmaps display actual CPM at each of these 9,574 regions. All data from three biological replicates.

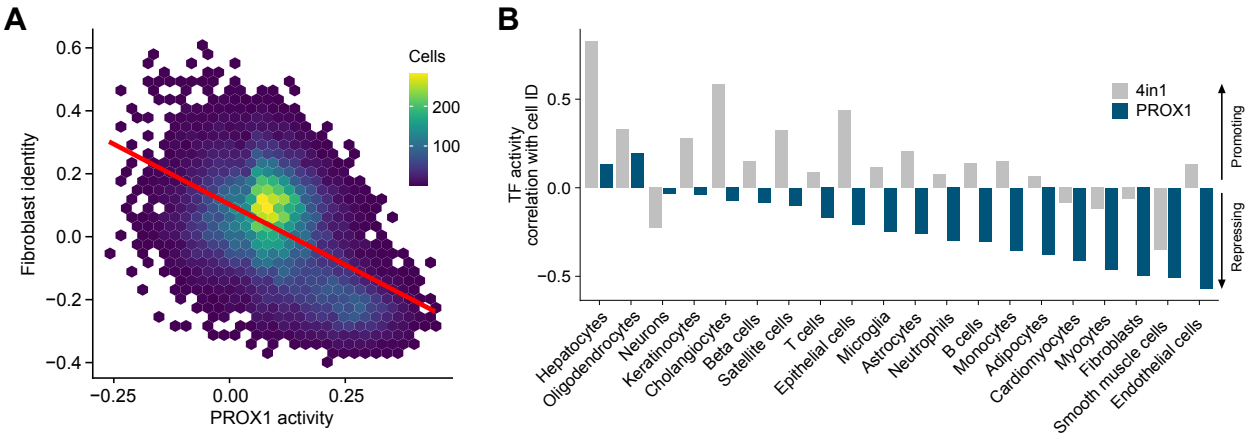

Figure 2.3.4. PROX1 activity correlates inversely with many non-hepatocyte cell types.
(A) Fibroblast identity score correlates inversely with PROX1 activity across all single cells. PROX1 activity was calculated as the extent of the repression of its target genes. Method for calculating TF activity is described in Section 4.4.4 (B) Correlation of 4in1 or PROX1 activity against multiple cell identity scores based on marker genes from Franzén et al., 2019 shows that high PROX1 activity is correlated with decreased cell identity for all cell types except hepatocytes and oligodendrocytes.

to perform this in neuronal and myocyte reprogramming, since both the neuronal and myocyte identities were anti-correlated with PROX1 activity (Figure 2.3.4B). Additionally, neurons derive from the ectoderm, and myocytes from the mesoderm. As such, both cell types also serve as representatives of the two non-endodermal germ layers. *Prox1* is also known to be expressed in some neuronal subtypes, particularly in the hippocampus (Stergiopoulos et al., 2015).

Firstly, I overexpressed *Prox1* or GFP during direct reprogramming from MEFs to iNs using the pro-neuronal transcription factor ASCL1, which can convert fibroblasts into functional neurons (Chanda et al., 2014; Vierbuchen et al., 2010). Unlike direct reprogramming to hepatocytes, which was significantly enhanced by *Prox1* overexpression, neuronal reprogramming was almost entirely abolished. This was true both in terms of protein expression of Tubulin Beta 3 Class III (*TUBB3*), a marker of neuronal identity, as measured by Western blot (Figure 2.3.5),

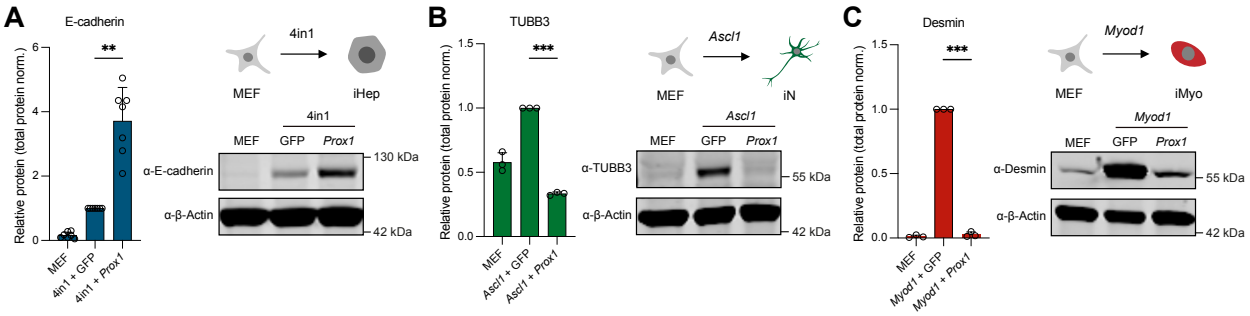

Figure 2.3.5. WB analysis shows PROX1 suppresses neuronal and myocyte reprogramming.
Reprogramming of MEFs to induced hepatocytes (A), neurons (B), or myocytes (C) using 4in1, *Ascl1*, or *Myod1* overexpression, respectively. Protein quantification of E-cadherin (hepatocyte), TUBB3 (neuronal), or Desmin (myocyte) marker proteins at day 14 of respective reprogramming protocols with or without *Prox1* overexpression using Western blot analysis. Bar graphs show mean values from three biological replicates. Error bars = SD. Dunnett's test, ** $p < 0.01$, *** $p < 0.001$.

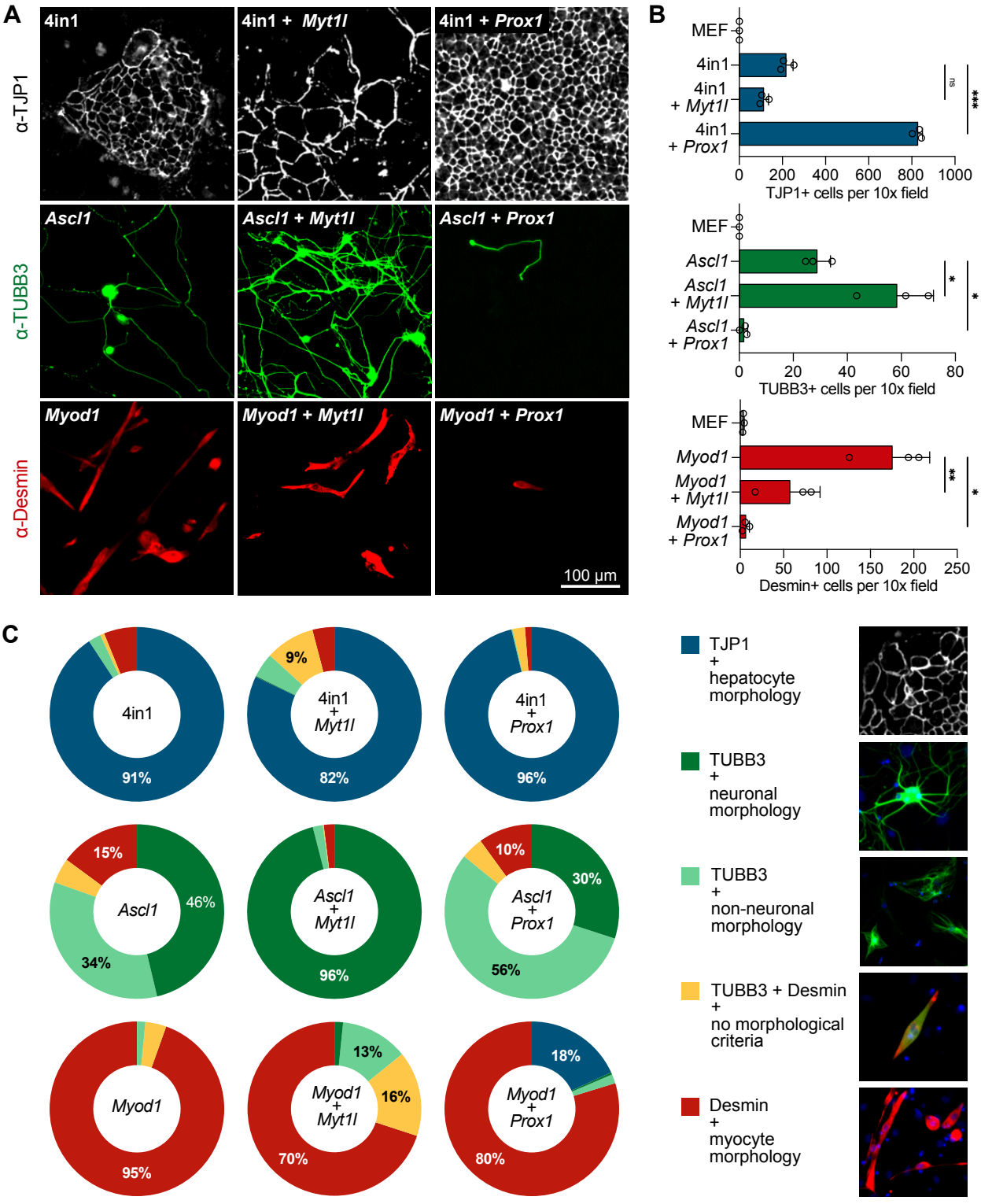

Figure 2.3.6. PROX1 reduces neuronal and myocyte reprogramming efficiency.
(A) Representative immunofluorescence microscopy of cells reprogrammed towards hepatocytes, neurons, or myocytes, with or without *Myt1l* or *Prox1* overexpression in respective reprogramming protocols using indicated antibodies. Scale bar = 100 µm. (B) Quantification of successfully reprogrammed cells in A with *Myt1l* overexpression, based on displayed immunofluorescence markers and morphology. (C) Percentage of reprogrammed cells in A positive for indicated markers and morphological criteria. Bar graphs show mean values from three biological replicates. Error bars = SD. Dunnett's test, * p-adj < 0.05. ** p < 0.01, *** p < 0.001.

and in terms of neuronal morphology acquisition as determined by immunofluorescence against TUBB3 (Figure 2.3.6).

Next, I determined the effects of *Prox1* overexpression in myocyte fate acquisition. Myocyte identity was among the top six signatures repressed by PROX1 (Figure 2.3.4B). I therefore overexpressed *Prox1* or GFP during direct reprogramming from MEFs to myocytes using *Myod1* (Davis et al., 1987; Weintraub et al., 1989). As with neuronal reprogramming, PROX1 almost entirely abolished myocyte reprogramming. Desmin protein expression as measured by Western blot was significantly decreased compared to controls (Figure 2.3.5). Similarly, the number of Desmin positive cells observed by immunofluorescence was significantly lower than in controls (Figure 2.3.6).

Interestingly, while MYOD1 alone did not induce any liver-like cells, immunofluorescence showed that 18% of reprogrammed cells had some level of TJP1 protein expression with hepatocyte-like morphology upon *Prox1* co-overexpression. This finding may reflect the described promiscuity of MYOD1 (Lee et al., 2020), and suggests that safeguard repressors such as PROX1 or MYT1L might be able to redirect this activity to generate the cell identity that is not actively suppressed.

Along with the *Prox1* overexpression experiments towards hepatocytes, neurons, and myocytes, I also overexpressed the neuronal safeguard repressor, *Myt1l*. This resulted in an analogous effect, with *Myt1l* overexpression causing a reduction in efficiency of myocyte and hepatocyte reprogramming, and an increase in efficiency of neuronal reprogramming (Figure 2.3.6). As such, PROX1 and MYT1L both display similar safeguard activity by suppressing alternative cell identities and promoting reprogramming to the safeguarded cell type.

2.4 PROX1 target gene repression induces hepatocyte identity

PROX1 facilitates the acquisition of hepatocyte fate during reprogramming. Additionally, PROX1 activity is correlated with a decrease in non-hepatocyte gene expression. To demonstrate that these effects are caused by direct transcriptional repression by PROX1 of its target genes, I first overexpressed fusion proteins of PROX1 to directly activate or repress its bound genes (Section 2.4.1). I then determined the binding sites of PROX1, as well as its effects on chromatin accessibility, to dissect the effects of PROX1 upon its target genes during the reprogramming process (Section 2.5). With these data, I also identified key PROX1 target genes. These included *Pparg* and *Prrx1*, which encode for the transcription factors PPARγ and Paired Related Homeobox 1 (PRRX1), and which respectively drive adipocyte and fibroblast identity. I then tested their roles by overexpressing or knocking them down during hepatocyte reprogramming (Section 2.5.2).

2.4.1 PROX1 target gene repression enhances hepatocyte reprogramming

PROX1 can act as an activator or repressor, depending on the presence of cell type-specific cofactors (Johnson et al., 2008; Karalay et al., 2011; Lee et al., 2009). I therefore sought to determine whether PROX1 target gene regulation can promote or repress specific cell fates during reprogramming.

For this, I generated fusion proteins in which the DNA-binding domain (DBD) of PROX1 is merged with either a strong transcriptional activator domain (Viral Protein 64 (VP64)) or the Engrailed repressor domain (EnR) (Figure 2.4.1A). In this way, the DBD of each construct provides sequence-specific chromatin binding ability to the fusion protein, where the effector domain can then directly activate or repress PROX1 target genes. As a control, I used the DNA-binding domain alone, with no effector domain. This would identify steric effects as the cause of any phenotypes observed following overexpression of the fusion proteins. Based upon reprogramming efficiency, as determined by TJP1-positive hepatocyte-like cells measured by immunofluorescence, 4in1 with DBD overexpression did not affect hepatocyte reprogramming relative to 4in1 alone (Figure 2.4.1B and C). I also saw no significant difference in Albumin secretion per cell as measured by an ELISA between 4in1 with DBD and control (Figure 2.4.1D). As such, I concluded that the 4in1 + DBD condition could serve as a suitable control since steric hindrance by the DBD did not generate any observable phenotypic changes.

The repressor fusion improved hepatocyte fate induction compared to the DBD, as determined by the number of TJP1-positive hepatocyte-like cells measured by immunofluorescence (Figure 2.4.1B and C). Similarly, there was an increase in the amount of Albumin secretion per cell as measured by an ELISA (Figure 2.4.1D). Conversely, the activator fusion had significantly impaired hepatocyte reprogramming by the same two metrics. Furthermore, the fraction of reprogrammed cells that expressed alternative neuronal and myocyte markers was lower with the

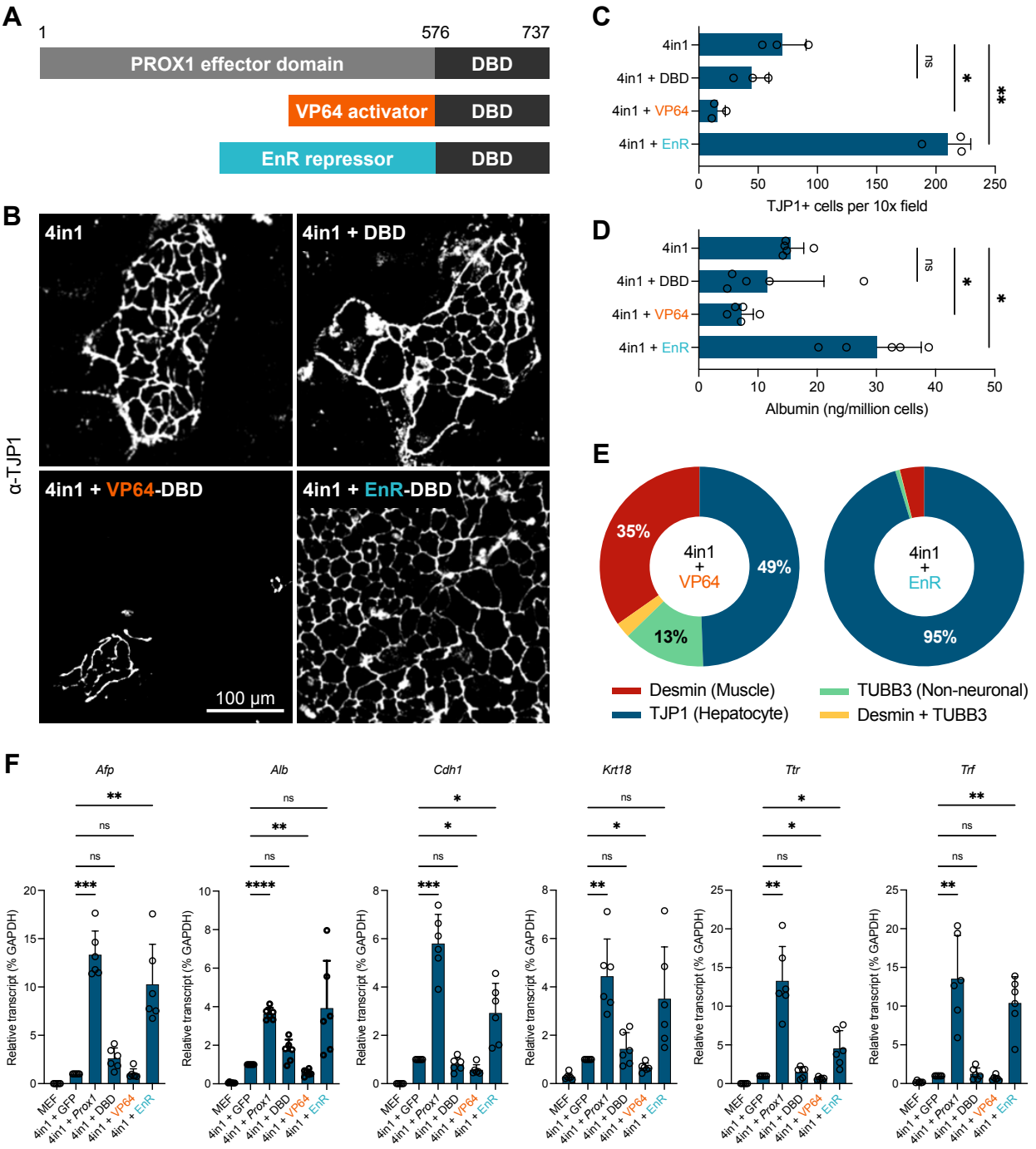

Figure 2.4.1. PROX1-DBD repressor fusion protein promotes hepatocyte reprogramming.
(A) Fusion proteins containing the PROX1-DBD and the VP64 activator or EnR repressor domains (not to scale). (B) Representative TJP1 immunofluorescence upon 4in1-induced hepatocyte reprogramming with indicated PROX1 fusion constructs at day 14. (C) Quantification of TJP1+ induced hepatocytes generated in B (n=3). (D) Normalised Albumin secretion of cells in B (n=5). (E) Proportion of reprogrammed cells in B positive for indicated cell type-specific markers and morphology. (F) Expression of indicated hepatocyte marker genes at day 7 of 4in1-induced hepatocyte reprogramming upon overexpression of indicated PROX1 fusion constructs based on qRT-PCR (n=6). Bar graphs show mean values from indicated replicates. Error bars = SD. Dunnett's test, * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.

repressor fusion, and higher with the activator fusion (Figure 2.4.1E). I additionally performed qRT-PCR to determine the mRNA expression levels of the hepatocyte markers *Afp*, *Alb*, *Cdh1*, *Krt18*, *Trf* and *Ttr* (Figure 2.4.1F). In all cases, only PROX1 and the repressor fusion led to a significant increase in hepatocyte marker gene expression.

2.4.2 Repressor fusion protein mimics full-length PROX1

Finally, I performed RNA-seq analysis at day 7 of hepatocyte reprogramming with PROX1 fusion proteins or controls. With this, I verified that the repressor fusion triggered a similar response in global gene expression as full-length PROX1. I found a strong induction of hepatocyte genes, such as *Hnf4a* and *Krt18*, in the repressor fusion condition. Alongside this, I observed repression of non-hepatocyte genes such as the cholangiocyte marker Keratin 19 (*Krt19*) and the fibroblast transcription factor *Prrx1* (Figure 2.4.2A). The activator fusion instead had the opposite effect, decreasing hepatocyte marker gene expression and increasing alternative fate gene expression.

I then performed a gene set overlap analysis and found a significant overlap between genes down-regulated by PROX1 and genes down-regulated by the repressor fusion. I also found that genes normally down-regulated by PROX1 were up-regulated by the activator fusion, while those genes up-regulated by PROX1 were down-regulated by the activator fusion (Figure 2.4.2B). These experiments altogether demonstrate that PROX1 promotes hepatocyte identity by directly binding to and actively repressing genes that mark or drive alternative cell types.

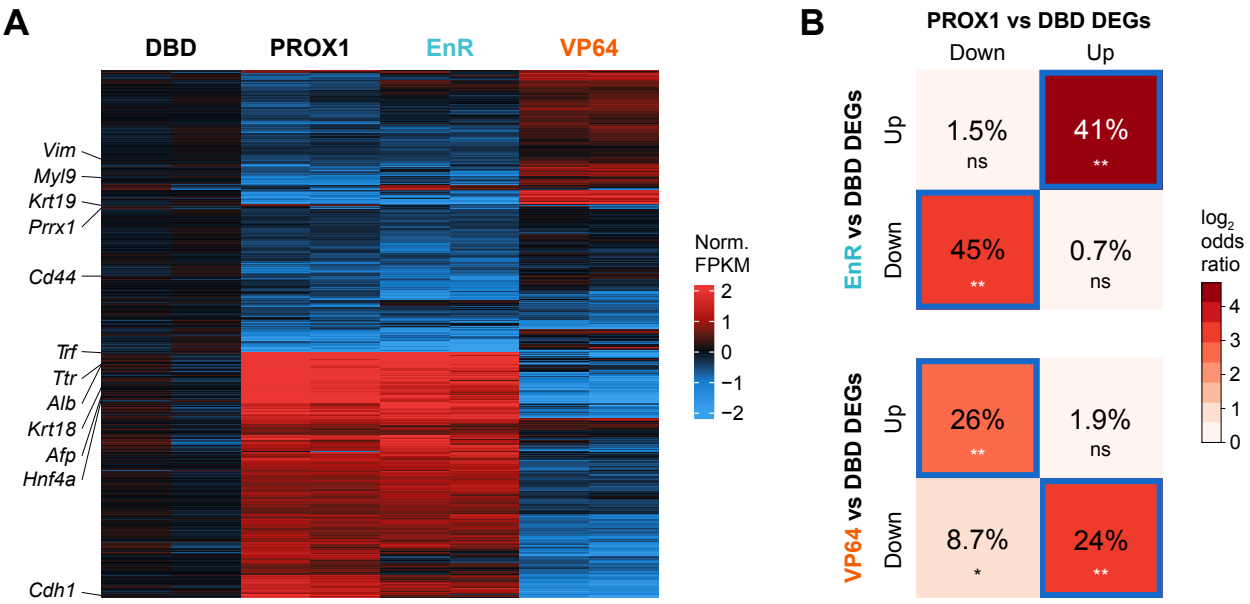

Figure 2.4.2. Global transcriptional effects of PROX1-DBD repressor fusion protein mimic those of full-length PROX1.
(A) Normalised counts of differentially-expressed genes from RNA-seq of induced hepatocytes with over-expression of PROX1-fusion proteins, relative to DBD as a control, at day 7 (n=2). (B) Percent overlap of up- and down-regulated genes in A. Fisher test, * p < 1e-03, ** p < 1e-06.

2.4.3 Repressor fusion protein impairs alternative fate reprogramming

Since *Prox1* overexpression inhibits neuronal and myocyte reprogramming, I hypothesised that overexpression of the repressor fusion protein would also have a similar effect. To test this, I also overexpressed the activator and repressor fusions in the contexts of neuronal or myocyte reprogramming using *Ascl1* or *Myod1* overexpression, respectively.

As expected, the repressor fusion reduced the efficiency of neuronal and myocyte reprogramming, as determined by Western blot analysis of protein expression of TUBB3 in iNs and MYH or Desmin in induced myocytes (Figure 2.4.3). Unexpectedly, the DBD alone led to a decrease in reprogramming efficiency towards both neurons and myocytes. In addition, the activator fusion had the interesting effect of promoting reprogramming towards myocytes.

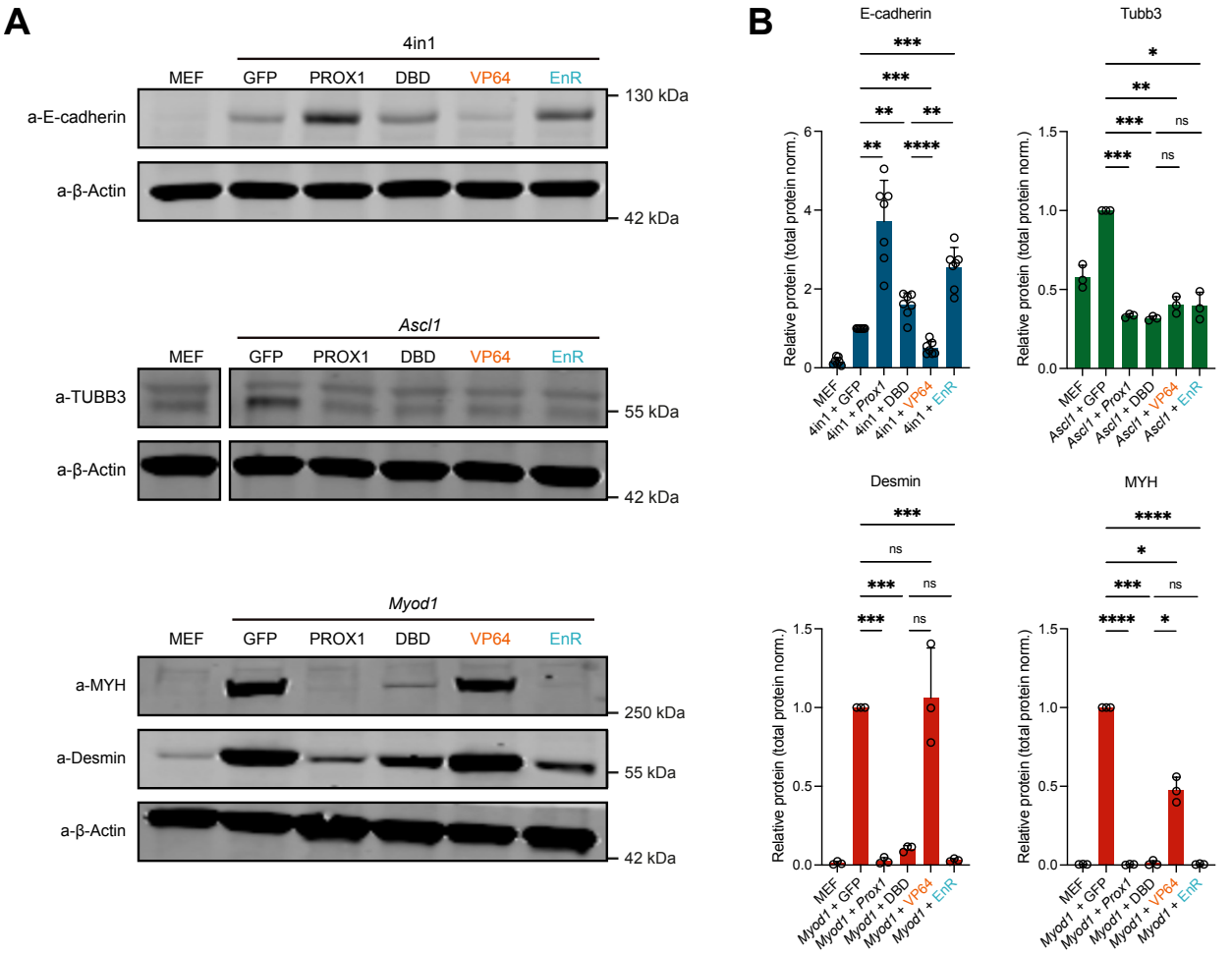

Figure 2.4.3. PROX1-DBD repressor fusion protein promotes hepatocyte reprogramming and inhibits neuronal and myocyte reprogramming.
(A) Representative Western blots of E-cadherin (hepatocyte, n=7), TUBB3 (neuronal, n=3), or MYH and Desmin (muscle, n=3) in induced hepatocytes, neurons, or myocytes generated using 4in1, *Ascl1*, or *MyoD1* overexpression, respectively. All samples gathered after 7 days of reprogramming. PROX1 target gene activation was induced by fusion of PROX1 DNA-binding domain (DBD) to the VP64 activator, while silencing was induced by EnR repressor fusion. Full-length PROX1 and DBD as controls. (B) Protein quantification of Western blots in A. Bar graphs show mean values from indicated number of biological replicates. Error bars = SD. Dunnett's test, * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.

Immunofluorescence analysis of protein expression and cellular morphology with the same conditions produced similar results (Figure 2.4.4). As in the Western blot analysis, immunofluorescence analysis showed that the repressor fusion and DBD alone led to a decrease in reprogramming efficiency towards both neurons and myocytes (Figure 2.4.4A and B). Despite this, the

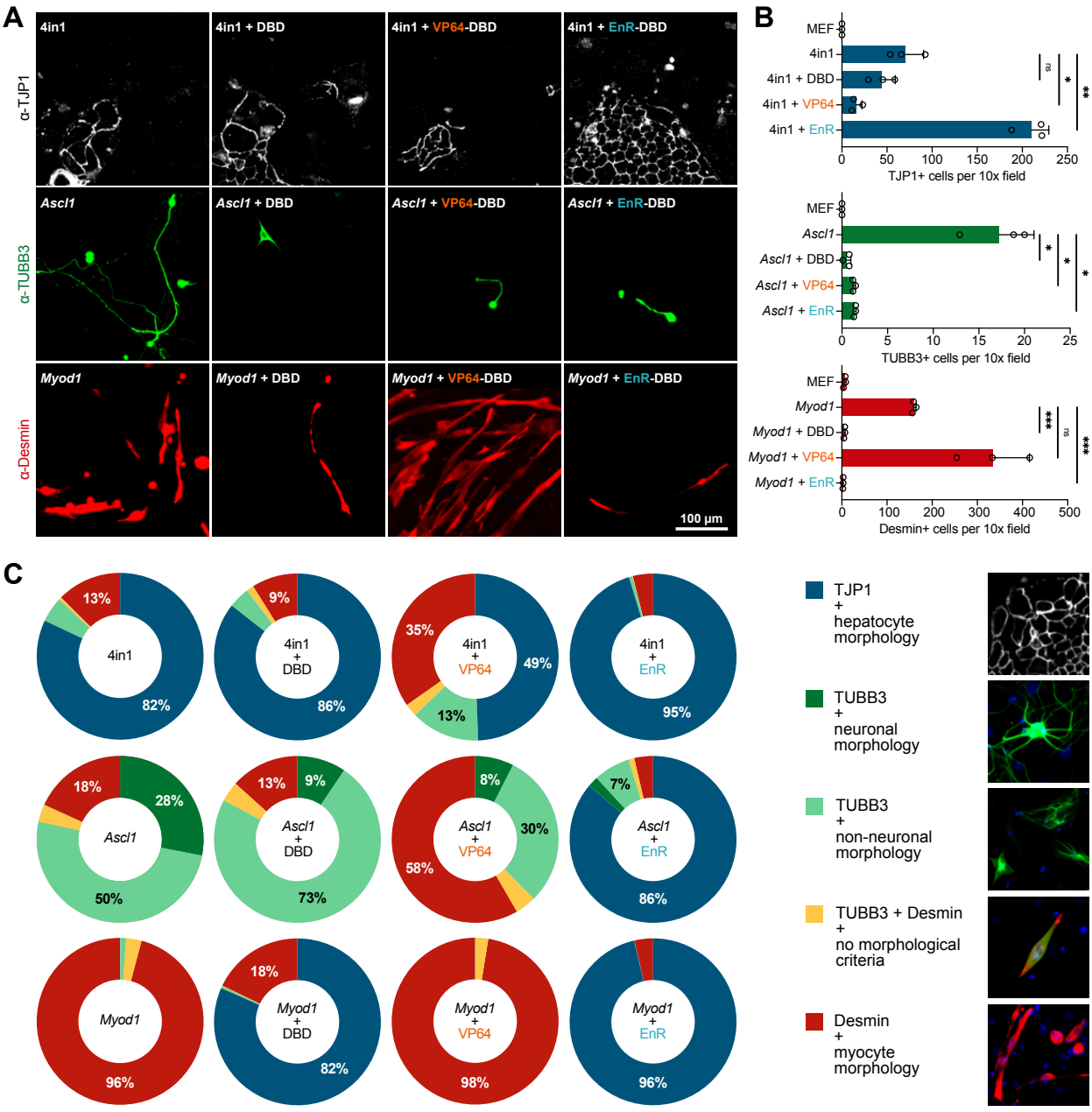

Figure 2.4.4. PROX1-DBD repressor fusion protein reduces alternative fate induction in hepatocyte reprogramming.
(A) Representative immunofluorescence microscopy images of cells at day 14 of reprogramming towards induced hepatocytes, neurons, or myocytes using indicated antibodies. Scale bar = 100 μm. (B) Quantification of successfully reprogrammed cells in A based on displayed immunofluorescence markers and morphology. (C) Percentage of reprogrammed cells in A positive for indicated markers and morphological criteria. Bar graphs show mean values from three biological replicates. Error bars = SD. Dunnett's test, * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.

fraction of TJP1-positive hepatocyte-like cells was increased upon repressor fusion expression during both neuronal and myocyte reprogramming (Figure 2.4.4C). While the activator fusion reduced neuronal cell induction, it did increase myocyte reprogramming efficiency, albeit not significantly. It also increased the fraction of Desmin-positive cells in both reprogramming contexts relative to DBD.

2.5 PROX1 represses alternative fate inducers

To investigate the molecular mechanisms underlying the repressive activity of PROX1, I sought to characterise the effects of PROX1 on chromatin organisation and to identify key PROX1 target genes that may explain its suppression of alternative cell identities. PROX1 is reported to interact with chromatin modifying co-factors which differ depending on the cellular context. For instance, HDAC3 is a repressive co-factor that is known to associate with PROX1 in the liver (Armour et al., 2017). PROX1 is also known to interact with the repressive Nucleosome Remodeling and Deacetylase (NuRD) complex in liver and in colorectal cancer cells (Högström et al., 2018; Ouyang et al., 2013). Interestingly, in lymphatic endothelial cells, PROX1 interacts with factors such as the histone acetyltransferase E1A-associated Protein p300 (EP300), which acts as a transcriptional activator (Wong et al., 2017), and with Nuclear Receptor Subfamily 2 Group F Member 2 (NRF2F2) (Lee et al., 2009). This may explain its differential activity between different cell types.

2.5.1 PROX1 decreases chromatin accessibility

To determine if PROX1 closes chromatin in the context of hepatocyte reprogramming, I performed an assay for transposase-accessible chromatin with sequencing (ATAC-seq). Following two days of 4in1-induced hepatocyte reprogramming, I found 111,411 differentially-accessible peaks (padj < 0.05), of which 85,140 (76.4%) were differentially closed upon overexpression of *Prox1* (Figure B.2.2). I also overexpressed *Prox1* or GFP as a control without 4in1, and obtained a similar number of differentially-accessible peaks (109,917, padj < 0.05), of which 84,800 (77.1%) were differentially closed upon overexpression of *Prox1*. I therefore concluded from this that, at the early time point of 48 hours, 4in1 does not strongly influence chromatin accessibility. On the other hand, PROX1 closes a large number of chromatin regions. This is reflected by principal component analysis, which showed that the presence of PROX1 explains 80% of the variation between conditions (Figure B.2.2).

Following initial quality controls, I proceeded to characterise the chromatin remodelling at PROX1-bound target genes using the previously-obtained PROX1 DNA binding data (Figures 2.3.3 and B.2.1). This analysis revealed a repressive PROX1 signature, with decreased accessibility on average across all PROX1-bound sites upon *Prox1* overexpression (Figure B.2.2E) (Berest et al., 2019). Furthermore, the majority of genes containing both a PROX1-bound and a differentially-accessible region in their promoters were down-regulated upon *Prox1* overexpression (Figure B.2.2F). From these observations, I concluded that PROX1 primarily closes chromatin at its directly-bound target genes in fibroblasts and early stages of hepatocyte reprogramming. This is likely a major mechanism by which PROX1 represses the expression of its target genes.

I next sought to understand how PROX1 promotes hepatocyte cell identity acquisition at later stages of reprogramming. In addition, I wanted to identify key PROX1 target genes whose

repression is responsible for hepatocyte fate stabilisation. For this, I analysed the transcriptomic changes that occur across a time course of hepatocyte reprogramming using bulk-RNA-seq. Specifically, I assessed differential gene expression during hepatocyte reprogramming with either GFP (control) or *Prox1* overexpression at days 7, 14, and 28 (Figure 2.5.1). This analysis showed that a total 8,036 protein-coding genes had an absolute $\log_2$FC of 0.75 between *Prox1* overexpression and control in at least one of the three time points analysed. Of these, 3,629 were up-regulated and 3,264 were down-regulated consistently in all three timepoints upon *Prox1* overexpression, relative to control.

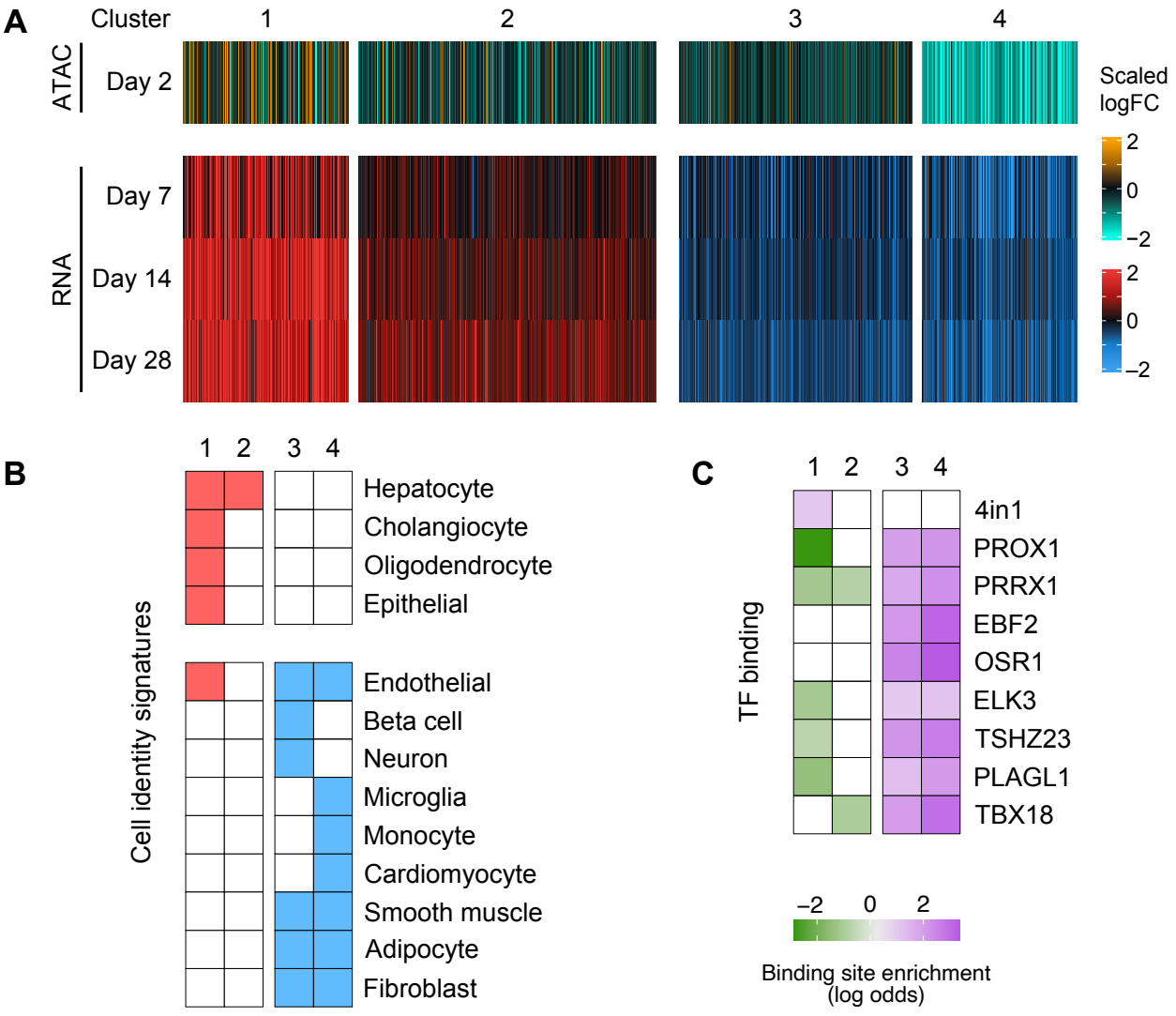

Figure 2.5.1. PROX1 represses alternative cell identities by chromatin closure.
(A) Chromatin accessibility (n=3) and gene expression analysis (n=2) comparing hepatocyte reprogramming with or without *Prox1* at indicated time points. Genes clustered based on expression and chromatin accessibility. Scaled logFC relative to control shown. (B) Overlap of cell identity marker genes with genes in each cluster in A. Overlaps with p-adj < 0.01 shown. (C) Enrichment or depletion of TF binding from CUT&RUN (for PROX1) or motif presence (for all other TFs) in the promoters of genes within each cluster in A. Log_2 odds ratios with p-adj < 0.05 shown.

I then grouped these 8,036 genes into four clusters based on the changes in promoter accessibility at day 2, and differential gene expression over the three time points (Figure 2.5.1A). Clusters 1 and 2 contained genes which were up-regulated upon *Prox1* overexpression. Cluster 2 genes were generally up-regulated at later time points and to a lesser extent, and also showed no differential promoter accessibility. Most cluster 1 genes, on the other hand, were up-regulated throughout the 4 week reprogramming period and displayed increased promoter accessibility at day 2. I also determined the enrichment of various cell type marker genes in each cluster (Figure 2.5.1B). Clusters 1 and 2 were both enriched for hepatocyte cell identity genes. However, cluster 1 also contained some non-hepatic genes such as cholangiocyte and oligodendrocyte markers.

Clusters 3 and 4 were down-regulated upon *Prox1* overexpression throughout the 28-day reprogramming experiment and also had reduced promoter accessibility at day 2 (Figure 2.5.1A). These two clusters were enriched for many non-hepatocyte cell identity marker genes, including neuronal, myocyte, fibroblast, and adipocyte genes (Figure 2.5.1B). The expression levels of genes in clusters 3 and 4 generally decreased over the 28-day reprogramming period upon *Prox1* overexpression.

Next, I set out to understand whether any of these clusters were directly regulated by the overexpressed liver reprogramming factors FOXA3, GATA4, HNF1A, and HNF4A (4in1). For this, with Aryan Kamal, I retrieved the target genes of the four transcription factors from the DoRothEA database and combined them into a unified 4in1 regulon containing 145 target genes (Garcia-Alonso et al., 2019).

In addition, I identified 448 high-confidence PROX1 direct target genes which were bound by PROX1 based on CUT&RUN analysis, and which were directly regulated by PROX1 as determined by activator or repressor fusion transcriptomics analysis (Figure 2.3.3 and Figure 2.4.2). The hepatocyte marker gene-enriched cluster 1 was significantly enriched for 4in1 target genes, and was also strongly induced during reprogramming (Figure 2.5.1). Interestingly, cluster 2, which is also hepatocyte-specific, was not enriched for direct 4in1 targets. This might imply that *Prox1* overexpression can enhance hepatocyte maturation independently of 4in1-induced gene activation. However, PROX1 target genes were neither enriched nor significantly depleted in either of clusters 1 or 2. Since both clusters contain hepatocyte marker genes, it is more likely that the increased expression of cluster 2 is induced by factors downstream of 4in1 targets. On the other hand, clusters 3 and 4 contain 9 alternative fate signatures and are significantly enriched in direct PROX1 targets. Since these clusters were down-regulated upon *Prox1* overexpression, this analysis lends further credence to the hypothesis that PROX1 mediates active repression of unwanted fates to promote liver cell induction.

2.5.2 PROX1 represses transcription factors driving alternative fates

To address how PROX1 silences multiple non-hepatocyte cell identities, and to identify key PROX1 target genes, I and Aryan Kamal used GRaNPA, a computational method which predicts the importance of specific transcription factors based on differential target gene expression (Kamal et al., 2023). For this, I constructed gene regulatory networks for PROX1 and for direct PROX1 targets that were annotated as transcription factors.

As expected, differential expression at day 2 was almost exclusively explained by PROX1 itself (Figure 2.5.2A). Importantly, PROX1 remained important at days 14 and 28, as did its direct targets, PRRX1 and Early B-cell Factor Transcription Factor 2 (EBF2). The observation that PROX1 remains important throughout the 4-week experiment supports its potential function in hepatocyte fate maintenance.

At day 7, GRaNPA predicted additional transcription factors, including the direct PROX1 targets, PRRX1 and EBF2, as being responsible for regulating a large proportion of the observed differential gene expression (Figure 2.5.2A). Starting at day 14, I found several additional transcription factors predicted to regulate liver reprogramming in a PROX1-dependent manner, but which are not direct PROX1 targets, such as Odd-Skipped-Related 1 (OSR1) and Erythroblast Transformation Specific Transcription Factor ELK3 (ELK3) (Fig. 4D). Indeed, *Osr1* and *Elk3* are targets of PRRX1, potentially explaining additional indirect effects downstream of PROX1 target genes later in reprogramming.

Strikingly, I found that the target genes of transcription factors predicted to be downstream of PROX1 were enriched in clusters 3 and 4, which contained many non-hepatocyte genes (Fig-

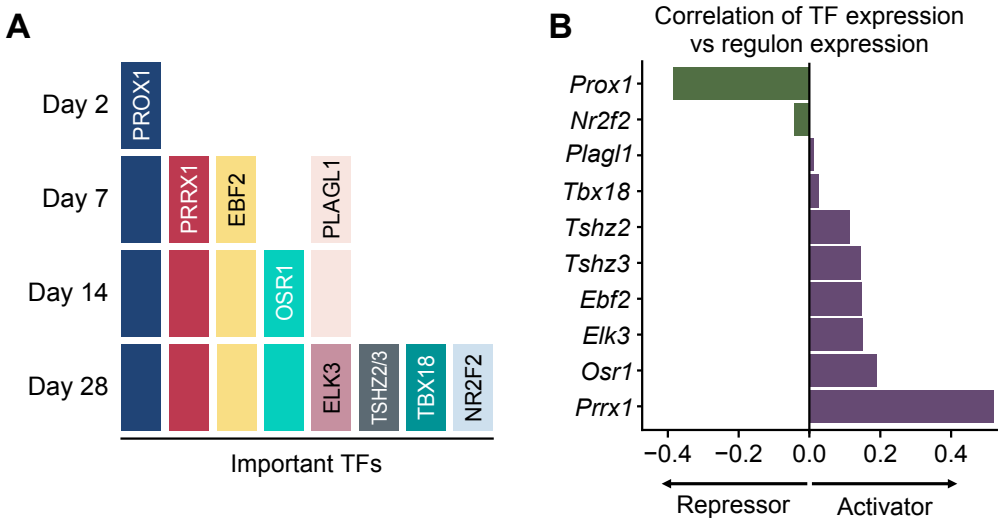

Figure 2.5.2. PROX1 induces gene regulatory network changes via transcriptional repression. (A) Computational prediction of TFs most important for PROX1-enhanced hepatocyte reprogramming reveals sequential induction of downstream transcription factors over the 28-day time-course. (B) Correlation of the expression levels of indicated transcription factors and the expression levels of their target genes predicts activator or repressor function. Negative correlations imply repressor function, while positive correlations imply activator function.

ure 2.5.1C and Figure 2.5.2A). Unlike PROX1, almost all of its target transcription factors are predicted to be activators based on correlation of their expression with the mean expression of their target genes (Figure 2.5.2B). This gene regulatory network analysis therefore suggests that PROX1 safeguards hepatocyte induction by silencing alternative cell fate genes, both directly and by repression of transcription factors that activate non-hepatocyte gene expression.

2.5.3 PROX1 directly represses fibroblast and adipocyte fate inducers

The donor fibroblast identity and alternative adipocyte identity were among the gene signatures most strongly repressed by PROX1, based on both bulk (Figure 2.5.1B) and single-cell transcriptomics analyses (Figure 2.3.4B) of hepatocyte reprogramming. Importantly, gene regulatory network analyses across the reprogramming time-course using GRaNPA identified master regulators of both fibroblasts and adipocytes as direct PROX1 targets.

One of the direct targets identified was EBF2, which is a co-activator of PPARG. Together, EBF2 and PPARG can cooperate to drive adipogenesis (Jimenez et al., 2007; Ma et al., 2018; Rajakumari et al., 2013; Rosen et al., 1999). Another direct target identified was PRRX1, which is a master transcription factor of stromal fibroblasts. PRRX1 regulates the differentiation of

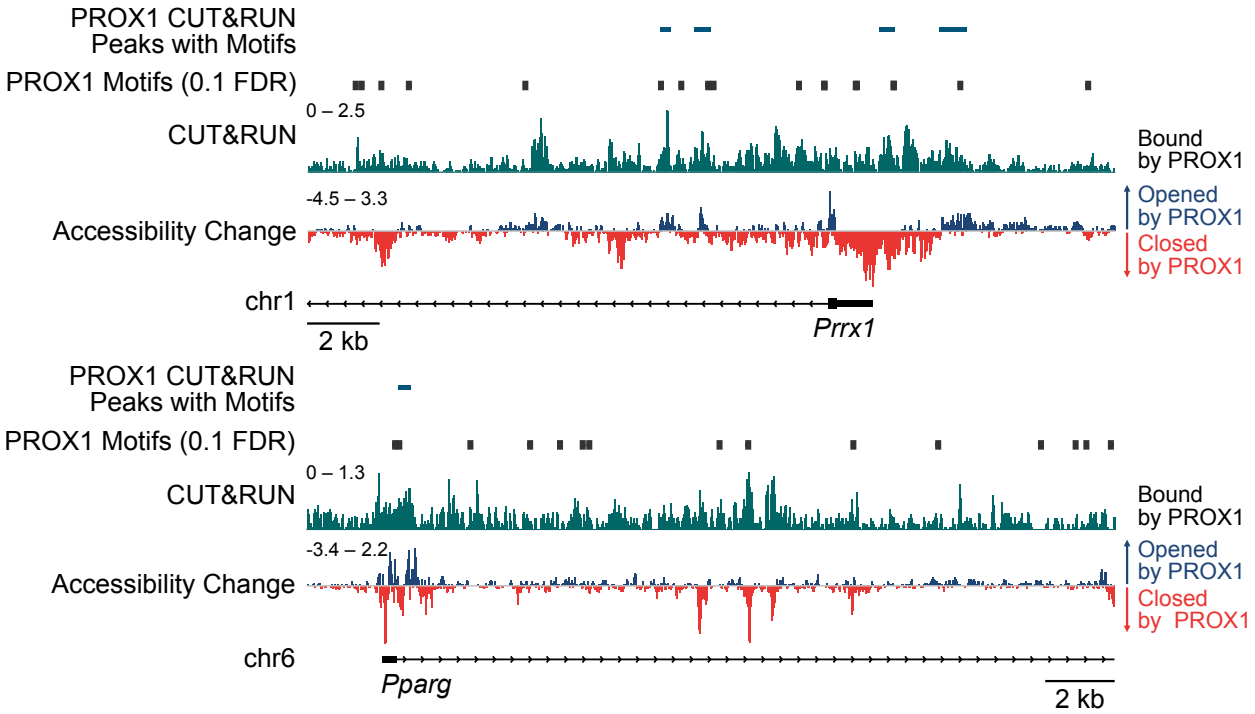

Figure 2.5.3. PROX1 binds and closes the promoter regions of *Prrx1* and *Pparg*.
Representative IGV tracks showing PROX1 chromatin binding based on CUT&RUN (n=3) and differential chromatin accessibility based on ATAC-seq (n=3). Data from day 2 of 4in1-induced hepatocyte reprogramming with or without *Prox1* overexpression. Regions shown are the promoters of *Prrx1* (top) or *Pparg* (bottom). PROX1 motifs and CUT&RUN peaks are displayed.

mesodermal cell types and driving fibrosis during wound healing (Leavitt et al., 2020; Lee et al., 2022).

Using CUT&RUN data, I confirmed that both *Prrx1* and *Pparg* are bound by PROX1 at their promoters. ATAC-seq data showed that both genes also displayed reduced chromatin accessibility near their transcriptional start sites upon *Prox1* overexpression, confirming that they are direct PROX1 targets (Figure 2.5.3). Therefore, I wanted to explore whether manipulating the expression levels of these master transcription factors, either by overexpression or knockdown, could explain the effects of PROX1.

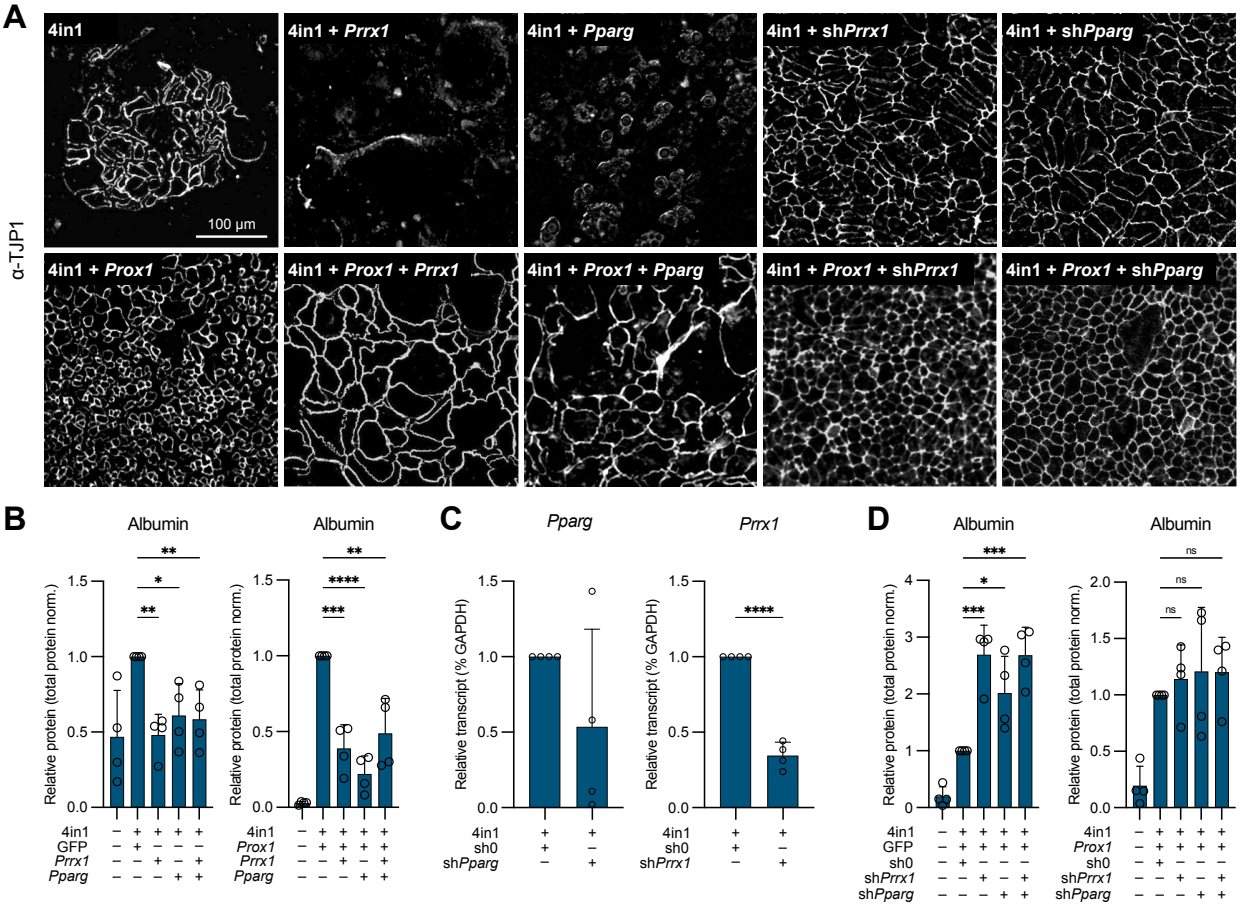

Figure 2.5.4. *Prrx1* and *Pparg* are directly repressed by PROX1 and their expression impairs reprogramming.
(A) Representative TJP1 immunofluorescence upon 4in1-induced hepatocyte reprogramming with overexpression or shRNA-mediated knockdown of *Prrx1* or *Pparg* at day 14, with or without *Prox1* co-overexpression (B) Quantification of Albumin protein levels at day 14 of hepatocyte reprogramming with overexpression of *Prrx1* or *Pparg* and co-overexpression of GFP control (left) or *Prox1* (right). (C) Expression of *Pparg* and *Prrx1* upon shRNA-knockdown at day 14 of hepatocyte reprogramming, determined by qRT-PCR. (D) Albumin protein levels determined by quantitative Western blot at day 14 of hepatocyte reprogramming upon knockdown of *Prrx1* or *Pparg*, with or without *Prox1* co-overexpression. Bar graphs show mean values normalised to total protein levels in (B and D), or GAPDH expression in (C). All data from four biological replicates. Error bars = SD. Dunnett's test (B and D) or two-tailed t-test (C), * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.

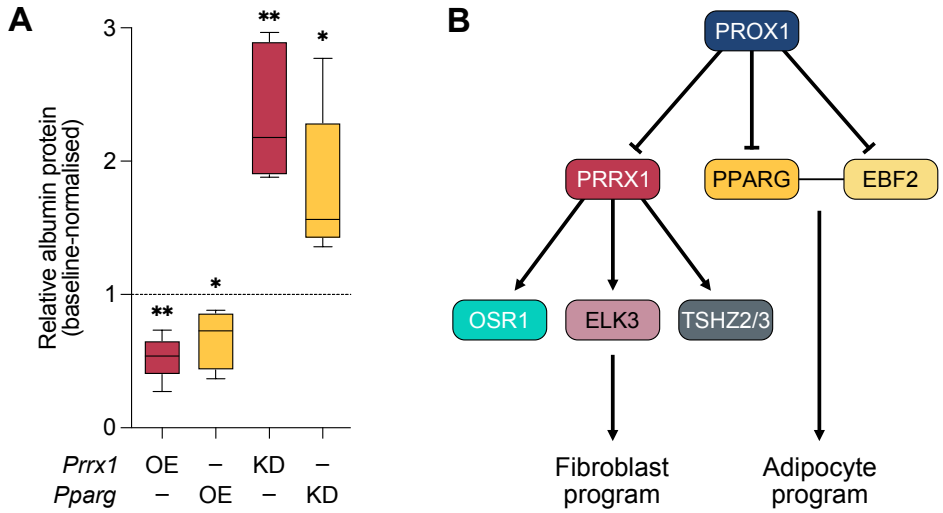

Figure 2.5.5. *Prrx1* and *Pparg* are key PROX1 target genes.
(A) Albumin protein quantification using Western blot of cells at day 14 of hepatocyte reprogramming with overexpression of *Prrx1* or *Pparg* or control. Data is normalised to respective controls in box plots (n=5). Fisher's LSD test, * p-adj < 0.05, ** p-adj < 0.01. (B) Proposed PROX1 gene regulatory network. Transcriptional repression of direct downstream transcription factors by PROX1 suppresses alternative cell identities.

First, I overexpressed *Prrx1* or *Pparg* during 4in1-mediated hepatocyte reprogramming. This impaired liver cell identity acquisition based on TJP1 expression and cellular morphology by immunofluorescence (Figure 2.5.4A). Both genes also reduced Albumin expression per cell by roughly 50% based on Western blot analysis (Figure 2.5.4B and Figure 2.5.5A). Furthermore, overexpression of *Prrx1* or *Pparg* in conjunction with *Prox1* overexpression partially negated the positive effects of PROX1 during liver reprogramming. This was observed again as decreased reprogramming efficiency by TJP1 expression and cellular morphology in immunofluorescence, and by Albumin expression by Western blot (Figure 2.5.4A and B).

Second, I performed shRNA-mediated depletion of *Prrx1* or *Pparg* during 4in1-mediated hepatocyte reprogramming, which was verified by qRT-PCR (Figure 2.5.4C). Similarly to *Prox1* overexpression, this led to approximately 1-2 fold higher Albumin expression per cell by Western blot (Figure 2.5.4D and Figure 2.5.5A) Additionally, the number of TJP1-positive hepatocyte-like cells was substantially higher (Figure 2.5.4A). Importantly, depletion of *Prrx1* or *Pparg* in conjunction with *Prox1* overexpression did not further enhance hepatocyte reprogramming based on Albumin protein levels and TJP1-immunofluorescence, providing further evidence that act downstream of PROX1 (Figure 2.5.4A and D).

In summary, PROX1 represses non-hepatic cell identities via direct transcriptional silencing of master regulators of alternative lineages. These PROX1 targets include the fibroblast fate driver *Prrx1*, and the adipocyte regulator *Pparg* (Figure 2.5.5B). The expression of both inhibits hepatocyte fate acquisition and their down-regulation, either by PROX1 or with shRNAs, leads to increased hepatocyte reprogramming efficiency.

2.6 PROX1 influences liver cancer initiation and development

Cell fate plasticity is an emerging hallmark of cancer. Inhibited differentiation, transdifferentiation, and dedifferentiation are potential drivers of neoplasia (Hanahan, 2022). The two dominant forms of primary liver cancer are cholangiocarcinoma (CCA) and hepatocellular carcinoma (HCC). Both differ in terms of morphology, metastatic proclivity, and response to therapy (Farazi & DePinho, 2006; Rizvi & Gores, 2013).

Interestingly, both CCA and HCC can arise from hepatocytes, and transdifferentiation between both subtypes has been reported (Fan et al., 2012; Seehawer et al., 2018). Despite this, the mechanisms underlying the capacity for the transdifferentiation and transformation of hepatic cells towards HCC or CCA are largely unknown. Based on our current understanding of HCC and CCA, and the results presented in previous sections, I hypothesised that safeguard repressors could act as tumour suppressors by maintaining cell fate fidelity in differentiation and by preventing cell fate plasticity. I therefore chose to explore the role of PROX1 as a hepatocyte cell fate safeguard in liver cancer.

2.6.1 *PROX1* expression levels segregate liver cancer patients

I first assessed whether *PROX1* expression differs between CCA and HCC samples from 153 liver cancer patients (Chaisaingmongkol et al., 2017). Using this dataset, I found that median PROX1 expression is 1.4-fold higher in HCC (n=62) patient samples than in CCA (n=91). Furthermore, *PROX1* expression in HCC tumour samples (n=62) is lower than in paired normal tissue controls (n=59) (Figure 2.6.1A and B).

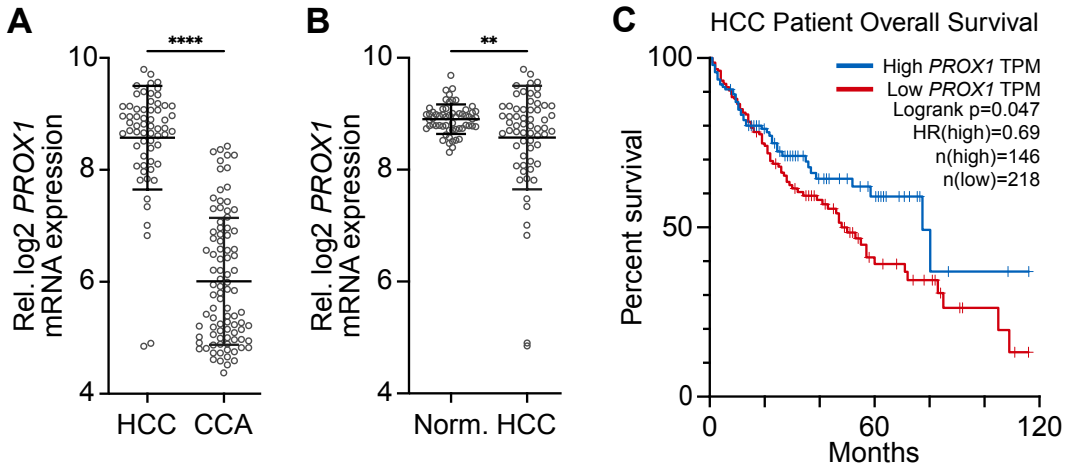

Figure 2.6.1. *PROX1* expression levels correlate with cancer subtype and survival.
(A) *PROX1* gene expression in tumour samples from cancer patients with HCC (n=62) and CCA (n=91) subtypes from the TIGER-LC dataset (Chaisaingmongkol et al., 2017). (B) *PROX1* expression in tumours from HCC patients in A (n=62) and paired normal tissue controls (n=59). (C) Overall survival of 364 HCC patients segregated by *PROX1* expression levels (40% cutoff for high-expression cohort) (Menyhárt et al., 2018). Log-rank test for survival. Scatter plots show mean values from specified biological replicates. Error bars = SD. Two-sided t-test, ** p-adj < 0.01, **** p-adj < 0.0001.

I proceeded to analyse whether *PROX1* expression levels in HCC patients correlates with disease prognosis. In a cohort of 364 patients with both transcriptomics and survival data (Menyhárt et al., 2018), I found that patients with high *PROX1* expression had a median survival of 81.9 months, while patients with low *PROX1* levels had a median survival of only 47.4 months (Figure 2.6.1C). These data collectively suggest that *PROX1* may act as a tumour suppressor in HCC.

2.6.2 PROX1 prevents HCC initiation

I next chose to determine whether PROX1 can influence liver tumorigenesis *in vivo*. To investigate this, I used a well-established HCC mouse model in which up-regulation of Myelo-Cytomatosis Proto-Oncogene (*Myc*) and down-regulation of Transformation-Related Protein 53 (*Trp53*) generates HCC-like tumours (Figure 2.6.2A). These genetic manipulations were performed by delivery of transposable elements into the liver via hydrodynamic tail-vein injection (HDTVI) which induced stable *Myc* overexpression and *Trp53* knockout (Revia et al., 2022).

After three weeks, mice developed multifocal liver carcinomas resembling HCC nodules with solid growth patterns. These tumours exhibited high protein expression of the hepatocyte-specific transcription factor, hepatocyte nuclear factor-4α (HNF4α). On the other hand, they did not express the biliary epithelial cell marker keratin 19 (KRT19) and lacked glandular structures typically found in CCA.

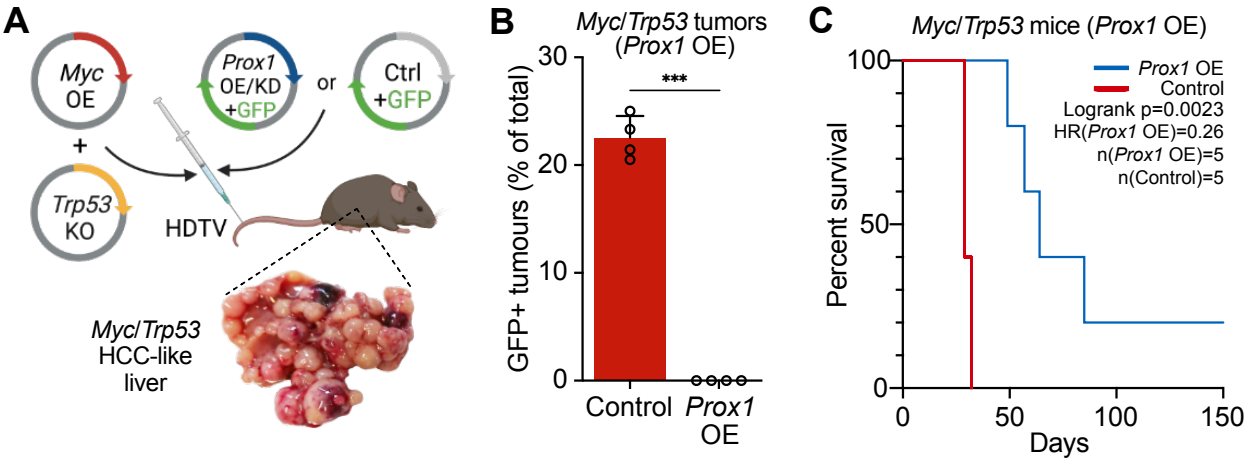

Figure 2.6.2. PROX1 prevents HCC tumour initiation and extends survival.
(A) Hydrodynamic tail vein injection (HDTVI) strategy induces HCC-like tumours by stable *Myc*-overexpression and *Trp53*-knockout. *Prox1* overexpression (OE) or knockdown (KD) constructs, or controls, were co-injected. (B) Percentage of GFP+ tumours in mice treated as in A, with *Prox1* or control overexpression (n=5). No *Prox1*-IRES-GFP positive tumours were present at the endpoint (n=4). (C) Overall survival of mice treated as in B following *Prox1* or control overexpression (n=5). Log-rank test for survival. Bar graph shows mean values from specified biological replicates. Error bars = SD. Two-sided t-test, *** p-adj < 0.001

Overexpression of *Prox1* in this model, within a construct that also contains an IRES-driven GFP marker, significantly decreased the number of tumour nodules at the endpoint (6.7 nodules vs 39.5 in control, $p < 0.001$). Importantly, all tumours in the animals injected with *Prox1*-IRES-GFP were GFP-negative at the endpoint (Figure 2.6.2B). This was despite detectable expression 48 hours following injection, suggesting that tumour formation was inhibited by *Prox1* expression. Strikingly, *Prox1* overexpression increased median survival from 29 days to 64 days, and caused a significant increase in overall survival duration (Figure 2.6.2C). Overall, these data suggest that PROX1 acts as a tumour suppressor by inhibiting tumour initiation.

I also performed an shRNA-mediated knockdown of *Prox1* in this model to investigate possible tumour suppression and inhibition of transdifferentiation. Depletion of *Prox1* led to a 10-fold increase in the number of detectable tumour nodules 14 days after HDTVI was performed. However, I did not observe any changes to tumour number, GFP-positive tumours, or survival at the endpoint (Figure C.1.1).

2.6.3 PROX1 inhibits HCC transdifferentiation

To determine whether PROX1 influences HCC identity *in vivo*, I performed immunohistochemistry analysis of mouse livers with HCC-like nodules following *Prox1* knockdown. I found that *Prox1* depletion induced a shift from HCC-like to CCA-like tumours. These nodules had significantly lower HNF4 expression and an accompanying increase in KRT19 expression. They also displayed a morphological shift, forming cholangiocyte-like duct structures (Figure 2.6.3A and B).

I then performed transcriptomics analysis of tumour nodules from the same mice using RNA-seq. This corroborated an HCC to CCA shift, with *Prox1* knockdown causing down-regulation of multiple hepatocyte markers, such as *Alb*, *Apoa1*, *Cdh1*, *Trf*, and *Ttr*. I also observed up-regulation of cholangiocyte markers such as *Krt19* and *Sox9* (Figure 2.6.3C).

This suggests that PROX1 can act as a safeguard repressor to maintain hepatocyte cell identity and prevent disease *in vivo*. Lower levels may permit plasticity and higher levels may reduce the potential for transformation and associated diseases.

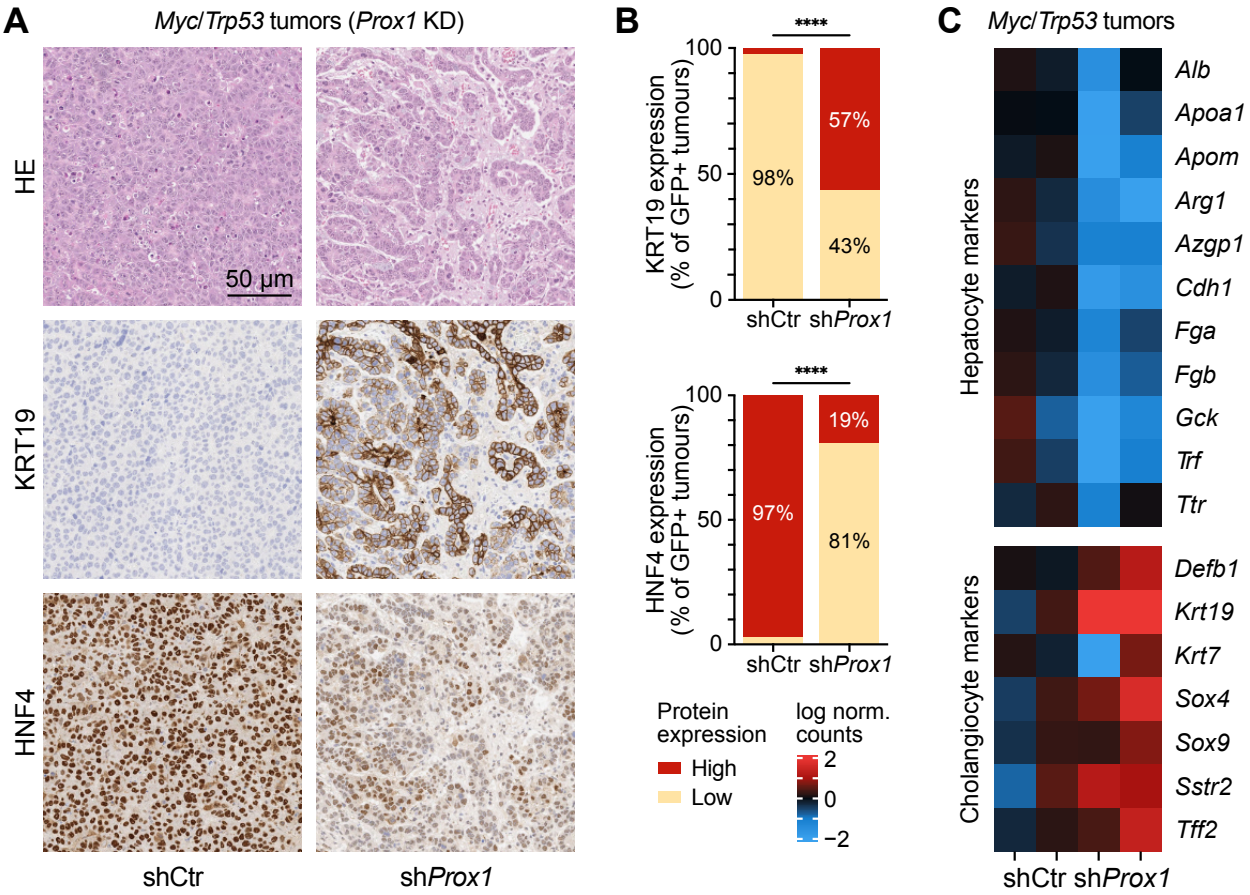

Figure 2.6.3. PROX1 prevents cell fate plasticity in HCC-like tumours.
(A) Immunohistochemistry of tumour models at the end point generated by HDTVI for stable *Myc*-overexpression and *Trp53*-knockout, with *Prox1* knockdown (KD) or controls. Hematoxylin and eosin (HE) staining as well as KRT19 and HNF4 antibodies following control or *Prox1*-knockdown shown. (B) Quantification of CCA (KRT19) and HCC (HNF4) marker protein levels in GFP+ tumors from A shown as percentage. Fisher's exact test, **** $p < 0.0001$. (C) Transcriptomics analysis of tumour nodules from mice in A using RNA-seq following knockdown of *Prox1* (n=2). Selected differentially-expressed cholangiocyte- and hepatocyte-related genes are shown. Bar graphs show mean values from specified biological replicates. Two-sided t-test, **** p-adj < 0.0001.

Chapter 3

Discussion

The activation of genes necessary for the induction of cell identity can be performed in most cases by a small number of transcription factors. However, repressive factors are likely to be responsible for fine-tuning the transcriptional output of these factors in order to generate more granular cell types and to reduce plasticity throughout life.

I propose that MYT1L, a neuron-specific transcriptional repressor, represents the founding member of a novel class of transcription factor: safeguard repressors (Mall et al., 2017). Loss of MYT1L results in the activation of various non-neuronal expression programs in mice and human induced neurons (Weigel et al., 2023). MYT1L depletion induced a confused transcriptional landscape within neurons throughout development, showing that MYT1L is necessary to suppress inappropriate gene expression.

3.1 Safeguard repression

In this study, I sought to establish whether other transcription factors might possess a similar cell identity safeguarding role to that of MYT1L. The identification of PROX1 as a repressive hepatocyte safeguard factor provides further evidence in support of the safeguard repressor hypothesis.

The importance of general repressive complexes has been well-researched. For instance, loss of repressive Polycomb group proteins in mouse embryonic stem cells (mESCs) causes activation of developmental transcription factors (Boyer et al., 2006). Additionally, deletion of Brahma (*Brm*), a BAF chromatin-remodelling complex member, induces neuronal transcription factor expression during directed differentiation of mESCs to cardiac myocytes. This indicates that *Brm* loss causes an identity shift from precardiac mesoderm to neural precursor cells (Hota et al., 2022).

Apart from these ubiquitously-expressed repressive chromatin regulators, only a few recent examples of cell type-specific safeguard transcription factors have been reported. For example,

Kmg, a male germ cell-specific repressor, down-regulates somatic lineage gene expression in *D. melanogaster* (Kim et al., 2017).

3.1.1 Computationally predicting safeguard repressors

To address the lack of known safeguard repressors, this thesis has used and validated a global computational screening strategy to identify potential repressive transcription factors with a role in cell identity maintenance. Suppression of inappropriate transcriptional activity can reduce cellular plasticity during development and disease. It has been suggested that epigenetic mechanisms may support canalisation to stabilise cell fate commitment (Pujadas & Feinberg, 2012). Safeguard repressors could serve a similar role by regulating target genes that are inappropriately accessible. I propose that they do so by binding to specific DNA motifs, and are expressed in a cell type-specific and continuous manner.

Some lineage-specific repressors may be temporally expressed, regulating lineage bifurcations by silencing specific alternative fates during development. They may also repress only one or a few alternative cell fates. One key example is that of REST, which represses neuronal gene expression in neuronal progenitor cells and non-neuronal cells (Schoenherr & Anderson, 1995). The criterion of continuous expression throughout life means that the screen used in this thesis identifies factors important not only in cell identity establishment, but also maintenance.

Safeguard repressors, by binding to genes involved in many unrelated cell identities, may suppress a larger number of alternative cell fates to maintain cell fate canalisation. Furthermore, safeguard factors that are expressed throughout life, or "terminal repressors", may inhibit cell fate plasticity not just in development but also in mature cells.

My findings lend credence to this model of cell type-specific safeguard repressors. Of 1,296 transcription factors analysed in this thesis for their potential to act as safeguard repressors, the computational screen identified 59 candidates in 18 cell types. More than 50% of these, including MYT1L, are previously described as promoting their associated cell type in development or reprogramming. While experimental validation of all 59 factors was not possible, I did investigate the hepatocyte factors, and especially PROX1.

3.2 PROX1 as a hepatocyte safeguard repressor

A key safeguard repressor candidate identified by this screen was PROX1. PROX1 fulfils the hypothesised criteria of a safeguard repressor, by being relatively specifically expressed throughout life in hepatocytes and acting via transcriptional repression to promote and maintain hepatocyte identity. The contextual implications of the experimental validation of the role of PROX1 in hepatocyte fate acquisition and maintenance, its cell type-specific repressive activity, and its potential role as a tumour suppressor in hepatocytes, will be the focus of the following section.

3.2.1 Validation of PROX1-driven alternative fate repression

Using MEF reprogramming towards induced hepatocytes, I identified PROX1 as a hepatocyte identity maintenance factor. *Prox1* showed high lifelong expression in hepatocytes and depletion of DNA binding motifs at hepatocyte-specific genes.

PROX1 has previously been described to regulate metabolic homeostasis in hepatocytes. *Prox1* depletion in adult mice causes the up-regulation of a broad set of fatty acid metabolism-related genes and subsequent dysregulation of lipid homeostasis, including increased triglyceride levels in the liver (Armour et al., 2017). Furthermore, loss of *Prox1* in development causes smaller livers in mice, and decreases hepatocyte commitment by bi-potent hepatic progenitors during development (Seth et al., 2014; Sosa-Pineda et al., 2000). PROX1 has also been reported to promote hepatocyte fate in human cells during reprogramming, although the mechanism behind this enhancement has not been investigated in detail (Du et al., 2014; Lu et al., 2021).

In this study, exogenous PROX1 was sufficient to enhance hepatocyte reprogramming over 10-fold and suppressed virtually all alternative cell identity gene signatures analysed. PROX1 was also necessary for efficient hepatocyte conversion. *Prox1* deletion reduced the number of reprogrammed hepatocytes 87%, and increased the fraction of cells that adopted alternative identities such as myocyte-like fates. This supports the hypothesis that PROX1 target genes must be silenced to facilitate proper hepatocyte fate and function.

PROX1 binding sites on chromatin were globally condensed. This led to repression of associated genes. This epigenetic regulation by PROX1, which primarily comprises transcriptional repression of donor and alternative fate genes, is most likely a major mechanism in the enhancement of hepatocyte identity acquisition.

Finally, I identified two transcription factors, *Prrx1* and *Pparg*, which are key PROX1 targets during hepatocyte reprogramming. These two transcription factors are master regulators of fibroblast and adipocyte lineages, respectively (Jimenez et al., 2007; Leavitt et al., 2020; Lee et al., 2022; Rosen et al., 1999). Their repression by PROX1 is responsible for a large proportion of the reprogramming improvement observed upon *Prox1* overexpression. This observation reflects the role of PROX1 as a direct transcriptional repressor of alternative fates.

3.2.2 Cell type specificity of safeguard repressors

The cellular context in which a transcription factor is expressed strongly defines its transcriptional output. As such, a cell type-specific experimental model is essential when studying transcription factor function. For instance, although the computational screen assumes a model where all binding motifs are equally accessible, this is almost certainly not the case. Additionally, expression of co-factors such as chromatin-modifying enzymes varies between cell types. This can also lead to different transcriptional effects by the same transcription factor in different cells.

PROX1 DNA binding motifs are enriched at many non-hepatocyte signature gene promoters, which reflects its potential role as a hepatocyte-specific repressive safeguard. However, this contrasts with other studies of PROX1 in other lineages. Notably, *Prox1* is expressed in some cells in the hippocampus and cerebellum in mice, in which it has been shown to promote neurogenesis (Karalay et al., 2011; Lavado & Oliver, 2007; Lavado et al., 2010).

PROX1 is also known to be the master regulator of lymphatic endothelial cell differentiation. For instance, *Prox1* ablation in mouse lymphatic vessels induces cell fate loss, with cells undergoing a transition towards a blood endothelial cell-like state (Johnson et al., 2008; Petrova, 2002). PROX1 binds to NR2F2 (a.k.a COUP-TFII) in lymphatic endothelial cells, forming a heterodimer complex. This complex is responsible for the expression of many lymphatic endothelial cell-specific genes directed by the sequence specificity of PROX1. The lack of PROX1 in blood endothelial cells means the NR2F2 homodimer complex, which has different target genes, is predominant (Aranguren et al., 2013; Lee et al., 2009). This, along with evidence for its transcriptional activator function, may suggest that PROX1 acts as a terminal selector in this lineage (Wigle, 2002).

Although PROX1 primarily activates gene expression in combination with co-activators in non-liver cellular contexts (Aranguren et al., 2013; Iwano et al., 2012; Lee et al., 2009), PROX1 is known to interact with and co-bind genomic loci with the co-repressor HDAC3 and the hepatocyte-driving transcription factor HNF4A in the liver (Armour et al., 2017). Even assuming that chromatin accessibility across different cell types is the same, the presence of different co-factors at the same PROX1-bound genes is likely to induce distinct alterations in chromatin accessibility. Such variability in co-factor expression could be responsible for repurposing the DNA sequence specificity of transcription factor binding to alter transcriptional output and fate specification between different cell types. This may explain how cell type-specific master regulators can convert a terminal selector into a safeguard repressor, as in the case of PROX1 in lymphatic endothelial cells and hepatocytes, respectively. It may also explain why PROX1 enhanced hepatocyte reprogramming but prevented neuronal reprogramming, which is interesting since PROX1 is involved in the development of some neuronal subtypes (Iwano et al., 2012; Miyoshi et al., 2015; Stergiopoulos et al., 2015). One explanation for this discrepancy may be that *Prox1* overexpression

in fibroblasts as a starting cell type may display primarily repressive effects due to low expression of activating co-factors of PROX1, with the result of suppressing neuronal reprogramming.

The use of synthetic biology approaches, whereby the PROX1 DNA-binding domain was fused with either an activator or repressor effector domain, was important in decoupling PROX1 target binding from the effects of variable expression of its co-factors. This experiment confirmed that PROX1 target gene repression promotes hepatic fate attainment in hepatocyte reprogramming, and importantly showed that target gene activation by the activator fusion induced alternative fates. Here, too, the effects on neuronal reprogramming were interesting and unexpected. Both the repressor and activator fusion, as well as the DBD alone, caused suppression of reprogramming. Once again, this could be due to a lack of activating co-factors of PROX1, though the increase in myocyte reprogramming efficiency induced by the activator fusion protein is an interesting contrast to the effect in neurons.

3.2.3 PROX1 as a hepatocyte tumour suppressor

Due to the role of cellular plasticity in many diseases, including cancer, I investigated the role of PROX1 in maintaining hepatocyte cell fate and preventing disease in hepatocellular carcinoma. PROX1 expression is decreased in HCC patients, and low expression is associated with poor survival. This would suggest that PROX1 prevents hepatocyte plasticity and transformation.

Our data in a mouse HCC model shows that *Prox1* overexpression completely prevents neoplastic transformation. This contradicts some previous studies, in which *Prox1* overexpression in established tumour cell lines led to higher levels of cell migration, proliferation, drug resistance, and metastasis upon transplantation (Liu et al., 2015b; Liu et al., 2013). Other studies report conflicting results, with *Prox1* knockdown in HCC cell lines leading to accelerated growth *in vitro* and overexpression leading to reduced proliferation (Chang & Hung, 2013; Shimoda et al., 2006). *PROX1* expression in HCC cell lines has also been observed to increase upon treatment with rapamycin. Interestingly, knockdown of *PROX1* causes a decrease in the anti-proliferative effect of rapamycin, suggesting that PROX1 facilitates the anti-tumour effect of rapamycin (Kwon et al., 2022). Loss of *PROX1* has also been observed in CCA patient samples, with either promoter silencing by methylation or genomic deletions causing a decrease in *PROX1* expression (Laerm et al., 2007). On the other hand, another study showed that HCC and CCA patient samples expressed *PROX1*, with differential splicing of *PROX1* mRNA correlating with the choice between HCC and CCA fate (Dudas et al., 2008).

PROX1 has also been implicated in cancers of other tissues (Ntikoudi et al., 2022). Here, its function is also unclear. Loss of *Prox1* in mouse models of colorectal cancer causes derepression of matrix metalloproteins, permitting invasion (Ragusa et al., 2020). In support of this, PROX1 represses matrix metalloprotease 14 expression in an HCC cell line, mouse dermal lymphatic vessels, human lymphatic endothelial cells, and breast cancer cells (Gramolelli et al., 2018). Furthermore, high *PROX1* expression has been shown to correlate with improved patient outcomes

in gastric cancer (Laitinen et al., 2017). On the other hand, PROX1 has been reported to suppress Notch pathway activation in colorectal cancer stem cells by interaction with the NuRD complex, preventing their differentiation and promoting tumour progression (Högström et al., 2018). The role of PROX1 in squamous cell carcinomas of the upper digestive tract is similarly controversial. Oral squamous cell carcinoma is suppressed by PROX1, while high PROX1 levels predict poorer outcomes in oesophageal squamous cell carcinoma (Rodrigues et al., 2014; Yokobori et al., 2015).

Unexpectedly, *Prox1* knockdown caused no changes to tumour initiation rates and the survival of mice in the same genetic HCC model. However, the resulting *Prox1* knockdown tumours exhibited an identity switch away from HCC and towards CCA. Hepatocytes are a potential cell of origin for both HCC and CCA (Fan et al., 2012; Mu et al., 2015; Seehawer et al., 2018; Sun et al., 2019). The determinants of liver tumour identity are not fully understood, but a necroptotic tumour microenvironment with high cytokine levels is associated with the development of CCA from hepatocytes that have been oncogenically transformed using HDTVI of *Myc* and *Ras* (Seehawer et al., 2018). On the other hand, an apoptotic microenvironment causes the same transformed hepatocytes to develop into HCC. Further mechanistic investigation found that the levels of two transcription factors, T-Box Transcription Factor 3 (*Tbx3*) and PR Domain 5 (*Prdm5*), were influenced by the different microenvironments and were responsible for regulating the expression of determinants of HCC or CCA fate (Seehawer et al., 2018). Our data indicate that PROX1 may be another potential determinant of HCC or CCA fate in transformed hepatocytes, further supporting the role that safeguard repressors may play in preventing cell fate plasticity and blocking cancer development *in vivo*.

3.3 Conclusion

This thesis presents several key findings. It shows that computational prediction is capable of generating testable lists of safeguard repressor candidates. With direct cell reprogramming, I demonstrate that PROX1 safeguards hepatocyte identity by directly suppressing alternative lineage programs, hence demonstrating safeguard repression in a second cell type and germ layer.

Using single-cell transcriptomics, I provide further evidence that PROX1 silences donor and alternative cell identities. PROX1 directly binds to promoters of mesodermal master regulators, including *Prrx1* and *Pparg*, decreases the chromatin accessibility at those loci, and reduces expression of those factors. This provides a mechanistic explanation for the repression of the fibroblast and adipocyte gene expression programs by PROX1. Importantly, repression of *Prrx1* and *Pparg* during hepatocyte reprogramming mimicked effects of *Prox1* overexpression. On the other hand, *Prox1* deletion or target gene activation permitted inappropriate gene expression.

Finally, I use a hepatocellular carcinoma mouse model to show that *Prox1* overexpression prevented tumour initiation, reducing overall tumour load and extending survival substantially. *Prox1* depletion in the same model induced cholangiocarcinoma-like morphology and gene expression, demonstrating the role of PROX1 in cell identity maintenance even during oncogenesis.

Overall, these results show that PROX1 is a repressive safeguard of hepatocyte identity. They also support a model whereby continuous silencing of alternative lineage gene expression programs by safeguard repressors prevents cell fate plasticity, thereby maintaining cell identity.

3.3.1 Significance

The significance of this study lies in its contribution to our understanding of the mechanisms involved in maintaining cell identity and preventing inappropriate cell fate plasticity. By demonstrating that PROX1 safeguards hepatocyte identity by directly suppressing alternative lineage programs, it adds to existing evidence indicating a need for continuous silencing of alternative lineage gene expression programs by cell type- and sequence-specific transcription factors. Such safeguard repressors, which include MYT1L, prevent cell fate plasticity, thereby maintaining cell identity. This has important implications for our understanding of how cell identity is maintained and how it can be disrupted in diseases such as cancer. This study's findings on the role of PROX1 in preventing liver tumorigenesis, and its utility in promoting hepatocyte reprogramming may also have potential therapeutic applications such as cancer treatment and cell therapy.

It is generally accepted that complexes of transcription factors and other regulatory cofactors coordinate to define a cell type-specific gene expression program, establishing a particular cell type (Arendt et al., 2016). Transcription factors within such complexes that help to maintain terminal cell identity have been referred to as terminal selectors (Hobert, 2008). New cell types have been proposed to form through the gain or loss of terminal selectors, or their co-factors, from cell type-defining complexes (Arendt et al., 2016). This may explain why PROX1 plays an

important role not just as an activating terminal selector in lymphatic endothelial cells (Johnson et al., 2008), but also, as proposed here, a repressive terminal selector in the liver. The mechanisms underlying this differential activity may include the presence of different transcriptional machinery in each cell type, which could alter co-factor recruitment. Another potential mechanism could be differential accessibility of PROX1 binding sites between cell types. PROX1 may target different genes in hepatocytes and lymphatic endothelial cells, for instance. This study could therefore serve as a basis for further investigation into how PROX1 and other transcription factors modulate their activity by interacting with different regulatory elements and proteins. If further evidence is found for the existence of safeguard repressors in other tissues, this study may be a basis for generalising the concept of *terminal repressors* – i.e., terminal selectors with a safeguarding and transcriptionally repressive function.

3.3.2 Limitations

One limitation of this research is that it focuses primarily on the role of PROX1 in safeguarding hepatocyte identity, and does not fully explore the mechanisms behind differential regulation of target genes by PROX1 in different tissues. For instance, PROX1 seems to be a transcriptional activator in neuronal subpopulations and lymphatic endothelial cells. Further study is necessary to develop an understanding of how different combinations of transcriptional regulators generate distinct gene regulatory networks. This includes not only transcriptional co-factors with no DNA sequence specificity, but also potential combinatorial effects of interactions with other DNA-binding transcription factors.

Another limitation is that while the study presents evidence for the role of PROX1 in preventing liver cancer, the mechanisms underlying this process and the role of PROX1 in tumour progression and therapy response are not fully understood. Further research is needed to explore these questions, possibly in models with other transgenes in HDTVI or in non-HDTVI-dependent models. Additional studies in human models would also be more relevant to clinical application.

This study makes several assumptions for the candidate safeguard repressor screen that may not be entirely accurate. Firstly, gene expression data and marker gene set lists were derived from mouse, not human. As such, it may make predictions that do not hold across both species. Secondly, it assumes that motif density is highly correlated with transcription factor binding, and that motifs in the promoter have a large influence on transcriptional regulation than those farther away. This may not be true due to the variability of chromatin landscapes in different cell types, and does not consider the role of chromatin looping and enhancers on gene expression. Thirdly, the screen did not consider the potential roles of co-factors and other transcription factors that may cooperate with a given candidate. Further work could build on the existing screen by using human data and curated marker gene lists, as well as incorporating a model to take into account chromatin accessibility and conformation in different cell types. It could also take a combinatorial approach, considering the effects of two or more factors in conjunction.

Finally, while this study presents a broad framework for the identification of repressive transcription factors with a role in cell fate maintenance, it does not validate them in more than one cell type. An extension to this study that investigates other cell types, especially those in the mesodermal germ layer, would greatly increase the amount of evidence supporting the existence of safeguard repressors. Nevertheless, it does recapitulate data about known repressors and cell identity inducers in a wide range of cell types.

3.3.3 Future perspectives

The advent of single-cell transcriptomics has shown that gene expression levels are heterogeneous even within populations of a single cell type (Treutlein et al., 2016). Alongside this heterogeneity in gene expression comes heterogeneity in all the phenotypes of a cell that are transcriptionally regulated, including cell identity and plasticity. Terminal repressors that are expressed throughout life and continuously silence inappropriate gene expression programs may be key regulators responsible for finely controlling cell type and preventing disease. Understanding how this process might manifest in other cell types could inform further work in development and disease, and might reveal safer reprogramming factors for better *in vitro* models and improved cell therapy techniques. Inappropriate gene expression due to loss of safeguard mechanisms such as repression of alternative fates may lead to other functional deficits, such as aging and loss of homeostasis leading to dysfunction. Further understanding the role of safeguard repressors such as PROX1 in maintaining proper biological function could lead to greater insight into transcriptional noise as a cause of disease.

Some potential future directions for research into PROX1 could include further investigation into the mechanisms by which PROX1 and other terminal repressors maintain cell identity by suppressing alternative lineage programs. This could involve studying the interactions between PROX1 and other transcription factors or epigenetic machinery, as well as the epigenetic changes that occur during cell fate determination. Such epigenetic changes may be more substantial than the chromatin closure observed in this study, and

Additionally, it would be interesting to explore the potential therapeutic applications of these findings, such as developing strategies to manipulate *PROX1* expression or activity in order to prevent or treat liver cancer. If PROX1 acts as a liver tumour suppressor in humans, as it seems to do in mice, its expression could be induced in HCC patients by gene therapy, perhaps delivered via adeno-associated virus vectors, or as an mRNA-based therapy. It could also serve as a diagnostic marker to determine treatment choice, since higher *PROX1* expression is associated with HCC as opposed to CCA, and may serve as a prognostic tool since higher expression is associated with improved patient survival. Other liver diseases, such as non-alcoholic fatty liver disease and cirrhosis, are also associated with cell identity loss. Studies into the roles of PROX1 in these diseases may also yield therapeutically-relevant results.

Chapter 4

Methods

4.1 *In vitro* experiments

4.1.1 Primary cell line derivation

Mouse embryonic fibroblasts (MEFs) were harvested from E13.5 mouse embryos derived from pregnant C57BL/6N (WT) or C57BL/6J (*Prox1*$^{fl/fl}$) mice (Martinez-Corral et al., 2015) (Table 4.1.1). For each MEF line, the distal portions of all four limbs from four mouse embryos were dissected, placed into 100 µL trypsin, and cut thoroughly with scissors. Additional trypsin was added to a total volume of 1 mL, and the suspension was then incubated (37 ℃, 15 min).

Trypsin was inactivated by addition of the cell suspension to 25 mL of MEF media (Table 4.1.2). MEFs were then cultured on one 15 cm dish per MEF line until confluent, after which cells were split 1:4 onto four 15 cm dishes. When confluent, cells were pooled then cryopreserved in FBS with 10% DMSO, or passaged once more before use in reprogramming experiments.

Cell line	Source	Culture medium
HEK-293T	ATCC (Manassas, USA)	MEF medium
MEFs	Prepped from E13.5 mouse embryos	MEF medium

Table 4.1.1. Sources of cell lines and their culture media

Medium	Components
MEF medium	DMEM, Cosmic Calf Serum (1:10), GlutaMAX (1x), Penicillin-Streptomycin (Pen-Strep) (1x), MEM-NEAA (1x), sodium pyruvate (1mM), β-mercaptoethanol (0.1 mM)
N3 medium	DMEM-F12 (L-Glutamine), N2 supplement (1x), B27 supplement (1x), Pen-Strep (1x), Insulin (20 μg/mL)

Table 4.1.2. Self-made cell culture media

Medium	Source
DMEM	Thermo Fisher Scientific (Carlsbad, USA)
DMEM/F12 (L-Glutamine)	Thermo Fisher Scientific (Carlsbad, USA)
HBSS	Thermo Fisher Scientific (Carlsbad, USA)
Hepatocyte Complete Medium	Lonza (Basel, Switzerland)
MEM	Thermo Fisher Scientific (Carlsbad, USA)
Opti-MEM	Thermo Fisher Scientific (Carlsbad, USA)

Table 4.1.3. Commercial cell culture media

4.1.2 Virus production

Lentivirus was produced through transfection of lentiviral backbones containing indicated transgenes and third-generation packaging plasmids (Table 4.6.1) into HEK293T cells (Table 4.1.1) according to the Trono laboratory protocol (Dull et al., 1998). Lentivirus was concentrated from HEK293T culture supernatant through ultra-centrifugation (23,000 rpm, 2 h, 4 ℃) and used immediately or stored for future use at -80 ℃.

4.1.3 Direct reprogramming from MEFs

Wild-type or *Prox1*$^{fl/fl}$ MEFs were transduced by incubation with lentivirus prepared from indicated plasmids in MEF medium with 8 µg/mL polybrene (Sigma) for 16 to 20 h. Medium was exchanged for fresh MEF medium containing 2 µg/mL doxycycline (Sigma) to induce transgene expression.

For myocyte and neuronal reprogramming, cells were kept in MEF medium for 48 hours. Medium was then exchanged for N3 medium with doxycycline (Adrian-Segarra et al., 2021). Medium was changed every 2 days for the remainder of reprogramming.

For hepatocyte reprogramming (Song et al., 2016), MEFs were seeded onto plates pre-coated for 1 hour with 10% collagen in 0.55% acetic acid in water. Medium was supplemented 24 hours following dox induction with 0.5x volume Hepatocyte Culture Medium containing 5% FBS (Life Technologies) and 2 µg/mL doxycycline. After another 24 hours, all medium was then exchanged for a mixture of 1/3 MEF medium and 1/3 HCM + 5% FBS with 2 µg/mL doxycycline. Medium was then changed every 2 days for the remainder of reprogramming.

Reagent	Source
Accutase	Thermo Fisher Scientific (Carlsbad, USA)
EDTA (Ultrapure, 0.5M, pH 8.0)	Thermo Fisher Scientific (Carlsbad, USA)
Phosphate-buffered saline	Thermo Fisher Scientific (Carlsbad, USA)
Polyethyleneimine (PEI)	Polysciences (Warrington, USA)
Poly-L-ornithine	Merck (Darmstadt, Germany)
Trypan blue	Thermo Fisher Scientific (Carlsbad, USA)
Trypsin-EDTA, 0.25%	Thermo Fisher Scientific (Carlsbad, USA)

Table 4.1.4. Cell culture reagents

Reagent	Source
B27 supplement	Thermo Fisher Scientific (Carlsbad, USA)
β-mercaptoethanol	Thermo Fisher Scientific (Carlsbad, USA)
Cosmic Calf Serum	Thermo Fisher Scientific (Carlsbad, USA)
Doxycycline	Merck (Darmstadt, Germany)
GlutaMAX	Thermo Fisher Scientific (Carlsbad, USA)
Insulin	Thermo Fisher Scientific (Carlsbad, USA)
L-Glutamine	Thermo Fisher Scientific (Carlsbad, USA)
MEM-NEAA	Thermo Fisher Scientific (Carlsbad, USA)
N2 supplement	Thermo Fisher Scientific (Carlsbad, USA)
Pen-Strep	Thermo Fisher Scientific (Carlsbad, USA)
Sodium pyruvate	Thermo Fisher Scientific (Carlsbad, USA)

Table 4.1.5. Cell culture medium additives

4.2 Molecular biology

4.2.1 SDS-PAGE and Western blot

Cells were washed once in PBS and harvested by addition of 1x SDS buffer in RIPA buffer. Samples were sonicated (6 cycles, 30 seconds on, 1 minute off). Samples were separated in NuPAGE 4-12% Bis-Tris gels at 100 to 150 V in MOPS buffer. Proteins were transferred onto nitrocellulose membranes (overnight at 75 mA or 2h at 300 mA) in transfer buffer.

A total protein stain was performed on membranes with the Revert 700 Total Protein Stain (LI-COR). according to manufacturer guidelines. Membranes were imaged with the ODYSSEY CLx Imaging System (LI-COR). Total Protein Stain was then removed by incubation of the membrane in reversal buffer for 5 min.

The membrane was blocked in 2% BSA in PBS-T for 1-2 h at RT or 4°C overnight. The membrane was then incubated with primary antibody diluted in 2% BSA in PBS-T (1 h at RT or overnight at 4 °C) on a shaking platform. Three washes were performed in PBS-T for 15 min each.

The membrane was incubated with secondary antibody diluted in 2% BSA in PBS-T (1 h at RT), shielded from light, on a shaking platform. Three washes were performed in PBS-T for 15 min each. Membranes were imaged with the ODYSSEY CLx Imaging System (LI-COR). Protein expression was determined by densitometry, and expression of target proteins normalised to that of respective total protein stains with the Image Studio Lite software.

4.2.2 Quantitative real-time polymerase chain reaction

Samples were harvested by addition of TRIzol and stored at -80°C until use. RNA was isolated with the Zymo Research RNA Miniprep kit following manufacturer instructions. cDNA synthesis was performed with the LunaScript RT Super Mix Kit according to manufacturer guidelines, and stored at -20°C until use.

Quantitative measurement of gene expression was performed by qPCR in 384-well plates on the QuantStudio 5 Real Time PCR system. Each reaction contained 5 µL Power SYBR Green PCR Master Mix, 2.1 µL nuclease-free water, 0.4 µL 5 µM primer mix comprising 5 µM forward and 5 µM reverse primer, and 2.5 µL of cDNA. DNA oligonucleotide primers for quantitative PCR were ordered from Sigma. All primers used in this study are listed in Table 4.2.3.

qPCR was performed under the conditions in Table 4.2.2. Data processing and visualisation were performed with the QuantStudio web application (Thermo Fisher), Microsoft Excel, and GraphPad Prism.

Buffer	Components
BSA	Merck (Darmstadt, Germany)
NuPAGE MOPS buffer (20x)	Thermo Fisher Scientific (Carlsbad, USA)
NuPAGE Transfer buffer (20x)	Thermo Fisher Scientific (Carlsbad, USA)
PageRuler Plus Pre-stained Protein Ladder	Thermo Fisher Scientific (Carlsbad, USA)
PBS-T	0.1% Tween-20 in PBS
Protease inhibitor	Merck (Darmstadt, Germany)
Revert-700 total protein stain	LI-COR Biosciences (Lincoln, USA)
RIPA buffer	150 mM NaCl, 50 mM Tris pH 7.4, 1% Triton X-100, 0.5% DOC, 0.1% SDS, 1 mM EDTA in water
SDS sample buffer (5x)	250 mM Tris pH 6.8, 50% (v/v) glycerol, 20% (w/v) SDS, 250 mM DTT, 0.1% (w/v) bromophenol blue in water
Total protein stain reversal solution	0.1 M sodium hydroxide, 30% (v/v) methanol in water
Total protein stain wash solution	6.7% (v/v) glacial acetic acid, 30% (v/v) methanol in water
WB transfer buffer	20% methanol in 1x NuPAGE transfer buffer

Table 4.2.1. SDS-PAGE and Western blot reagents

Step	Temperature	Duration
1	50°C	2 min
2	95°C	10 min
3	95°C	15 sec
4	60°C	1 min
		40 cycles

Table 4.2.2. qRT-PCR amplification conditions

4.2.3 Immunocytochemistry

All immunocytochemistry data are based upon cells fixed 14 days after transgene induction and imaged by immunofluorescence microscopy. Cells were washed twice with PBS, then fixed directly in 24-well plates with 4% PFA in PBS for 10 min. Cells were washed twice with PBS before further steps.

Cells were then permeabilised with 0.25% Triton X-100 in PBS for 10 min, blocked in a solution of PBS containing 2% BSA for 1 h, then incubated with primary antibody (Table 4.7.1) diluted in PBS containing 2% BSA overnight at 4 °C. Cells were washed three times in PBS, then incubated in the dark for 1 h at RT with secondary antibodies (Table 4.7.2) diluted in PBS containing 2% BSA. Cells were then washed three times in PBS containing 2% BSA, then incubated

Target	Forward Primer (5'-3')	Reverse Primer (5'-3')
mmAfp	TGGTTACACGAGGAAAGCCC	AATGTCGGCCATTCCCTCAC
mmAlb	CCCACTAGCCTCTGGCAAAA	ACTCTTGTGTGCTTCTCGGC
mmCol1a1	GAGAGGTGAACAAGGTCCCG	AAACCTCTCTCGCCTCTTGC
mmDesmin	AACTTCCGAGAAACCAGCCC	CACAAAGGGGTGATCGGTGA
mmEcadh	AACCCAAGCACGTATCAGGG	GAGTGTTGGGGGCATCATCA
mmGAPDH	TATGATGACATCAAGAAGGTGG	CATTGTCATACCAGGAAATGAG
mmKrt18	AGGGCTCAGATCTTTGCGAAT	CATGGATGTCGCTCTCCACA
mmMap2	TCACAGAAGGAGTGGGCCTT	CTGAGGAACCTTAATTCTTGCCA
mmMyh3	AAATGAAGGGGACGCTGGAG	CAGCTGGAAGGTGACTCTGG
mmPparg	GCCTATGAGCACTTCACAAGAAAT	GGAATGCGAGTGGTCTTCCA
Flag-*mmProx1*	AAGGACGACGATGACAAGCC	CAGATGCTGTCCCTACCGTC
mmProx1 exon2	ACTCATCCGAGCAATCTGCC	GGAGCACCTAGTGGACTCTG
mmProx1-3'UTR	TCTCGGACGTGAAGTTCAACA	GCAACTTCCAGGAATCTCTCT
mmPrrx1	CAAGTTCCGCAGGAATGAGC	GGACGAGGTACGATGGGTTG
mmTnnc2	CAACCATGACGGACCAACAG	GTGTCTGCCCTAGCATCCTC
mmTnnt2	GTAGAGGACACCAAACCCAAG	GAGTCTGTAGCTCATTCAGGTC
mmTrf	GTGTGACGAGTGGAGCATCA	TCCGCTTCTCCGTTCACAAT
mmTtr	AAGTCCTGGATGCTGTCCGA	GCGGTCTTCCCAGAGGCAAA
mmTubb3	TTTTCGTCTCTAGCCGCGTG	GATGACCTCCCAGAACTTGGC

Table 4.2.3. Oligonucleotides for qPCR

for 5 min with 100 ng/mL DAPI diluted in 4% PFA in PBS. Finally, cells were washed with PBS and 1 mL of PBS per well (in a 24-well plate) was added before plates were sealed with Parafilm.

4.2.4 Fluorescence microscopy and quantification

15 random fluorescence micrographs per well were captured with a Nikon Ti2 microscope and Ti-HCS system, the Nikon S Plan Fluor ELWD 20x NA 0.45 objective, Nikon DS-Qi2 CMOS camera (2404x2404), and Lumencor Sola SE II light source.

To calculate the efficiency of neuronal induction, the total number of TUBB3-expressing cells with complex neurite outgrowth (cells with a round cell body and at least one thin process with a length at least double the diameter of the cell body) was counted manually. Any TUBB3-positive and Desmin-negative cells that did not meet the morphological criterion considered TUBB3-positive, non-neuronal cells. To calculate the efficiency of myocyte cell induction, the total number of Desmin-expressing cells was counted manually. Dual TUBB3- and Desmin-

positive cells were considered as a separate category of cells. To calculate the efficiency of hepatocyte reprogramming, the total number of cells for which TJP1 staining formed a complete border around the nucleus (as stained by DAPI) was counted manually.

Quantifications were based on the mean number of positive cells across 10 randomly selected 20x magnification fields of view per biological replicate, with at least 3 biological replicates. The number of reprogrammed cells in each treatment condition was then normalised to the number of reprogrammed cells in the control condition. To calculate the fraction of total reprogrammed cells positive for indicated markers and morphological criteria, the number of cells in each of the five categories was divided by the total number of cells present in any of the five categories, pooled across all replicates.

4.2.5 Enzyme-linked immunosorbent assay

Enzyme-linked immunosorbent assay (ELISA) for mouse Albumin was performed using the Mouse Albumin ELISA Kit (Bethyl Laboratories, Montgomery, TX, USA) according to the manufacturer's instructions.

4.3 *In Silico* Screen

4.3.1 Data sources

Using RNA-seq expression data and tissue and cell type annotations from Tabula Muris, we classified genes as specific to a cell type using Seurat `FindMarkers()`. The top 1000 genes belonging to one specific cell type were selected for further analysis.

To generate safeguard repressor scores for each TF in each cell type, median normalized CPM units for Tabula Muris scRNA-seq data was obtained through Seurat. Motif density was determined from mean numbers of TF-specific motifs in promoters of genes versus non-target genes. Motifs from CIS-BP build 1.94d (Lambert et al., 2018; Weirauch et al., 2014) were mapped to mm10 and motif fold change in the promoter region (±2 kb around the TSS) of cell type markers versus markers of other cell types was calculated.

4.3.2 Data analysis

For each TF, comparisons for all possible cell type-cell type pairings were performed. Genes common to both cell types in each pairwise comparison were removed. To summarise all motif fold-changes from these pairwise comparisons as a single value, the median of all motif fold-changes per TF-cell type pair was calculated.

Cell type-specific transcription factor expression, or $Z_{\text{expression}}$, was determined as a Z-score for each transcription factor across all cell types. The relative enrichment of transcription factor binding motifs in promoters of cell type-specific genes, or Z_{motif}, was calculated for each cell type. A higher $Z_{\text{expression}}$ indicates more specific expression of a transcription factor in the corresponding cell type relative to the other 17 cell types. A positive Z_{motif} predicts an activator, while a negative Z_{motif} predicts a repressor.

Z_{max} was then determined for $Z_{\text{expression}}$ and Z_{motif} for each cell type and transcription factor. Finally, a score for each transcription factor, summarising its likelihood of being a terminal repressor, was defined as the sum of $Z_{\text{expression}}$ and $-1 \cdot Z_{\text{motif}}$ following Z_{max}-normalisation. This Z_{max}-normalisation was intended to ensure equal weighting of expression or motif bias in the final score (fig. S1C). The top transcription factors per cell type, up to a maximum of six, were then shortlisted as highest-confidence candidates.

4.4 Genomics

4.4.1 Bulk RNA-seq

Library generation

Primary mouse liver tumor samples were dissected and placed into TRIzol (Invitrogen). Samples were crushed and then homogenized using QIAshredder (Qiagen). Reprogrammed MEF cells were harvested from pure cultures at day 7 or day 14 by addition of TRIzol to the cultures (Invitrogen). Wild-type hepatocytes were harvested from mouse livers at day 14 following hydrodynamic tail vein injection by perfusion with Liver Perfusion Medium (Gibco) and gradient centrifugation with Percoll (90% in PBS) as described (Charni-Natan & Goldstein, 2020).

Hepatocytes were fixed in 80% methanol and stored at -80 ℃. RNA extraction was performed by addition of TRIzol to the cells following three washes in PBS for rehydration and pelleting (Invitrogen). RNA harvested in TRIzol (Invitrogen) was isolated using the RNA Miniprep kit (Zymo Research). For bulk RNA sequencing, libraries were prepared according to the dUTP protocol (Levin et al., 2010) and paired-end sequencing (2x100bp) was performed on the NovaSeq 6000 platform (Illumina).

Data processing

Raw reads were mapped to the reference genome mm10 using STAR (Dobin et al., 2013). Differential expression was determined using DESeq2 (Love et al., 2014) (R package version 1.28.1) with size factor normalization and Wald significance tests. For MEF reprogramming data, we used primary MEF line as a covariate. EnhancedVolcano (version 1.8.0) (https://github.com/kevinblighe/EnhancedVolcano) was used to make volcano plots, and ComplexHeatmap (version 2.12.1) was used to generate heatmaps. Odds ratio and geneset overlap analyses were performed using the GeneOverlap R package (version 1.26.0) (Li, 2017). Sources for software packages are listed in Table 4.10.3 and Table 4.10.2.

4.4.2 CUT&RUN

Library preparation

Cells were harvested with accutase and strained through a 70 µm strainer. 100,000 cells were then washed twice with wash buffer (Table 4.4.1), then resuspended in wash buffer and added to Concanavalin-A beads (Table 4.8.1) pre-activated in ice-cold binding buffer (Table 4.4.1). The bead-cell suspension was then rotated (RT, 10 min). The supernatant was removed on magnet and beads were resuspended in antibody buffer (Table 4.4.1).

Primary antibody (mouse anti-FLAG-M2) or control (mouse IgG) was added to 1:100 v/v, and cells were incubated on a nutator (overnight, 4 ℃). Beads were washed twice in digitonin-wash

buffer (Table 4.4.1), resuspended in 700 ng/mL pAG-MNase (Protein Expression and Purification Core Facility, EMBL, Heidelberg) in digitonin-wash buffer, and rotated (1 h, 4 ℃). Beads were washed twice in digitonin-wash buffer, resuspended in digitonin-wash buffer, and placed on ice.

1 µL of 100 mM $CaCl_2$ was added to induce chromatin digestion. This mixture was incubated on ice (30 min). 50 µL of 2X STOP buffer (Table 4.4.1) was added. The suspension was incubated (10 mins, 37 ℃) to release chromatin fragments from cells. The supernatant was subjected to phenol-chloroform extraction. Purified DNA fragments were used for library preparation with NEBNext DNA Library Prep Kit for Illumina (NEB E7645). Libraries were sequenced (paired-end, 2x40 bp) on the NextSeq 550 platform (Illumina).

Buffer	Components
Antibody buffer	0.2 mM EDTA in Digitonin wash buffer
Binding buffer	20 mM HEPES-KOH pH 7.5, 10 mM KCl, 1 mM $CaCl_2$,1 mM $MnCl_2$ in water
Digitonin wash buffer	50 mg/mL Digitonin in CUT&RUN wash buffer
2X STOP buffer	340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% Digitonin, $50\,\mu g/mL$ RNase A, 50 μg/mL glycogen, $0.5\,ng/mL$ CUTANA E.coli spike-in DNA in water
Wash buffer	20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1X protease inhibitor in water

Table 4.4.1. CUT&RUN buffer compositions

Data processing

CUT&RUN data was analyzed using the nf-core/cutandrun pipeline v1.0.0 with Nextflow version 21.05.0. Reads were aligned to mm10. Software versions: bedtools (v2.30.0), bowtie2 (2.4.2), deeptools (v3.5.0), deseq2 (v1.28.0), fastqc (v0.11.9), multiqc (v1.11), picard (v2.23.9), python (v3.8.3), samtools (v1.10), SEACR (v1.3), TrimGalore (v0.6.6), ucsc (v377). Sources for software packages are listed in Table 4.10.3.

Consensus peaks were defined as any peaks that were present in at least 2 of the 3 replicates, where the corresponding peaks in each replicate overlapped by at least 50% of the length of the peaks. Consensus peaks were generated by running `bedmap --count --echo-map-range --fraction-both 0.5` and then merging adjacent peaks with bedtools merge after bed files for each replicate were merged and sorted.

To determine consensus CUT&RUN peaks containing motifs, I mapped the PROX1 binding motif (CIS-BP ID: M03445_2.00) onto mm10. I then extended each resulting peak to a total width of 50bp, and ran `bedtools intersect -wa` on the CUT&RUN consensus peaks and motif peaks, respectively. Genes defined as direct target genes based on CUT&RUN were determined by running `Homer annotatePeaks.pl` on the final peakset and filtering for genes with a peak within ±2000 bp of their TSS.

4.4.3 ATAC-seq

Sample preparation and tagmentation

Cells on culture plates were washed twice with PBS and detached by accutase digestion (4-6 min, RT). Cell suspensions were placed into an equal volume of MEF medium then centrifuged (500g, 5 min, 4 ℃) before resuspension in ice-cold PBS + 1% BSA. Cells were then strained through a 70 µm filter, centrifuged (500g, 5 min, 4 ℃), and 50,000 cells resuspended in 50 µL of ice-cold nuclei prep buffer containing 0.1% NP-40 and 0.01% digitonin in wash buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, and 1% BSA in ddH2O) for 3-5 minutes.

1 mL of ice-cold wash buffer was then added to stop cell lysis, and nuclei were centrifuged (500g, 10 min, 4 ℃). 4.5 µL of each of the tagmentation oligos, Tn5-ME and Tn5-R1N, were annealed in oligo annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, and 10 mM EDTA final concentration in ddH2O) by heating (3 min, 95 ℃) followed by a ramp down by 1 ℃/min to 25 ℃. The same procedure was performed for Tn5-ME and Tn5-R2N.

Tn5 was assembled with annealed oligos by combining 50 µL Tn5 (1 mg/mL stock) with 25 µL annealed Tn5-ME and Tn5-R1N and 25 µL Tn5-ME and Tn5-R2N (Table 4.4.2). Nuclei were resuspended in 40 µL tagmentation buffer (38.8 mM Tris-acetate, 77.6 mM K-acetate, 11.8 Mg-acetate, 18.8% dimethylformamide, and 0.12% NP-40 in ddH2O), to which 5 µL ice-cold PBS + 1% BSA and 5 µL pre-assembled Tn5 was added. Samples were incubated on a Thermomixer (500 rpm, 30 min, 37 ℃) before being subjected to MinElute (Qiagen) cleanup to extract tagmented DNA.

Oligo	Sequence	Description
Tn5-ME	[Phos]CTGTCTCTTATACACATCT	Used in oligoannealing/Tn5 assembly
Tn5-R1N	TCGTCGGCAGCGTCAGATGTGT ATAAGAGACAG	Used in oligoannealing/Tn5 assembly
Tn5-R2N	GTCTCGTGGGCTCGGAGATGTG TATAAGAGACAG	Used in oligoannealing/Tn5 assembly
P5	AATGATACGGCGACCACCGAGA TCTACACTCGTCGGCAGCGTC	Used in PCR amplification to add P5 sequencing adapter
P7-index00X	CAAGCAGAAGACGGCATACGAG AT[i7]GTCTCGTGGGCTCGG	Used in PCR amplification to add sample index (8 bp) and P7 sequencing adapter

Table 4.4.2. Oligonucleotides for ATAC-seq

Library preparation

Eluted DNA was pre-amplified with P5 and P7 primers using the NEBNext HF 2x PCR Master Mix (New England Biolabs) in a thermocycler set to 72 ℃ (5 min), 98 ℃ (30 sec), and 5 cycles of 98 ℃ (10 sec), 63 ℃ (30 sec), and 72 ℃ (1 min). A qPCR reaction was performed with the resulting pre-amplified libraries to determine the necessary additional cycles for complete amplification.

This was calculated as 5 cycles fewer than the number of cycles corresponding to 1/3 of max fluorescence. If the result was positive, pre-amplified libraries were amplified for the required number of additional cycles.

After finishing amplification, 50 µL of each library was subjected to 2-sided size selection. 27.5 µL (0.55x) AMPure XP beads was added to the libraries, followed by incubation (5 min), transfer of supernatant to new tubes, addition of 42.5 µL (1.4x) AMPure XP beads to the supernatant, incubation (5 min), three washes with 80% ethanol, and elution of DNA. Resulting libraries were then sequenced on the NextSeq 2000 platform (Illumina).

Data processing

ATAC-seq data was analysed with a custom Snakemake pipeline courtesy of Christian Arnold (EMBL). Raw reads were quality-checked with fastqc (v0.11.8), trimmed with trimmomatic (v0.38), and aligned to UCSC mm10 with bowtie2 (v2.3.4.3). Aligned reads were cleaned and bases were recalibrated (to take account of Tn5 insertion biases) with samtools (v1.10) and picard (v2.18.16). Reads were filtered with bedtools (v2.27.1), samtools, and picard. Peaks were called using macs2 (2.1.2) and coverage calculated with deeptools (v3.1.3). Final quality checks were performed with multiqc (v1.6).

Differential peak analysis was performed with DiffBind (v3.4.11) as described in the authors' vignette (same version). To determine the PROX1 footprint, 11 bp-wide motif peaks (CIS-BP ID: M03445_2.00, mapped on mm10) were intersected with CUT&RUN consensus peaks, respectively, using `bedtools intersect -wa`. This generated peaks present in the CUT&RUN consensus peakset which contain motifs and are centred on the motif. DiffTF (v1.8) was then run using ATAC-seq bam files and the motif-centred peakset (Berest et al., 2019). Sources for software packages are listed in Table 4.10.3.

4.4.4 Single-cell RNA-seq

Library generation

For single cell RNA-sequencing, reprogrammed hepatocytes on culture dishes were dissociated into single cells after two PBS washes by accutase digestion. Cells were lifted using cell scrapers, then suspensions were then passed through 70 µm filters into 9 mL of pre-warmed MEF medium. Cells were pelleted (800 RPM, 5 min) and resuspended in 1 mL HBSS + 0.04% BSA to > 3.5 million cells/mL, then fixed through addition of 4× volume of ice-cold methanol to a final concentration of 80% methanol. Cells were then stored at -20° C.

Barcoding oligonucleotides were designed based upon the ClickTag scheme (Gehring et al., 2020) and ordered with a 5' amine group label. A methyltetrazine group was conjugated to the oligos via the 5'-amine group. Membrane proteins on the fixed cells were conjugated to

an amine-trans-cyclooctene group. Subsequent incubation of oligos and cells according to the ClickTag protocol (Gehring et al., 2020) allowed chemical labeling of cells.

10,000 cells were multiplexed and loaded per GEM well in the Chromium Controller (10x Genomics), and the Chromium Single Cell 3' v2 reagent kit was used according to the manufacturer's instructions for gene expression library generation. We performed modifications at the cDNA and library preparation steps as suggested in the ClickTag protocol to generate barcoded oligonucleotide libraries in parallel. Gene expression and barcode libraries were then diluted to equimolar amounts, pooled at a 9 to 1 ratio, and 26 + 98 bp paired-end sequencing performed using NovaSeq 6000 (Illumina).

Data analysis

Single-cell RNA-seq data was analyzed using 10x Genomics Cell Ranger (v4.0.0) and Seurat (v4.0). Cells containing fewer than 1000 features, containing fewer than 2000 reads, or with mitochondrial genes comprising over 20% of genes, were discarded. Doublet cells were removed using Scrublet (Wolock et al., 2019) with a threshold of 0.35. Expression values of cell cycle genes were regressed out using `vars.to.regress` in `ScaleData()` to reduce heterogeneity caused by cycling cells.

Prox1-labeled cells were removed from the analysis if, based on PROX1 activity, they had a lower empirical probability of belonging to the Prox1-labeled than to the GFP-labeled population. This empirical probability was determined as the ratio of the number of Prox1- to GFP-labeled cells present at each PROX1 activity level, with a 1:1 ratio being the minimum threshold for removal. 4in1 activity was similarly inferred from expression of the 4in1 regulon. Based on 4in1 activity, cells in the 4in1 condition but with low 4in1 activity compared to MEFs were removed.

Regulon generation

For all TFs except PROX1, Dorothea (v1.7.2, all confidence levels) (Garcia-Alonso et al., 2019) was used to assemble their regulon. Genes were included in the PROX1 regulon if a CUT&RUN peak with an overlapping PROX1 motif was present within 1 kb of the gene's TSS, and if the gene was either downregulated in EnR vs DBD or upregulated in VP64 vs DBD.

Quantification of TF activity

4in1 activity was inferred from aggregate expression of all genes within the 4in1 regulon, calculated using Seurat `addModuleScore()`. PROX1 activity was calculated by taking the inverse of the aggregate expression, calculated with `addModuleScore()`, of a set of 48 high-confidence repressed PROX1 target genes. To generate this geneset, we filtered differentially-accessible ATAC-seq peaks at day 2 using a log2FC threshold (GFP vs *Prox1*) of > 1. We then selected only those peaks which overlapped with CUT&RUN peaks containing a PROX1 DNA binding motif.

If a peak meeting these criteria was found within 2 kb of a gene's TSS, that gene was included in the geneset. PROX1 activity scores were used for removal of Prox1-labeled cells with no *Prox1* expression, and for correlation of PROX1 activity with cell identity scores.

4.5 *In vivo* experiments

4.5.1 Hydrodynamic tail vein injection

For hydrodynamic tail vein injection (HDTVI), 2 mL of sterile 0.9% NaCl solution, corresponding to 10% of body weight, containing the plasmids of interest was injected into the tail vein of 8 week-old female C57BL/6N mice within 5 to 7 seconds. Depending on the vector used, this technique allowed for liver-specific gene knockouts and/or overexpression by in vivo transfection of hepatocytes.

Each mouse was injected with 20 µg of pX330-based plasmid for sgRNA- and shRNA-mediated gene knockdown, and 10 µg of pT3-EF1a-based plasmid for transposon-mediated gene overexpression. CMV-Sleeping Beauty transposase was co-injected to facilitate genomic integration of the transposon and was added in the injection cocktail at 20% of the total mass of pT3-EF1a plasmids injected. All HDTVI plasmids used can be found in Table 4.6.3. All injections in mice were performed by Kai Volz. Each experimental group involving HDTVI contained at least 5 mice.

Histology

Livers were embedded in paraffin and 2 µm slices were processed as previously described for immunohistochemistry (IHC) staining (Gallage et al., 2022). BOND-MAX (Leica Biosystems) was used to automate staining. BondTM citrate solution (AR9961, Leica), BondTM EDTA solution (AR9640, Leica), or BondTM proteolytic enzyme kit (AR9551, Leica) were used for antigen retrieval. Sections were incubated in antibodies diluted in BondTM primary antibody diluent (AR9352, Leica Biosystems) followed by secondary antibody (Leica Biosystems) incubation and staining with Bond Polymer Refine Detection Kit (DS9800, Leica Biosystems). Slides were scanned with an Aperio AT2 slide scanner (Leica Biosystems) at 20x, then annotated and analyzed with Aperio ImageScope (v12.4.0.5043, Leica) and Fiji (ImageJ v1.53q, National Institutes of Health USA).

4.6 Plasmids

Plasmids and cloning oligonucleotides were designed in Benchling. Vector backbones were digested with restriction enzymes for 4 h or overnight at 37 ℃. DNA constructs were generated by DNA synthesis (Sigma) or PCR amplification of cDNA with Q5 polymerase (NEB), using designed oligonucleotides, and according to manufacturer protocols. Linearised vectors were 5' dephosphorylated with Antarctic Phosphatase for 10 min at 37 ℃. Enzymes were inactivated by incubation for 20 minutes at 65 ℃. PCR-amplified inserts were digested in the same way as the vector backbone. Both were then run in a 2% agarose gel and extracted using the NucleoSpin Gel and PCR Clean-up Kit.

For generation of shRNA constructs, oligonucleotides were designed using pSicoOligoMaker 3.0. 10 µM of the forward and reverse oligo were mixed with 1x T4 ligation buffer and 0.5 µL T4 PNK and filled up with nuclease-free water to a final volume of 10 µL. The mix was incubated for 30 min at 37°C to be phosphorylated, 5 min at 95°C and then ramped down to 25°C at 5°C per minute for annealing.

Insertion of amplified constructs into linearised plasmids was performed by ligation into restriction-digested vectors using indicated enzymes and T4 DNA ligase for 20 min at room temperature or overnight at 16°C (Table 4.7.3). All lentiviral plasmids generated in this study can be found in Table 4.6.1. For shRNA cloning into the pSicoR backbone, 5'-phosphorylated DNA oligonucleotides were ordered from Sigma, annealed, and ligated with T4 DNA ligase into restriction-digested vectors. All shRNA plasmids generated in this study can be found in Ta-

Plasmid	Insert	Source or Insert Sequence
pMDLg/pRRE	Gag and Pol	Addgene #12259
pRSV-Rev	Rev	Addgene #12259
pCMV-VSV-G	VSVG	Addgene #8454
FUW-GFP-Cre	Cre recombinase	Pak et al., 2015
FUW-GFP-dCre	truncated Cre recombinase	Pak et al., 2015
FUW-rtTA	rtTA	Addgene #20342
4in1 (LV.4TF)	*Foxa3*, *Gata4*, *Hnf1a*, & *Hnf4a*	Song et al., 2016
tetO-mmMyoD1	mmMyoD1	Mall et al., 2017
tetO-mmAscl11	mmAscl1	Mall et al., 2017
tetO-GFP	GFP	Mall et al., 2017
tetO-*mmMyt1l*	*mmMyt1l*	Mall et al., 2017
tetO-*mmProx1*	*mmProx1*	this study
tetO-*mmPrrx1*	*mmPrrx1*	this study
tetO-*mmPparg*	*mmPparg*	this study

Table 4.6.1. Plasmids for lentiviral transduction

Plasmid	Insert	Source or Insert Sequence
pSicoR-PGK-Puro	Backbone only	Addgene #12084
pSicoR-PGK-Puro-mmProx1-shRNA_1	*mmProx1* shRNA_1	GGACCAAGATGTCATCACA
pSicoR-PGK-Puro-mmProx1-shRNA_2	*mmProx1* shRNA_2	GGAAGAAGGCCATTTACAA
pSicoR-PGK-Puro-mmPrrx1-shRNA_1	*mmPrrx1* shRNA_1	GAAACAGGACAACATTCAA
pSicoR-PGK-Puro-mmPrrx1-shRNA_2	*mmPrrx1* shRNA_2	GAGCCATGCTGGCCAATAA
pSicoR-PGK-Puro-mmPparg-shRNA_1	*mmPparg* shRNA_1	GCGGAGATCTCCAGTGATA
pSicoR-PGK-Puro-mmPparg-shRNA_2	*mmPparg* shRNA_2	GGCCTCCCTGATGAATAA

Table 4.6.2. Plasmids for shRNA-mediated gene knockdown

ble 4.6.2. All bacterial culture was performed using standard protocols, using reagents found in Table 4.6.4.

Bacterial culture

JM109 competent cells were used for plasmid transformation and expansion. Following thawing of bacteria on ice, 5 µl of the ligated plasmid was added to 50 µL of the cells and incubated for 20 min on ice. The mixture was subjected to heat shock for 1 min at 42°C, followed by 2 min on ice. 1 mL LB medium without antibiotics was then added to the mix under sterile conditions. This was then shaken at 750 rpm on a Thermomixer for 1 h at 37°C.

Bacteria were then centrifuged for 3 min at 3000 rpm, and 900 µL of the supernatant removed. Bacteria were then plated on an LB-ampicillin agar plate for incubation at 37°C overnight. The next day, single colonies were transferred into 2 mL of LB media with ampicillin (100 µg/µL)

Plasmid	Insert	Source or Insert Sequence
CMV-SB13	SB13 Transposase	N/A
pT3-EF1a-Myc	MYC	N/A
pX330 sgTp53	sgTp53	N/A
pT3-EF1a-mmProx1-IRES-GFP	*mmProx1*	N/A
miR-E shProx1	*mmProx1* shRNA_3	TGCTGTTGACAGTGAGCGCGAGGA CCAAGATGTCATCACATAGTGAAG CCACAGATGTATGTGATGACATCTT GGTCCTCATGCCTACTGCCTCGGA

Table 4.6.3. Plasmids for hydrodynamic tail vein injection

and incubated for 16 h at 37°C. Plasmid DNA was extracted from 1 mL of this bacterial culture with the GenElute Plasmid Miniprep Kit. Isolated constructs were test digested and run on an agarose gel, then sent for Sanger sequencing (Genewiz). The remaining 1 mL of bacterial culture was stored at 4°C until sequence-verified.

Bacterial cultures corresponding to verified plasmids were added to 200 mL LB medium with ampicillin (100 µg/µl) then incubated overnight at 37°C. Bacteria were then centrifuged, the supernatant discarded, and plasmids purified with the NucleoBond Xtra Midi Kit. For plasmids to be used in hydrodynamic tail-vein injection (see Section 4.5.1), endotoxins were removed using Zymo Research EndoZero II Spin-Columns. Plasmid concentrations were determined using a NanoDrop spectrophotometer.

Item	Source
Ampicillin	Merck (Darmstadt, Germany)
JM109 competent cells	Promega (Madison, USA)
LB Broth	Merck (Darmstadt, Germany)
LB-Agar Ampicillin	Carl Roth (Karlsruhe, Germany)

Table 4.6.4. Bacterial culture reagents

4.7 Antibodies and enzymes

All primary antibodies used in this study can be found in Tables 4.7.1 and 4.7.2. Secondary Alexa-conjugated antibodies for immunofluorescence were used at 1:2,000 and secondary IRDye-conjugated antibodies for Western blot were used at 1:10,000.

Target (host)	Supplier (cat. number)	Dilution (application)
A6 (rat)	Hybridoma Bank	1:50 (IHC)
Alb (goat)	Bethyl (A90-134A)	1:1,000 (WB)
B220 (rat)	BD Biosciences (553084)	1:3,000 (IHC)
β-Actin (mouse)	Sigma-Aldrich (A5441)	1:10,000 (WB)
Casp3 (rabbit)	Cell Signaling (9661)	1:300 (IHC)
CD3 (rabbit)	Invitrogen (MA1-90582)	1:500 (IHC)
CD8 (rat)	Invitrogen (14-0808-82)	1:200 (IHC)
CK19 (rat)	Hybridoma Bank	1:500 (IHC)
Clec4F (goat)	R&D (AF2784-SP)	1:1,000 (IHC)
Desmin (rabbit)	Abcam (ab32362)	1:500 (ICC), 1:1,000 (WB)
Ecadh (rabbit)	Cell Signaling Technology (24E10)	1:1,000 (WB)
F4/80 (rat)	Linaris (T2006)	1:250 (IHC)
FLAG M2 (mouse)	Sigma-Aldrich (F1804)	1:1,000 (ICC), 1:1,000 (WB)
GFP (rabbit)	Invitrogen (A11122)	1:250 (IHC)
HNF4 (rabbit)	Abcam (ab181604)	1:2,000 (IHC)
Ki-67 (rabbit)	Thermo Scientific (RM-9106-S1)	1:200 (IHC)
PROX1 (rabbit)	ReliaTech (RLT-102-PA32S)	1:500 (WB)
Sox9 (rabbit)	Abcam (ab185230)	1:2,000 (IHC)
TJP1 (rat)	Invitrogen (14-9776-80)	1:500 (ICC)
TUBB3 (mouse)	BioLegend (#801202)	1:1,000 (ICC), 1:1,000 (WB)

Table 4.7.1. Primary antibodies

Target (target species)	Supplier	Dilution
Alexa Fluor 488 Goat (anti-guinea pig)	Thermo Fisher Scientific (Carlsbad, USA)	1:2,000 (ICC)
Alexa Fluor 488 Donkey (anti-mouse)	Thermo Fisher Scientific (Carlsbad, USA)	1:2,000 (ICC)
Alexa Fluor 555 Donkey (anti-rabbit)	Thermo Fisher Scientific (Carlsbad, USA)	1:2,000 (ICC)
Alexa Fluor 647 Donkey (anti-rat)	Thermo Fisher Scientific (Carlsbad, USA)	1:2,000 (ICC)
IRDye 680RD Donkey (anti-mouse)	LI-COR Biosciences (Lincoln, USA)	1:10,000 (WB)
IRDye 800CW Donkey (anti-rabbit)	LI-COR Biosciences (Lincoln, USA)	1:10,000 (WB)

Table 4.7.2. Secondary antibodies

Enzyme	Supplier
Antarctic Phosphatase	New England Biolabs (Ipswich, USA))
Benzonase	Merck (Darmstadt, Germany)
LunaScript RT SuperMix	New England Biolabs (Ipswich, USA)
pAG-MNase	Protein Expression and Purification Core Facility, EMBL (Heidelberg, Germany)
Restriction enzymes	New England Biolabs (Ipswich, USA))
RNAse A	Thermo Fisher Scientific (Carlsbad, USA)
T4 DNA Ligase	New England Biolabs (Ipswich, USA))

Table 4.7.3. Molecular biology enzymes

4.8 Reagents

Buffer	Components
1 kb plus DNA ladder	Thermo Fisher Scientific (Carlsbad, USA)
Agarose	Carl Roth (Karlsruhe, Germany)
BSA	Merck (Darmstadt, Germany)
Concanavalin-A beads	Polysciences (Pennsylvania, USA)
CUTANA E. coli Spike-in DNA	Biocat (Heidelberg, Germany)
DAPI	Thermo Fisher Scientific (Carlsbad, USA)
dNTP mix, 10 mM	Thermo Fisher Scientific (Carlsbad, USA)
Gel Loading Dye, Purple (6X)	New England Biolabs (Ipswich, USA)
Nuclease-free water	Qiagen (Hilden, Germany)
Power SYBR Green PCR Master Mix	Thermo Fisher Scientific (Carlsbad, USA)
Protease inhibitor	Merck (Darmstadt, Germany)
Protein A-coupled beads	Thermo Fisher Scientific (Carlsbad, USA)
QuickExtract DNA Extraction Solution	Lucigen (Middleton, USA)
RNAse A	Thermo Fisher Scientific (Carlsbad, USA)
Roti®-GelStain Red	Carl Roth (Karlsruhe, Germany)

Table 4.8.1. Commercial molecular biology reagents

Buffer	Components
Lysis buffer (IP)	0.5% Tween-20, 50 mM Tris pH 7.5, 2 mM EDTA, 1 mM DTT, 1 mM PMSF, 5 mM NaF, complete protease inhibitor in water
NP40 Lysis buffer (IP)	0.5% NP-40, 50 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1 mM DTT, 1 mM PMSF, 5 mM NaF, 5 µg/ml cytochalasin B, complete protease inhibitor, benzonase in water
50xTAE buffer	2 M tris base, 50 mM EDTA pH 8.0, 5.75% (v/v) acetic acid in water

Table 4.8.2. General molecular biology buffer compositions

Reagent	Source
Acetic acid	Honeywell (Charlotte, USA)
Cytochalasin B	Merck (Darmstadt, Germany)
Digitonin	Merck (Darmstadt, Germany)
Dithiothreitol	Merck (Darmstadt, Germany)
EDTA	GERBU Biotechnik (Heidelberg, Germany)
EGTA	Merck (Darmstadt, Germany)
Ethanol (absolute)	VWR International (Radnor, USA)
Glacial Acetic Acid	Thermo Fisher Scientific (Carlsbad, USA)
Glycerol	Merck (Darmstadt, Germany)
Glycogen	Merck (Darmstadt, Germany)
Isopropanol	Merck (Darmstadt, Germany)
Methanol	VWR International (Radnor, USA)
IGEPAL CA630 (AKA NP-40)	Merck (Darmstadt, Germany)
Paraformaldehyde, 4%	Boster Biological Technology (Pleasanton, USA)
PMSF	Merck (Darmstadt, Germany)
Spermidine	MP Biomedicals (Eschwege, Germany)
Sodium dodecyl sulphate	GERBU Biotechnik (Heidelberg, Germany)
Trisma base (Tris base)	Honeywell (Charlotte, USA)
Triton X-100	Merck (Darmstadt, Germany)
TRIzol	Thermo Fisher Scientific (Carlsbad, USA)
Tween-20	Merck (Darmstadt, Germany)

Table 4.8.3. Organic buffer components

Reagent	Source
Calcium chloride ($CaCl_2$)	Carl Roth (Karlsruhe, Germany)
HEPES, pH 7.2	Rockland (Pennsylvania, USA)
Hydrochloric acid (HCl)	Thermo Fisher Scientific (Carlsbad, USA)
Manganese(II)-chloride ($MnCl_2$)	Honeywell (Charlotte, USA)
NaF	Merck (Darmstadt, Germany)
Nuclease-free water	Qiagen (Hilden, Germany)
NuPAGE MOPS buffer (20x)	Thermo Fisher Scientific (Carlsbad, USA)
NuPAGE Transfer buffer (20x)	Thermo Fisher Scientific (Carlsbad, USA)
Phosphate-buffered saline pellets	Th. Geyer (Renningen, Germany)
Potassium chloride (KCl)	Merck (Darmstadt, Germany)
Potassium hydroxide (KOH)	Merck (Darmstadt, Germany)
Sodium bicarbonate ($NaHCO_3$)	Santa Cruz Biotechnology (Dallas, USA)
Sodium chloride (NaCl)	Merck (Darmstadt, Germany)
Sodium hydroxide solution (NaOH)	Thermo Fisher Scientific (Carlsbad, USA)

Table 4.8.4. Inorganic buffer components

4.9 Kits and consumables

Kit	Supplier
Direct-zol RNA MiniPrep	Zymo Research (Irvine, USA)
GenElute Plasmid Miniprep Kit	Macherey-Nagel (Düren, Germany)
NEBNext Ultra II DNA Library Prep Kit	New England Biolabs (Ipswich, USA)
NucleoSpin Gel and PCR Clean-up Kit	Macherey-Nagel (Düren, Germany)
NucleoBond Xtra Midi Kit	Macherey-Nagel (Düren, Germany)
Phusion High-Fidelity PCR Kit	Thermo Fisher Scientific (Carlsbad, USA)
QiaPrep Spin Miniprep Kit	Qiagen (Hilden, Germany)

Table 4.9.1. Commercial kits

Software	Source
6-well tissue culture plates	Thermo Fisher Scientific (Carlsbad, USA)
12-well tissue culture plates	Corning Inc. (Corning, USA)
24-wel tissue culture plates	Corning Inc. (Corning, USA)
384-well qPCR plate	Biozym (Hessisch Oldendorf, Germany)
Bacterial culture tubes	VWR International (Radnor, PA, USA)
Cell culture dish (10 cm)	Thermo Fisher Scientific (Carlsbad, USA)
Cell culture dish (15 cm)	Greiner Bio-One (Kremsmünster, Austria)
Cell lifters	Santa Cruz Biotechnology (Dallas, USA)
Conical centrifuge tubes (15 mL, 50 mL)	Corning (Corning, USA)
Countess Cell Counting Chamber Slides	Thermo Fisher Scientific (Carlsbad, USA)
Microcentrifuge tubes (1.5 mL)	Thermo Fisher Scientific (Carlsbad, USA)
Microcentrifuge tubes (2.0 mL)	Eppendorf (Hamburg, Germany)
Millex syringe filters (45 µm)	Merck (Darmstadt, Germany)
Parafilm	Merck (Darmstadt, Germany)
PCR tubes	Merck (Darmstadt, Germany)
Petri dishes	Greiner Bio-One (Kremsmünster, Austria)
Pipette tips (10, 20, 200, and 1000 µL)	Mettler Toledo (Columbus, USA)
Pipette tips (filter) (10, 20, 200, and 1000 µL)	Starlab (Hamburg, Germany)
qPCR Seal Sheets	Biozym (Hessisch Oldendorf, Germany)
Serological pipettes (5, 10, 25, and 50 mL)	Greiner Bio-One (Kremsmünster, Austria)
Syringe (30 mL)	TH. Gyer (Renningen, Germany)

Table 4.9.2. Plasticware

4.10 Devices and Software

Device	Manufacturer
BioRad T100 Thermal Cycler	BioRad (Hercules, USA)
Centrifuge S424	Eppendorf (Hamburg, Germany)
Contrast microscope BX51WI	Olympus-Lifescience (Tokyo, Japan)
Countess cell counter	Thermo Fisher Scientific (Carlsbad, USA)
DM IL LED Fluorescence Microscope	Leica (Oberkochen, Germany)
EasyPhor gel electrophoresis chamber	Biozym (Hessisch Oldendorf, Germany)
Electrophoresis power supply	BioRad (Hercules, USA)
Fusion Orbitrap	Thermo Fisher Scientific (Carlsbad, USA)
GelDoc XR+ Molecular Imager	BioRad (Hercules, USA)
HERA cell 240i CO2 incubator	Thermo Fisher Scientific (Carlsbad, USA)
Mini Gel Tank	Thermo Fisher Scientific (Carlsbad, USA)
NanoDrop One	Thermo Fisher Scientific (Carlsbad, USA)
neoBlock 1 heatblock	neoLab (Heidelberg, Germany)
NextSeq 550	Illumina (San Diego, USA)
NextSeq 2000	Illumina (San Diego, USA)
NovaSeq 6000	Illumina (San Diego, USA)
ODYSSEY CLx Imaging System	LI-COR Biosciences (Lincoln, USA)
QuantStudio 5	Thermo Fisher Scientific (Carlsbad, USA)
Rotanta 460R centrifuge	Hettich (Kirchlengern, Germany)
Thermomixer compact	Eppendorf (Hamburg, Germany)
Vortex Genie 2	Scientific Industries (Bohemia, USA)

Table 4.10.1. Devices

Software	Developer or Citation
ComplexHeatmap	Gu et al., 2016
deseq2	Love et al., 2014
DiffBind	Stark and Brown, 2011
EnhancedVolcano	Blighe, 2023
GeneOverlap	Li, 2017
R (v4.2.0)	RStudio (Boston, USA)
Seurat (v4.0)	Hao et al., 2021

Table 4.10.2. Software (R and R libraries)

Software	Developer or Citation
bedtools	Quinlan and Hall, 2010
bowtie2	Langmead and Salzberg, 2012
CellRanger (v4.0.0)	10x Genomics
deeptools	Ramírez et al., 2016
deseq2	Love et al., 2014
DiffTF	Berest et al., 2019
fastqc	Simon Andrews (Babraham Institute)
macs2	Zhang et al., 2008
multiqc	Ewels et al., 2016
nf-core/cutandrun	Cheshire et al., 2022
picard	Broad Institute, 2019
python	Python Software Foundation
samtools	Danecek et al., 2021
Scrublet	Wolock et al., 2019
SEACR	Meers et al., 2019
STAR	Dobin et al., 2013
TrimGalore	Krueger et al., 2021
trimmomatic	Bolger et al., 2014

Table 4.10.3. Software (command line)

Software	Developer
Benchling	Benchling Inc. (San Francisco, CA, USA)
BioRender	BioRender.com. (Toronto, Canada)
Fiji ImageJ (v1.53t)	National Institutes of Health (Maryland, USA)
GraphPad Prism (v9.2.0)	GraphPad Software (San Diego, USA)
Image Lab (v5.2.1)	BioRad (Hercules, USA)
Image Studio Lite (v5.2.5)	LI-COR Biosciences (Lincoln, USA)
Integrative Genomics Viewer (v2.8.2)	Broad Institute (Cambridge, USA)
Leica Application Suite X (v3.7.1)	Leica (Wetzlar, Germany)
Microsoft Office	Microsoft Corporation (Redmond, USA)
pSicoOligoMaker (v3.0)	Andrea Ventura
QuantStudio Design and Analysis software (v1.3.1)	Thermo Fisher Scientific (Carlsbad, USA)
Zotero	Corporation for Digital Scholarship (Fairfax, USA)

Table 4.10.4. Software (GUI)

4.11 Statistics

Data are presented as mean ± SD unless stated otherwise. No statistical methods were used to predetermine sample size. The exact number of technical and biological replicates, and the statistical tests used, are indicated in figure legends. The biological replicates for RNA-sequencing data are listed.

Primary MEF cultures were randomly prepared. Other experiments were not randomised. IF quantification was performed following blinding to the overexpression conditions of the cells. For all other experiments and subsequent analyses, I was not blinded.

Statistical analysis was performed in R, GraphPad Prism, and with command line tools. Statistical testing was performed using t-test, Dunnett's test, or analysis of variance (ANOVA) test with correction for multiple testing depending on study design.

References

Adrian-Segarra, J. M., Weigel, B., & Mall, M. (2021). Isolation and Neuronal Reprogramming of Mouse Embryonic Fibroblasts. In H. Ahlenius (Ed.), *Neural Reprogramming* (pp. 1–12, Vol. 2352). Springer US.

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and Methylation of Histones and their Possible Role in the Regulation of RNA Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(5), 786–794.

Aranguren, X. L., Beerens, M., Coppiello, G., Wiese, C., Vandersmissen, I., Nigro, A. L., Verfaillie, C. M., Gessler, M., & Luttun, A. (2013). COUP-TFII orchestrates venous and lymphatic endothelial identity by homo- or hetero-dimerisation with PROX1. *Journal of Cell Science*, *126*(5), 1164–1175. https://doi.org/10.1242/jcs.116293

Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D., & Wagner, G. P. (2016). The origin and evolution of cell types. *Nature Reviews Genetics*, *17*(12), 744–757. https://doi.org/10.1038/nrg.2016.127

Armour, S. M., Remsberg, J. R., Damle, M., Sidoli, S., Ho, W. Y., Li, Z., Garcia, B. A., & Lazar, M. A. (2017). An HDAC3-PROX1 corepressor module acts on HNF4α to control hepatic triglycerides. *Nature Communications*, *8*(1), 549. https://doi.org/10.1038/s41467-017-00772-5

Baker, N. E. (2001). Master regulatory genes; telling them what to do. *BioEssays*, *23*(9), 763–766. https://doi.org/10.1002/bies.1110

Balsalobre, A., & Drouin, J. (2022). Pioneer factors as master regulators of the epigenome and cell fate. *Nature Reviews Molecular Cell Biology*, *23*(7), 449–464. https://doi.org/10.1038/s41580-022-00464-z

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2 Pt 1), 299–308. https://doi.org/10.1016/0092-8674(81)90413-x

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Benevolenskaya, E. V. (2007). Histone H3K4 demethylases are essential in development and differentiation. *Biochemistry and Cell Biology*, *85*(4), 435–443. https://doi.org/10.1139/O07-057

Berasain, C., & Avila, M. A. (2015). Regulation of hepatocyte identity and quiescence. *Cellular and Molecular Life Sciences*, *72*(20), 3831–3851. https://doi.org/10.1007/s00018-015-1970-7

Berest, I., Arnold, C., Reyes-Palomares, A., Palla, G., Rasmussen, K. D., Giles, H., Bruch, P.-M., Huber, W., Dietrich, S., Helin, K., & Zaugg, J. B. (2019). Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Reports*, *29*(10), 3147–3159.e12. https://doi.org/10.1016/j.celrep.2019.10.106

Blighe, K. (2023, May). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., & Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, *441*(7091), 349–353. https://doi.org/10.1038/nature04733

Britten, R. J., & Davidson, E. H. (1969). Gene Regulation for Higher Cells: A Theory: New facts regarding the organization of the genome provide clues to the nature of gene regulation. *Science*, *165*(3891), 349–357. https://doi.org/10.1126/science.165.3891.349

Bulger, M., & Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell*, *144*(3), 327–339. https://doi.org/10.1016/j.cell.2011.01.024

Buyl, K., De Kock, J., Bolleyn, J., Rogiers, V., & Vanhaecke, T. (2015). Measurement of Albumin Secretion as Functionality Test in Primary Hepatocyte Cultures. *Methods in Molecular Biology (Clifton, N.J.)*, *1250*, 303–308. https://doi.org/10.1007/978-1-4939-2074-7_22

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., Mazin, P. V., Xenarios, I., Harshman, K., Mort, M., Cooper, D. N., Sandi, C., Soares, M. J., Ferreira, P. G., Afonso, S., Carneiro, M., Turner, J. M. A., VandeBerg, J. L., Fallahshahroudi, A., Jensen, P., Behr, R., Lisgo, S., Lindsay, S., Khaitovich, P., Huber, W., Baker, J., Anders, S., Zhang, Y. E., & Kaessmann, H. (2019). Gene expression across mammalian organ development. *Nature*, *571*(7766), 505–509. https://doi.org/10.1038/s41586-019-1338-5

Chaisaingmongkol, J., Budhu, A., Dang, H., Rabibhadana, S., Pupacdi, B., Kwon, S. M., Forgues, M., Pomyen, Y., Bhudhisawasdi, V., Lertprasertsuke, N., Chotirosniramit, A., Pairojkul, C., Auewarakul, C. U., Sricharunrat, T., Phornphutkul, K., Sangrajrang, S., Cam, M., He, P., Hewitt, S. M., Ylaya, K., Wu, X., Andersen, J. B., Thorgeirsson, S. S., Waterfall, J. J., Zhu, Y. J., Walling, J., Stevenson, H. S., Edelman, D., Meltzer, P. S., Loffredo, C. A., Hama, N., Shibata, T., Wiltrout, R. H., Harris, C. C., Mahidol, C., Ruchirawat, M., Wang, X. W., & TIGER-LC Consortium. (2017). Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. *Cancer Cell*, *32*(1), 57–70.e3. https://doi.org/10.1016/j.ccell.2017.05.009

Chan, S. S.-K., & Kyba, M. (2013). What is a Master Regulator? *Journal of stem cell research & therapy*, *3*, 114. https://doi.org/10.4172/2157-7633.1000e114

Chanda, S., Ang, C. E., Davila, J., Pak, C., Mall, M., Lee, Q. Y., Ahlenius, H., Jung, S. W., Südhof, T. C., & Wernig, M. (2014). Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. *Stem Cell Reports*, *3*(2), 282–296. https://doi.org/10.1016/j.stemcr.2014.05.020

Chang, T.-M., & Hung, W.-C. (2013). The homeobox transcription factor Prox1 inhibits proliferation of hepatocellular carcinoma cells by inducing p53-dependent senescence-like phenotype. *Cancer Biology & Therapy*, *14*(3), 222–229. https://doi.org/10.4161/cbt.23293

Charni-Natan, M., & Goldstein, I. (2020). Protocol for Primary Mouse Hepatocyte Isolation. *STAR Protocols*, *1*(2), 100086. https://doi.org/10.1016/j.xpro.2020.100086

Chen, J., Lambo, M. E., Ge, X., Dearborn, J. T., Liu, Y., McCullough, K. B., Swift, R. G., Tabachnick, D. R., Tian, L., Noguchi, K., Garbow, J. R., Constantino, J. N., Gabel, H. W., Hengen, K. B., Maloney, S. E., & Dougherty, J. D. (2021). A MYT1L syndrome mouse model recapitulates patient phenotypes and reveals altered brain development due to disrupted neuronal maturation. *Neuron*, *109*(23), 3775–3792.e14. https://doi.org/10.1016/j.neuron.2021.09.009

Cheshire, C., Charlotte-West, Bot, N.-C., Ladd, D., Fields, C., Patel, H., Deu-Pons, J., Ewels, P., & Menden, K. (2022, June). Nf-core/cutandrun: Nf-core/cutandrun v2.0 Copper Cobra.

Cobaleda, C., Jochum, W., & Busslinger, M. (2007). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature*, *449*(7161), 473–477. https://doi.org/10.1038/nature06159

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, *51*(6), 987–1000. https://doi.org/10.1016/0092-8674(87)90585-X

De Paepe, C., Krivega, M., Cauffman, G., Geens, M., & Van de Velde, H. (2014). Totipotency and lineage segregation in the human embryo. *Molecular Human Reproduction*, *20*(7), 599–618. https://doi.org/10.1093/molehr/gau027

de Wit, E., & de Laat, W. (2012). A decade of 3C technologies: Insights into nuclear organization. *Genes & Development*, *26*(1), 11–24. https://doi.org/10.1101/gad.179804.111

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., & Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends in genetics: TIG*, *23*(12), 614–622. https://doi.org/10.1016/j.tig.2007.09.001

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dorigo, B., Schalch, T., Bystricky, K., & Richmond, T. J. (2003). Chromatin fiber folding: Requirement for the histone H4 N-terminal tail. *Journal of Molecular Biology*, *327*(1), 85–96. https://doi.org/10.1016/s0022-2836(03)00025-1

Drouin, J. (2014). Minireview: Pioneer Transcription Factors in Cell Fate Specification. *Molecular Endocrinology*, *28*(7), 989–998. https://doi.org/10.1210/me.2014-1084

Du, Y., Wang, J., Jia, J., Song, N., Xiang, C., Xu, J., Hou, Z., Su, X., Liu, B., Jiang, T., Zhao, D., Sun, Y., Shu, J., Guo, Q., Yin, M., Sun, D., Lu, S., Shi, Y., & Deng, H. (2014). Human Hepatocytes with Drug Metabolic Function Induced from Fibroblasts by Lineage Reprogramming. *Cell Stem Cell*, *14*(3), 394–403. https://doi.org/10.1016/j.stem.2014.01.008

Dubois, V., Staels, B., Lefebvre, P., Verzi, M., & Eeckhoute, J. (2020). Control of Cell Identity by the Nuclear Receptor HNF4 in Organ Pathophysiology. *Cells*, *9*(10), 2185. https://doi.org/10.3390/cells9102185

Dudas, J., Mansuroglu, T., Moriconi, F., Haller, F., Wilting, J., Lorf, T., Füzesi, L., & Ramadori, G. (2008). Altered regulation of Prox1-gene-expression in liver tumors. *BMC Cancer*, *8*(1), 92–92. https://doi.org/10.1186/1471-2407-8-92

Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., & Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *Journal of Virology*, *72*(11), 8463–8471. https://doi.org/10.1128/JVI.72.11.8463-8471.1998

Elsir, T., Smits, A., Lindström, M. S., & Nistér, M. (2012). Transcription factor PROX1: Its role in development and cancer. *Cancer and Metastasis Reviews*, *31*(3-4), 793–805. https://doi.org/10.1007/s10555-012-9390-8

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fan, B., Malato, Y., Calvisi, D. F., Naqvi, S., Razumilava, N., Ribback, S., Gores, G. J., Dombrowski, F., Evert, M., Chen, X., & Willenbring, H. (2012). Cholangiocarcinomas can originate from hepatocytes in mice. *Journal of Clinical Investigation*, *122*(8), 2911–2915. https://doi.org/10.1172/JCI63212

Farazi, P. A., & DePinho, R. A. (2006). Hepatocellular carcinoma pathogenesis: From genes to environment. *Nature Reviews Cancer*, *6*(9), 674–687. https://doi.org/10.1038/nrc1934

Fong, A. P., & Tapscott, S. J. (2013). Skeletal muscle programming and re-programming. *Current Opinion in Genetics & Development*, *23*(5), 568–573. https://doi.org/10.1016/j.gde.2013.05.002

Franzén, O., Gan, L.-M., & Björkegren, J. L. M. (2019). PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, *2019*. https://doi.org/10.1093/database/baz046

Fuda, N. J., Ardehali, M. B., & Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, *461*(7261), 186–192. https://doi.org/10.1038/nature08449

Gallage, S., Ali, A., Barragan Avila, J. E., Herebian, D., Karimi, M. M., Irvine, E. E., McHugh, D., Schneider, A. T., Vucur, M., Keitel, V., Gil, J., Withers, D. J., Luedde, T., & Heikenwalder, M. (2022). Spontaneous Cholemia in C57BL/6 Mice Predisposes to Liver Cancer in NASH. *Cellular and Molecular Gastroenterology and Hepatology*, *13*(3), 875–878. https://doi.org/10.1016/j.jcmgh.2021.11.012

Galle, P. R., Forner, A., Llovet, J. M., Mazzaferro, V., Piscaglia, F., Raoul, J.-L., Schirmacher, P., & Vilgrain, V. (2018). EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *Journal of Hepatology*, *69*(1), 182–236. https://doi.org/10.1016/j.jhep.2018.03.019

Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., & Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Research*, *29*(8), 1363–1375. https://doi.org/10.1101/gr.240663.118

Gehring, J., Hwee Park, J., Chen, S., Thomson, M., & Pachter, L. (2020). Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. *Nature Biotechnology*, *38*(1), 35–38. https://doi.org/10.1038/s41587-019-0372-z

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S., & Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research*, *17*(6), 669–681. https://doi.org/10.1101/gr.6339607

Graf, T. (2011). Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell*, *9*(6), 504–516. https://doi.org/10.1016/j.stem.2011.11.012

Gramolelli, S., Cheng, J., Martinez-Corral, I., Vähä-Koskela, M., Elbasani, E., Kaivanto, E., Rantanen, V., Tuohinto, K., Hautaniemi, S., Bower, M., Haglund, C., Alitalo, K., Mäkinen, T., Petrova, T. V., Lehti, K., & Ojala, P. M. (2018). PROX1 is a transcriptional regulator of MMP14. *Scientific Reports*, *8*(1), 9531. https://doi.org/10.1038/s41598-018-27739-w

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, *32*(18), 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Gurdon, J. B. (1960). The developmental capacity of nuclei taken from differentiating endoderm cells of Xenopus laevis. *Journal of Embryology and Experimental Morphology*, *8*, 505–526.

Gurdon, J. B. (1962a). Adult frogs derived from the nuclei of single somatic cells. *Developmental Biology*, *4*, 256–273. https://doi.org/10.1016/0012-1606(62)90043-x

Gurdon, J. B. (1962b). The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. *Development*, *10*(4), 622–640. https://doi.org/10.1242/dev.10.4.622

Gurdon, J. B., & Uehlinger, V. (1966). "Fertile" Intestine Nuclei. *Nature*, *210*(5042), 1240–1241. https://doi.org/10.1038/2101240a0

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews. Molecular Cell Biology*, *19*(10), 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hansen, J. L., Loell, K. J., & Cohen, B. A. (2022). A test of the pioneer factor hypothesis using ectopic liver gene activation. *eLife*, *11*, e73358. https://doi.org/10.7554/eLife.73358

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., McElrath, J. M., Blish, C. A., Gottardo, R., Smibert, P., & Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Hobert, O. (2008). Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. *Proceedings of the National Academy of Sciences*, *105*(51), 20067–20071. https://doi.org/10.1073/pnas.0806070105

Hobert, O. (2010). Neurogenesis in the nematode Caenorhabditis elegans. *WormBook*. https://doi.org/10.1895/wormbook.1.12.2

Hobert, O. (2016). Terminal Selectors of Neuronal Identity. In *Current Topics in Developmental Biology* (pp. 455–475, Vol. 116). Elsevier.

Hobert, O., & Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. *Current Opinion in Neurobiology*, *56*, 97–105. https://doi.org/10.1016/j.conb.2018.12.006

Högström, J., Heino, S., Kallio, P., Lähde, M., Leppänen, V.-M., Balboa, D., Wiener, Z., & Alitalo, K. (2018). Transcription factor PROX1 suppresses Notch pathway activation via the nucleosome remodeling and deacetylase complex in colorectal cancer stem-like cells. *Cancer Research*, *78*(20), canres.0451.2018. https://doi.org/10.1158/0008-5472.can-18-0451

Hota, S. K., Rao, K. S., Blair, A. P., Khalilimeybodi, A., Hu, K. M., Thomas, R., So, K., Kameswaran, V., Xu, J., Polacco, B. J., Desai, R. V., Chatterjee, N., Hsu, A., Muncie, J. M., Blotnick, A. M., Winchester, S. A. B., Weinberger, L. S., Hüttenhain, R., Kathiriya, I. S., Krogan, N. J., Saucerman, J. J., & Bruneau, B. G. (2022). Brahma safeguards canalization of cardiac mesoderm differentiation. *Nature*, *602*(7895), 129–134. https://doi.org/10.1038/s41586-021-04336-y

Hu, J., Ho, A. L., Yuan, L., Hu, B., Hua, S., Hwang, S. S., Zhang, J., Hu, T., Zheng, H., Gan, B., Wu, G., Wang, Y. A., Chin, L., & DePinho, R. A. (2013). Neutralization of terminal differentiation in gliomagenesis. *Proceedings of the National Academy of Sciences*, *110*(36), 14520–14527. https://doi.org/10.1073/pnas.1308610110

Hyun, J., Oh, S.-H., Premont, R. T., Guy, C. D., Berg, C. L., & Diehl, A. M. (2019). Dysregulated activation of fetal liver programme in acute liver failure. *Gut*, *68*(6), 1076–1087. https://doi.org/10.1136/gutjnl-2018-317603

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., Lee, D., Kaestner, K. H., & Zaret, K. S. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Molecular cell*, *62*(1), 79–91. https://doi.org/10.1016/j.molcel.2016.03.001

Iwano, T., Masuda, A., Kiyonari, H., Enomoto, H., & Matsuzaki, F. (2012). Prox1 postmitotically defines dentate gyrus cells by specifying granule cell identity over CA3 pyramidal cell fate in the hippocampus. *Development*, *139*(16), 3051–3062. https://doi.org/10.1242/dev.080002

Jiang, M., Azevedo-Pouly, A. C., Deering, T. G., Hoang, C. Q., DiRenzo, D., Hess, D. A., Konieczny, S. F., Swift, G. H., & MacDonald, R. J. (2016). MIST1 and PTF1 Collaborate in Feed-Forward Regulatory Loops That Maintain the Pancreatic Acinar Phenotype in Adult Mice. *Molecular and Cellular Biology*, *36*(23), 2945–2955. https://doi.org/10.1128/MCB.00370-16

Jimenez, M. A., Akerblad, P., Sigvardsson, M., & Rosen, E. D. (2007). Critical role for Ebf1 and Ebf2 in the adipogenic transcriptional cascade. *Molecular and Cellular Biology*, *27*(2), 743–757. https://doi.org/10.1128/MCB.01557-06

Johnson, N. C., Dillard, M. E., Baluk, P., McDonald, D. M., Harvey, N. L., Frase, S. L., & Oliver, G. (2008). Lymphatic endothelial cell identity is reversible and its maintenance requires Prox1 activity. *Genes & Development*, *22*(23), 3282–3291. https://doi.org/10.1101/gad.1727208

Kalra, A., Yetiskul, E., Wehrle, C. J., & Tuma, F. (2023). Physiology, Liver. In *StatPearls*. StatPearls Publishing.

Kamal, A., Arnold, C., Claringbould, A., Moussa, R., Servaas, N. H., Kholmatov, M., Daga, N., Nogina, D., Mueller-Dott, S., Reyes-Palomares, A., Palla, G., Sigalova, O., Bunina, D., Pabst, C., & Zaugg, J. B. (2023). GRaNIE and GRaNPA: Inference and evaluation of enhancer-mediated gene regulatory networks. *Molecular Systems Biology*, e11627. https://doi.org/10.15252/msb.202311627

Karalay, Ö., Doberauer, K., Vadodaria, K. C., Knobloch, M., Berti, L., Miquelajauregui, A., Schwark, M., Jagasia, R., Taketo, M. M., Tarabykin, V., Lie, D. C., & Jessberger, S. (2011). Prospero-related homeobox 1 gene (Prox1) is regulated by canonical Wnt signaling and has a stage-specific role in adult hippocampal neurogenesis. *Proceedings of the National Academy of Sciences*, *108*(14), 5807–5812. https://doi.org/10.1073/pnas.1013456108

Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, *9*(10), 770–780. https://doi.org/10.1038/nrm2503

Kaufman, C. K., Mosimann, C., Fan, Z. P., Yang, S., Thomas, A. J., Ablain, J., Tan, J. L., Fogley, R. D., van Rooijen, E., Hagedorn, E. J., Ciarlo, C., White, R. M., Matos, D. A., Puller, A.-C., Santoriello, C., Liao, E. C., Young, R. A., & Zon, L. I. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science*, *351*(6272), aad2197. https://doi.org/10.1126/science.aad2197

Kiecker, C., Bates, T., & Bell, E. (2016). Molecular specification of germ layers in vertebrate embryos. *Cellular and Molecular Life Sciences*, *73*(5), 923–947. https://doi.org/10.1007/s00018-015-2092-y

Kim, J., Lu, C., Srinivasan, S., Awe, S., Brehm, A., & Fuller, M. T. (2017). Blocking promiscuous activation at cryptic promoters directs cell type–specific gene expression. *Science*, *356*(6339), 717–721. https://doi.org/10.1126/science.aal3096

Knezetic, J. A., & Luse, D. S. (1986). The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell*, *45*(1), 95–104. https://doi.org/10.1016/0092-8674(86)90541-6

Koch, C. M., Andrews, R. M., Flicek, P., Dillon, S. C., Karaöz, U., Clelland, G. K., Wilcox, S., Beare, D. M., Fowler, J. C., Couttet, P., James, K. D., Lefebvre, G. C., Bruce, A. W., Dovey, O. M., Ellis, P. D., Dhami, P., Langford, C. F., Weng, Z., Birney, E., Carter, N. P., Vetrie, D., & Dunham, I. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Research*, *17*(6), 691–707. https://doi.org/10.1101/gr.5704207

Köhler, C., Nittner, D., Rambow, F., Radaelli, E., Stanchi, F., Vandamme, N., Baggiolini, A., Sommer, L., Berx, G., van den Oord, J. J., Gerhardt, H., Blanpain, C., & Marine, J.-C. (2017). Mouse Cutaneous Melanoma Induced by Mutant BRaf Arises from Expansion and Dedifferentiation of Mature Pigmented Melanocytes. *Cell Stem Cell*, *21*(5), 679–693.e6. https://doi.org/10.1016/j.stem.2017.08.003

Kornberg, R. D. (1977). Structure of Chromatin. *Annual Review of Biochemistry*, *46*(1), 931–954. https://doi.org/10.1146/annurev.bi.46.070177.004435

Krah, N. M., De La O, J.-P., Swift, G. H., Hoang, C. Q., Willet, S. G., Chen Pan, F., Cash, G. M., Bronner, M. P., Wright, C. V., MacDonald, R. J., & Murtaugh, L. C. (2015). The acinar differentiation determinant PTF1A inhibits initiation of pancreatic ductal adenocarcinoma. *eLife*, *4*, e07125. https://doi.org/10.7554/eLife.07125

Krueger, F., James, F., Ewels, P., Afyounian, E., & Schuster-Boeckler, B. (2021, July). TrimGalore v0.6.7.

Kuo, M.-H., & Allis, C. D. (1998). Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays*, *20*(8), 615–626. https://doi.org/10.1002/(SICI)1521-1878(199808)20:8<615::AID-BIES4>3.0.CO;2-H

Kwon, S., Ban, K., Hong, Y.-K., Sung, J.-S., & Choi, I. (2022). PROX1, a Key Mediator of the Anti-Proliferative Effect of Rapamycin on Hepatocellular Carcinoma Cells. *Cells*, *11*(3), 446. https://doi.org/10.3390/cells11030446

Laerm, A., Helmbold, P., Goldberg, M., Dammann, R., Holzhausen, H.-J., & Ballhausen, W. G. (2007). Prospero-related homeobox 1 (PROX1) is frequently inactivated by genomic deletions and epigenetic silencing in carcinomas of the bilary system. *Journal of Hepatology*, *46*(1), 89–97. https://doi.org/10.1016/j.jhep.2006.07.033

Laitinen, A., Böckelman, C., Hagström, J., Kokkola, A., Kallio, P., & Haglund, C. (2017). High PROX1 expression in gastric cancer predicts better survival. *PLOS ONE*, *12*(8), e0183868. https://doi.org/10.1371/journal.pone.0183868

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*(4), 650–665. https://doi.org/10.1016/j.cell.2018.01.029

Lammerding, J. (2011). Mechanics of the nucleus. *Comprehensive Physiology*, *1*(2), 783–807. https://doi.org/10.1002/cphy.c100038

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton,

R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., … International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lavado, A., & Oliver, G. (2007). Prox1 expression patterns in the developing and adult murine brain. *Developmental Dynamics*, *236*(2), 518–524. https://doi.org/10.1002/dvdy.21024

Lavado, A., Lagutin, O. V., Chow, L. M. L., Baker, S. J., & Oliver, G. (2010). Prox1 Is Required for Granule Cell Maturation and Intermediate Progenitor Maintenance During Brain Neurogenesis (T. D. Palmer, Ed.). *PLoS Biology*, *8*(8), e1000460. https://doi.org/10.1371/journal.pbio.1000460

Leavitt, T., Hu, M. S., Borrelli, M. R., Januszyk, M., Garcia, J. T., Ransom, R. C., Mascharak, S., desJardins-Park, H. E., Litzenburger, U. M., Walmsley, G. G., Marshall, C. D., Moore, A. L., Duoto, B., Adem, S., Foster, D. S., Salhotra, A., Shen, A. H., Griffin, M., Shen, E. Z., Barnes, L. A., Zielins, E. R., Maan, Z. N., Wei, Y., Chan, C. K., Wan, D. C., Lorenz, H. P., Chang, H. Y., Gurtner, G. C., & Longaker, M. T. (2020). Prrx1 Fibroblasts Represent a Pro-fibrotic Lineage in the Mouse Ventral Dermis. *Cell Reports*, *33*(6), 108356. https://doi.org/10.1016/j.celrep.2020.108356

Lee, K.-W., Yeo, S.-Y., Gong, J.-R., Koo, O.-J., Sohn, I., Lee, W. Y., Kim, H. C., Yun, S. H., Cho, Y. B., Choi, M.-A., An, S., Kim, J., Sung, C. O., Cho, K.-H., & Kim, S.-H. (2022). PRRX1 is a master transcription factor of stromal fibroblasts for myofibroblastic lineage progression. *Nature Communications*, *13*(1), 2793. https://doi.org/10.1038/s41467-022-30484-4

Lee, Q. Y., Mall, M., Chanda, S., Zhou, B., Sharma, K. S., Schaukowitch, K., Adrian-Segarra, J. M., Grieder, S. D., Kareta, M. S., Wapinski, O. L., Ang, C. E., Li, R., Südhof, T. C., Chang, H. Y., & Wernig, M. (2020). Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. *Nature Cell Biology*, *22*(4), 401–411. https://doi.org/10.1038/s41556-020-0490-3

Lee, S., Kang, J., Yoo, J., Ganesan, S. K., Cook, S. C., Aguilar, B., Ramu, S., Lee, J., & Hong, Y.-K. (2009). Prox1 physically and functionally interacts with COUP-TFII to specify lymphatic endothelial cell fate. *Blood*, *113*(8), 1856–1859. https://doi.org/10.1182/blood-2008-03-145789

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, *23*(20), 4051–4060. https://doi.org/10.1038/sj.emboj.7600385

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. *Nature Reviews. Genetics*, *13*(4), 233–245. https://doi.org/10.1038/nrg3163

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. https://doi.org/10.1093/hmg/ddg180

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A., & Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, *7*(9), 709–715. https://doi.org/10.1038/nmeth.1491

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics*, *3*(9), 662–673. https://doi.org/10.1038/nrg887

Li, S. (2017). GeneOverlap.

Li, W., Li, L., & Hui, L. (2020). Cell Plasticity in Liver Regeneration. *Trends in Cell Biology*, *30*(4), 329–338. https://doi.org/10.1016/j.tcb.2020.01.007

Liu, C.-Y., Chen, K.-F., & Chen, P.-J. (2015a). Treatment of Liver Cancer. *Cold Spring Harbor Perspectives in Medicine*, *5*(9), a021535. https://doi.org/10.1101/cshperspect.a021535

Liu, Y., Ye, X., Zhang, J.-B., Ouyang, H., Shen, Z., Wu, Y., Wang, W., Wu, J., Tao, S., Yang, X., Qiao, K., Zhang, J., Liu, J., Fu, Q., & Xie, Y. (2015b). PROX1 promotes hepatocellular carcinoma proliferation and sorafenib resistance by enhancing β-catenin expression and nuclear translocation. *Oncogene*, *34*(44), 5524–5535. https://doi.org/10.1038/onc.2015.7

Liu, Y., Zhang, J.-B., Qin, Y., Wang, W., Wei, L., Teng, Y., Guo, L., Zhang, B., Lin, Z., Liu, J., Ren, Z.-G., Ye, Q.-H., & Xie, Y. (2013). PROX1 promotes hepatocellular carcinoma metastasis by way of up-regulating hypoxia-inducible factor 1α expression and protein stability: Hepatology. *Hepatology*, *58*(2), 692–705. https://doi.org/10.1002/hep.26398

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., & Finn, R. S. (2021). Hepatocellular carcinoma. *Nature Reviews Disease Primers*, *7*(1), 1–28. https://doi.org/10.1038/s41572-020-00240-3

Loebel, D. A. F., Watson, C. M., De Young, R. A., & Tam, P. P. L. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Developmental Biology*, *264*(1), 1–14. https://doi.org/10.1016/S0012-1606(03)00390-7

Lorch, Y., & Kornberg, R. D. (2017). Chromatin-remodeling for transcription. *Quarterly Reviews of Biophysics*, *50*, e5. https://doi.org/10.1017/S003358351700004X

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–210. https://doi.org/10.1016/0092-8674(87)90561-7

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Z., Priya Rajan, S. A., Song, Q., Zhao, Y., Wan, M., Aleman, J., Skardal, A., Bishop, C., Atala, A., & Lu, B. (2021). 3D scaffold-free microlivers with drug metabolic function generated by lineage-reprogrammed hepatocytes from human fibroblasts. *Biomaterials*, *269*, 120668. https://doi.org/10.1016/j.biomaterials.2021.120668

Luger, K., & Richmond, T. J. (1998). DNA binding within the nucleosome core. *Current Opinion in Structural Biology*, *8*(1), 33–40. https://doi.org/10.1016/s0959-440x(98)80007-9

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Ma, X., Wang, D., Zhao, W., & Xu, L. (2018). Deciphering the Roles of PPARγ in Adipocytes via Dynamic Change of Transcription Complex. *Frontiers in Endocrinology*, *9*, 473. https://doi.org/10.3389/fendo.2018.00473

Mall, M., Kareta, M. S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S. D., Ge, X., Drake, S., Ang, C. E., Walker, B. M., Vierbuchen, T., Fuentes, D. R., Brennecke, P., Nitta, K. R., Jolma, A., Steinmetz, L. M., Taipale, J., Südhof, T. C., & Wernig, M. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature*, *544*(7649), 245–249. https://doi.org/10.1038/nature21722

Mannervik, M., Nibu, Y., Zhang, H., & Levine, M. (1999). Transcriptional Coregulators in Development. *Science*, *284*(5414), 606–609. https://doi.org/10.1126/science.284.5414.606

Marrero, J. A., Kulik, L. M., Sirlin, C. B., Zhu, A. X., Finn, R. S., Abecassis, M. M., Roberts, L. R., & Heimbach, J. K. (2018). Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology*, *68*(2), 723–750. https://doi.org/10.1002/hep.29913

Martinez-Corral, I., Ulvmar, M. H., Stanczuk, L., Tatin, F., Kizhatil, K., John, S. W., Alitalo, K., Ortega, S., & Makinen, T. (2015). Nonvenous Origin of Dermal Lymphatic Vasculature. *Circulation Research*, *116*(10), 1649–1654. https://doi.org/10.1161/CIRCRESAHA.116.306170

Mattick, J. S., Taft, R. J., & Faulkner, G. J. (2010). A global view of genomic information–moving beyond the gene and the master regulator. *Trends in genetics: TIG*, *26*(1), 21–28. https://doi.org/10.1016/j.tig.2009.11.002

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., & Bentley, D. L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, *385*(6614), 357–361. https://doi.org/10.1038/385357a0

Meers, M. P., Tenenbaum, D., & Henikoff, S. (2019). Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics & Chromatin*, *12*(1), 42. https://doi.org/10.1186/s13072-019-0287-4

Melhuish, T. A., Kowalczyk, I., Manukyan, A., Zhang, Y., Shah, A., Abounader, R., & Wotton, D. (2018). Myt1 and Myt1l transcription factors limit proliferation in GBM cells by repressing YAP1 expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1861*(11), 983–995. https://doi.org/10.1016/j.bbagrm.2018.10.005

Menyhárt, O., Nagy, Á., & Győrffy, B. (2018). Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. *Royal Society Open Science*, *5*(12), 181006. https://doi.org/10.1098/rsos.181006

Mirny, L. A. (2011). The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *19*(1), 37–51. https://doi.org/10.1007/s10577-010-9177-0

Miyoshi, G., Young, A., Petros, T., Karayannis, T., Chang, M. M., Lavado, A., Iwano, T., Nakajima, M., Taniguchi, H., Huang, Z. J., Heintz, N., Oliver, G., Matsuzaki, F., Machold, R. P., & Fishell, G. (2015). Prox1 Regulates the Subtype-Specific Development of Caudal Ganglionic Eminence-Derived GABAergic Cortical Interneurons. *The Journal of Neuroscience*, *35*(37), 12869–12889. https://doi.org/10.1523/jneurosci.1164-15.2015

Mohn, F., & Schübeler, D. (2009). Genetics and epigenetics: Stability and plasticity during cellular differentiation. *Trends in Genetics*, *25*(3), 129–136. https://doi.org/10.1016/j.tig.2008.12.005

Mu, X., Español-Suñer, R., Mederacke, I., Affò, S., Manco, R., Sempoux, C., Lemaigre, F. P., Adili, A., Yuan, D., Weber, A., Unger, K., Heikenwälder, M., Leclercq, I. A., & Schwabe, R. F. (2015). Hepatocellular carcinoma originates from hepatocytes and not from the progenitor/biliary compartment. *The Journal of Clinical Investigation*, *125*(10), 3891–3903. https://doi.org/10.1172/JCI77995

Nkontchou, G., Tran Van Nhieu, J., Ziol, M., Tengher, I., Mahmoudi, A., Roulot, D., Bourcier, V., Ganne Carrie, N., Grando-Lemaire, V., Trinchet, J.-C., Cherqui, D., & Beaugrand, M. (2013). Peripheral intrahepatic cholangiocarcinoma occurring in patients without cirrhosis or chronic bile duct diseases: Epidemiology and histopathology of distant nontumoral liver in 57 White patients. *European Journal of Gastroenterology & Hepatology*, *25*(1), 94. https://doi.org/10.1097/MEG.0b013e328357cdd7

Ntikoudi, E., Pergaris, A., Kykalos, S., Politi, E., & Theocharis, S. (2022). The Role of PROX1 in Neoplasia: A Key Player Often Overlooked. *Diagnostics*, *12*(7), 1624. https://doi.org/10.3390/diagnostics12071624

Ohno, S. (1979). *Major Sex-Determining Genes* (F. Gross, A. Labhard, M. B. Lipsett, T. Mann, L. T. Samuels, & J. Zander, Eds.; Vol. 11). Springer Berlin Heidelberg.

Ouyang, H., Qin, Y., Liu, Y., Xie, Y., & Liu, J. (2013). Prox1 Directly Interacts with LSD1 and Recruits the LSD1/NuRD Complex to Epigenetically Co-Repress CYP7A1 Transcription (R. Mantovani, Ed.). *PLoS ONE*, *8*(4), e62192. https://doi.org/10.1371/journal.pone.0062192

Papizan, J. B., Singer, R. A., Tschen, S.-I., Dhawan, S., Friel, J. M., Hipkens, S. B., Magnuson, M. A., Bhushan, A., & Sussel, L. (2011). Nkx2.2 repressor complex regulates islet β-cell specification and prevents β-to-α-cell reprogramming. *Genes & Development*, *25*(21), 2291–2305. https://doi.org/10.1101/gad.173039.111

Paule, M. R., & White, R. J. (2000). Survey and summary: Transcription by RNA polymerases I and III. *Nucleic Acids Research*, *28*(6), 1283–1298. https://doi.org/10.1093/nar/28.6.1283

Pennacchio, L. A., Loots, G. G., Nobrega, M. A., & Ovcharenko, I. (2007). Predicting tissue-specific enhancers in the human genome. *Genome Research*, *17*(2), 201–211. https://doi.org/10.1101/gr.5972507

Perla Oliveira Schulz, F. G., alves Ferreira, M. d. F., Ara&uacute, T., jo Nascimento, A. V., & Ibrahim David, L. A. S. (2015). Association of nonalcoholic fatty liver disease and liver cancer. *World Journal of Gastroenterology*, *21*(3), 913–918. https://doi.org/10.3748/wjg.v21.i3.913

Petrova, T. V. (2002). Lymphatic endothelial reprogramming of vascular endothelial cells by the Prox-1 homeobox transcription factor. *The EMBO Journal*, *21*(17), 4593–4599. https://doi.org/10.1093/emboj/cdf470

Pujadas, E., & Feinberg, A. P. (2012). Regulated Noise in the Epigenetic Landscape of Development and Disease. *Cell*, *148*(6), 1123–1131. https://doi.org/10.1016/j.cell.2012.02.045

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Radman-Livaja, M., & Rando, O. J. (2010). Nucleosome positioning: How is it established, and why does it matter? *Developmental Biology*, *339*(2), 258–266. https://doi.org/10.1016/j.ydbio.2009.06.012

Radyk, M. D., Burclaff, J., Willet, S. G., & Mills, J. C. (2018). Metaplastic Cells in the Stomach Arise, Independently of Stem Cells, via Dedifferentiation or Transdifferentiation of Chief Cells. *Gastroenterology*, *154*(4), 839–843.e2. https://doi.org/10.1053/j.gastro.2017.11.278

Ragusa, S., Prat-Luri, B., González-Loyola, A., Nassiri, S., Squadrito, M. L., Guichard, A., Cavin, S., Gjorevski, N., Barras, D., Marra, G., Lutolf, M. P., Perentes, J., Corse, E., Bianchi, R., Wetterwald, L., Kim, J., Oliver, G., Delorenzi, M., Palma, M. D., & Petrova, T. V. (2020). Antiangiogenic immunotherapy suppresses desmoplastic and chemoresistant intestinal tumors in mice. *Journal of Clinical Investigation*, *130*(3), 1199–1216. https://doi.org/10.1172/jci129558

Rajakumari, S., Wu, J., Ishibashi, J., Lim, H.-W., Giang, A.-H., Won, K.-J., Reed, R. R., & Seale, P. (2013). EBF2 determines and maintains brown adipocyte identity. *Cell Metabolism*, *17*(4), 562–574. https://doi.org/10.1016/j.cmet.2013.01.015

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/nar/gkw257

Revia, S., Seretny, A., Wendler, L., Banito, A., Eckert, C., Breuer, K., Mayakonda, A., Lutsik, P., Evert, M., Ribback, S., Gallage, S., Chikh Bakri, I., Breuhahn, K., Schirmacher, P., Heinrich, S., Gaida, M. M., Heikenwälder, M., Calvisi, D. F., Plass, C., Lowe, S. W., & Tschaharganeh, D. F. (2022). Histone H3K27 demethylase KDM6A is an epigenetic gatekeeper of mTORC1 signalling in cancer. *Gut*, *71*(8), 1613–1628. https://doi.org/10.1136/gutjnl-2021-325405

Rizvi, S., & Gores, G. J. (2013). Pathogenesis, Diagnosis, and Management of Cholangiocarcinoma. *Gastroenterology*, *145*(6), 1215–1229. https://doi.org/10.1053/j.gastro.2013.10.013

Rodrigues, M. F., Rodini, C. d. O., Xavier, F. C. d. A., Paiva, K. B., Severino, P., Moyses, R. A., López, R. M., DeCicco, R., Rocha, L. A., Carvalho, M. B., Tajara, E. H., & Nunes, F. D. (2014). PROX1 Gene is Differentially Expressed in Oral Cancer and Reduces Cellular Proliferation. *Medicine*, *93*(28), e192. https://doi.org/10.1097/md.0000000000000192

Roeder, R. G., & Rutter, W. J. (1969). Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. *Nature*, *224*(5216), 234–237. https://doi.org/10.1038/224234a0

Rosen, E. D., Sarraf, P., Troy, A. E., Bradwin, G., Moore, K., Milstone, D. S., Spiegelman, B. M., & Mortensen, R. M. (1999). PPARγ Is Required for the Differentiation of Adipose Tissue In Vivo and In Vitro. *Molecular Cell*, *4*(4), 611–617. https://doi.org/10.1016/S1097-2765(00)80211-7

Rosenfeld, J. A., Wang, Z., Schones, D. E., Zhao, K., DeSalle, R., & Zhang, M. Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics*, *10*, 143. https://doi.org/10.1186/1471-2164-10-143

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., & Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development (Cambridge, England)*, *132*(4), 797–803. https://doi.org/10.1242/dev.01613

Schoenherr, C. J., & Anderson, D. J. (1995). The Neuron-Restrictive Silencer Factor (NRSF): A Coordinate Repressor of Multiple Neuron-Specific Genes. *Science*, *267*(5202), 1360–1363. https://doi.org/10.1126/science.7871435

Seehawer, M., Heinzmann, F., D'Artista, L., Harbig, J., Roux, P.-F., Hoenicke, L., Dang, H., Klotz, S., Robinson, L., Doré, G., Rozenblum, N., Kang, T.-W., Chawla, R., Buch, T., Vucur, M., Roth, M., Zuber, J., Luedde, T., Sipos, B., Longerich, T., Heikenwälder, M., Wang, X. W., Bischof, O., & Zender, L. (2018). Necroptosis microenvironment directs lineage commitment in liver cancer. *Nature*, *562*(7725), 69–75. https://doi.org/10.1038/s41586-018-0519-y

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*(3), 538–547. https://doi.org/10.1242/dev.099481

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J. R., Wu, C.-t., Bender, W., & Kingston, R. E. (1999). Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. *Cell*, *98*(1), 37–46. https://doi.org/10.1016/S0092-8674(00)80604-2

Shi, G., DiRenzo, D., Qu, C., Barney, D., Miley, D., & Konieczny, S. F. (2013). Maintenance of acinar cell organization is critical to preventing Kras-induced acinar-ductal metaplasia. *Oncogene*, *32*(15), 1950–1958. https://doi.org/10.1038/onc.2012.210

Shimoda, M., Takahashi, M., Yoshimoto, T., Kono, T., Ikai, I., & Kubo, H. (2006). A Homeobox Protein, Prox1, Is Involved in the Differentiation, Proliferation, and Prognosis in Hepatocellular Carcinoma. *Clinical Cancer Research*, *12*(20), 6005–6011. https://doi.org/10.1158/1078-0432.CCR-06-0712

Sirica, A. E., & Gores, G. J. (2014). Desmoplastic Stroma and Cholangiocarcinoma: Clinical Implications and Therapeutic Targeting. *Hepatology (Baltimore, Md.)*, *59*(6), 2397–2402. https://doi.org/10.1002/hep.26762

Si-Tayeb, K., Lemaigre, F. P., & Duncan, S. A. (2010). Organogenesis and Development of the Liver. *Developmental Cell*, *18*(2), 175–189. https://doi.org/10.1016/j.devcel.2010.01.011

Song, G., Pacher, M., Balakrishnan, A., Yuan, Q., Tsay, H.-C., Yang, D., Reetz, J., Brandes, S., Dai, Z., Pützer, B. M., Araúzo-Bravo, M. J., Steinemann, D., Luedde, T., Schwabe, R. F., Manns, M. P., Schöler, H. R., Schambach, A., Cantz, T., Ott, M., & Sharma, A. D. (2016). Direct Reprogramming of Hepatic Myofibroblasts into Hepatocytes In Vivo Attenuates Liver Fibrosis. *Cell Stem Cell*, *18*(6), 797–808. https://doi.org/10.1016/j.stem.2016.01.010

Sosa-Pineda, B., Wigle, J. T., & Oliver, G. (2000). Hepatocyte migration during liver development requires Prox1. *Nature Genetics*, *25*(3), 254–255. https://doi.org/10.1038/76996

Stark, R., & Brown, G. D. (2011). DiffBind: Differential binding analysis of ChIP-seq peak data.

Stergiopoulos, A., Elkouris, M., & Politis, P. K. (2015). Prospero-related homeobox 1 (Prox1) at the crossroads of diverse pathways during adult neural fate specification. *Frontiers in Cellular Neuroscience*, *8*, 454. https://doi.org/10.3389/fncel.2014.00454

Sun, L., Wang, Y., Cen, J., Ma, X., Cui, L., Qiu, Z., Zhang, Z., Li, H., Yang, R.-Z., Wang, C., Chen, X., Wang, L., Ye, Y., Zhang, H., Pan, G., Kang, J.-S., Ji, Y., Zheng, Y.-W., Zheng, S., & Hui, L. (2019). Modelling liver cancer initiation with organoids derived from directly reprogrammed human hepatocytes. *Nature Cell Biology*, *21*(8), 1015–1026. https://doi.org/10.1038/s41556-019-0359-5

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

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*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Talbert, P. B., Meers, M. P., & Henikoff, S. (2019). Old cogs, new tricks: The evolution of gene expression in a chromatin context. *Nature Reviews Genetics*, *20*(5), 283–297. https://doi.org/10.1038/s41576-019-0105-7

The Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, & Principal investigators. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*(7727), 367–372. https://doi.org/10.1038/s41586-018-0590-4

The Tabula Muris Consortium, Almanzar, N., Antony, J., Baghel, A. S., Bakerman, I., Bansal, I., Barres, B. A., Beachy, P. A., Berdnik, D., Bilen, B., Brownfield, D., Cain, C., Chan, C. K. F., Chen, M. B., Clarke, M. F., Conley, S. D., Darmanis, S., Demers, A., Demir, K., de Morree, A., Divita, T., du Bois, H., Ebadi, H., Espinoza, F. H., Fish, M., Gan, Q., George, B. M., Gillich, A., Gòmez-Sjöberg, R., Green, F., Genetiano, G., Gu, X., Gulati, G. S., Hahn, O., Haney, M. S., Hang, Y., Harris, L., He, M., Hosseinzadeh, S., Huang, A., Huang, K. C., Iram, T., Isobe, T., Ives, F., Jones, R. C., Kao, K. S., Karkanias, J., Karnam, G., Keller, A., Kershner, A. M., Khoury, N., Kim, S. K., Kiss, B. M., Kong, W., Krasnow, M. A., Kumar, M. E., Kuo, C. S., Lam, J., Lee, D. P., Lee, S. E., Lehallier, B., Leventhal, O., Li, G., Li, Q., Liu, L., Lo, A., Lu, W.-J., Lugo-Fagundo, M. F., Manjunath, A., May, A. P., Maynard, A., McGeever, A., McKay, M., McNerney, M. W., Merrill, B., Metzger, R. J., Mignardi, M., Min, D., Nabhan, A. N., Neff, N. F., Ng, K. M., Nguyen, P. K., Noh, J., Nusse, R., Pálovics, R., Patkar, R., Peng, W. C., Penland, L., Pisco, A. O., Pollard, K., Puccinelli, R., Qi, Z., Quake, S. R., Rando, T. A., Rulifson, E. J., Schaum, N., Segal, J. M., Sikandar, S. S., … Zou, J. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, *583*(7817), 590–595. https://doi.org/10.1038/s41586-020-2496-1

Treutlein, B., Lee, Q. Y., Camp, J. G., Mall, M., Koh, W., Shariati, S. A. M., Sim, S., Neff, N. F., Skotheim, J. M., Wernig, M., & Quake, S. R. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature*, *534*(7607), 391–395. https://doi.org/10.1038/nature18323

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: Function, expression and evolution. *Nature Reviews Genetics*, *10*(4), 252–263. https://doi.org/10.1038/nrg2538

Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Südhof, T. C., & Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, *463*(7284), 1035–1041. https://doi.org/10.1038/nature08797

Vos, S. M. (2021). Understanding transcription across scales: From base pairs to chromosomes. *Molecular Cell*, *81*(8), 1601–1616. https://doi.org/10.1016/j.molcel.2021.03.002

Waddington, C. H. (1957). *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. Allen & Unwin.

Waldron, L., Steimle, J. D., Greco, T. M., Gomez, N. C., Dorr, K. M., Kweon, J., Temple, B., Yang, X. H., Wilczewski, C. M., Davis, I. J., Cristea, I. M., Moskowitz, I. P., & Conlon, F. L. (2016). The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Developmental Cell*, *36*(3), 262–275. https://doi.org/10.1016/j.devcel.2016.01.009

Wang, H., Yang, Y., Liu, J., & Qian, L. (2021). Direct cell reprogramming: Approaches, mechanisms and progress. *Nature Reviews Molecular Cell Biology*, *22*(6), 410–424. https://doi.org/10.1038/s41580-021-00335-z

Wapinski, O. L., Vierbuchen, T., Qu, K., Lee, Q. Y., Chanda, S., Fuentes, D. R., Giresi, P. G., Ng, Y. H., Marro, S., Neff, N. F., Drechsel, D., Martynoga, B., Castro, D. S., Webb, A. E., Südhof, T. C., Brunet, A., Guillemot, F., Chang, H. Y., & Wernig, M. (2013). Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. *Cell*, *155*(3), 621–635. https://doi.org/10.1016/j.cell.2013.09.028

Weigel, B., Tegethoff, J. F., Grieder, S. D., Lim, B., Nagarajan, B., Liu, Y.-C., Truberg, J., Papageorgiou, D., Adrian-Segarra, J. M., Schmidt, L. K., Kaspar, J., Poisel, E., Heinzelmann, E., Saraswat, M., Christ, M., Arnold, C., Ibarra, I. L., Campos, J., Krijgsveld, J., Monyer, H., Zaugg, J. B., Acuna, C., & Mall, M. (2023). MYT1L haploinsufficiency in human neurons and mice causes autism-associated phenotypes that can be reversed by genetic and pharmacologic intervention. *Molecular Psychiatry*, 1–14. https://doi.org/10.1038/s41380-023-01959-7

Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proceedings of the National Academy of Sciences*, *86*(14), 5434–5438. https://doi.org/10.1073/pnas.86.14.5434

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., Najafabadi, H. S., Lambert, S. A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.-C., Galli, M., Lewsey, M. G., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J. S., Govindarajan, S., Shaulsky, G., Walhout, A. J. M., Bouget, F.-Y., Ratsch, G., Larrondo, L. F., Ecker, J. R., & Hughes, T. R. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, *158*(6), 1431–1443. https://doi.org/10.1016/j.cell.2014.08.009

Wigle, J. T. (2002). An essential role for Prox1 in the induction of the lymphatic endothelial cell phenotype. *The EMBO Journal*, *21*(7), 1505–1513. https://doi.org/10.1093/emboj/21.7.1505

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems*, *8*(4), 281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005

Wong, B. W., Wang, X., Zecchin, A., Thienpont, B., Cornelissen, I., Kalucka, J., García-Caballero, M., Missiaen, R., Huang, H., Brüning, U., Blacher, S., Vinckier, S., Goveia, J., Knobloch, M., Zhao, H., Dierkes, C., Shi, C., Hägerling, R., Moral-Dardé, V., Wyns, S., Lippens, M., Jessberger, S., Fendt, S.-M., Luttun, A., Noel, A., Kiefer, F., Ghesquière, B., Moons, L., Schoonjans, L., Dewerchin, M., Eelen, G., Lambrechts, D., & Carmeliet, P. (2017). The role of fatty acid β-oxidation in lymphangiogenesis. *Nature*, *542*(7639), 49–54. https://doi.org/10.1038/nature21028

Wu, W., He, X., Andayani, D., Yang, L., Ye, J., Li, Y., Chen, Y., & Li, L. (2017). Pattern of distant extrahepatic metastases in primary liver cancer: A SEER based study. *Journal of Cancer*, *8*(12), 2312–2318. https://doi.org/10.7150/jca.19056

Yokobori, T., Bao, P., Fukuchi, M., Altan, B., Ozawa, D., Rokudai, S., Bai, T., Kumakura, Y., Honjo, H., Hara, K., Sakai, M., Sohda, M., Miyazaki, T., Ide, M., Nishiyama, M., Oyama, T., & Kuwano, H. (2015). Nuclear PROX1 is Associated with Hypoxia-Inducible Factor 1α Expression and Cancer Progression in Esophageal Squamous Cell Carcinoma. *Annals of Surgical Oncology*, *22*(Suppl 3), 1566–1573. https://doi.org/10.1245/s10434-015-4831-6

Zaret, K. S. (2002). Regulatory phases of early liver development: Paradigms of organogenesis. *Nature Reviews Genetics*, *3*(7), 499–512. https://doi.org/10.1038/nrg837

Zaret, K. S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. *Annual Review of Genetics*, *54*(1), 367–385. https://doi.org/10.1146/annurev-genet-030220-015007

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zheng, C., Lu, X., Hansen, J. C., & Hayes, J. J. (2005). Salt-dependent Intra- and Internucleosomal Interactions of the H3 Tail Domain in a Model Oligonucleosomal Array*. *Journal of Biological Chemistry*, *280*(39), 33552–33557. https://doi.org/10.1074/jbc.M507241200

Zhou, Q., Choi, G., & Anderson, D. J. (2001). The bHLH Transcription Factor Olig2 Promotes Oligodendrocyte Differentiation in Collaboration with Nkx2.2. *Neuron*, *31*(5), 791–807. https://doi.org/10.1016/S0896-6273(01)00414-7

Zhu, J., Ordway, A., Weber, L., Buddika, K., & Kumar, J. P. (2018). Polycomb group (Pc-G) proteins and Pax6 cooperate to inhibit in vivo reprogramming of the developing Drosophila eye. *Development*, dev.160754. https://doi.org/10.1242/dev.160754

Zingg, J. M., Pedraza-Alva, G., & Jost, J. P. (1994). MyoD1 promoter autoregulation is mediated by two proximal E-boxes. *Nucleic Acids Research*, *22*(12), 2234–2241.

Appendix A

In silico screen supplemental data

A.1 Screen datasets and neuronal screen results

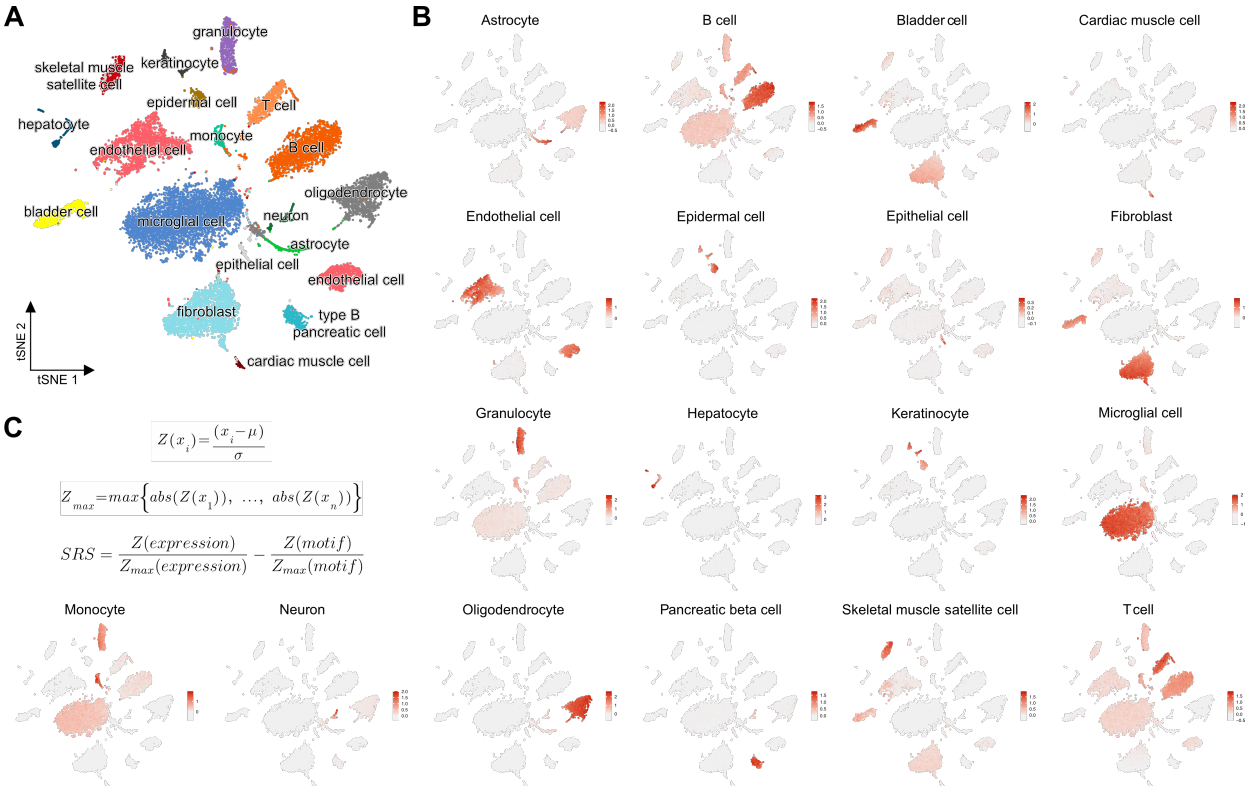

Figure A.1.1. *In silico* screen expression dataset and equations.
(A) Single cell UMAP of 18 cell types annotated by the Tabula Muris consortium. (B) Cell type-specific gene signatures used in this study are displayed across all cells in (A). (C) Equations used to calculate a safeguard repressor score for each TF (see Methods for details).

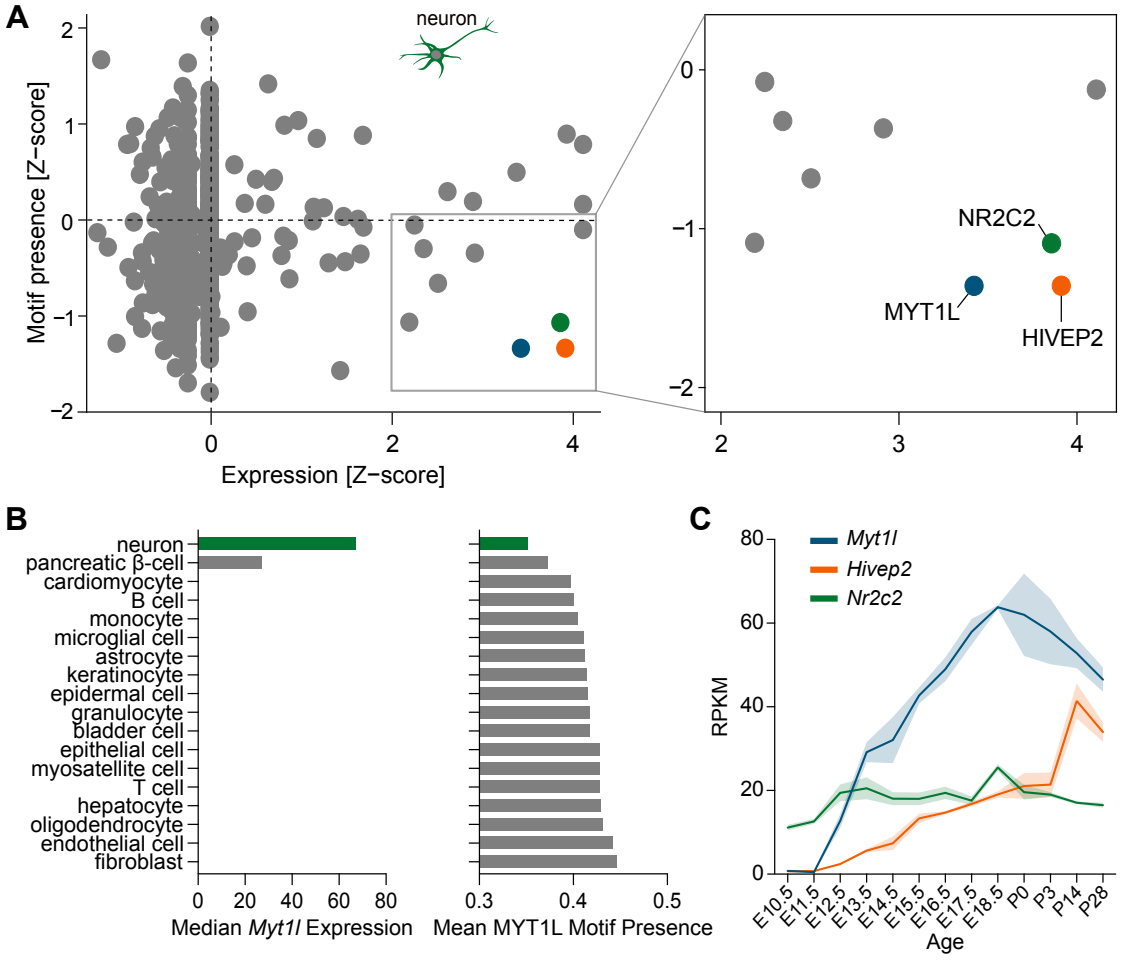

Figure A.1.2. *In silico* screen results for neurons.
(A) Expression and motif presence analysis of 1,296 TFs highlights three safeguard repressor candidates in neurons, including MYT1L. (B) *Myt1l* expression in 18 cell types from Tabula Muris and number of MYT1L motifs in promoters of cell type marker genes. (C) Developmental expression levels of the top three neuronal safeguard repressor candidates in the mouse brain.

Appendix B

Reprogramming supplemental data

B.1 scRNA-seq

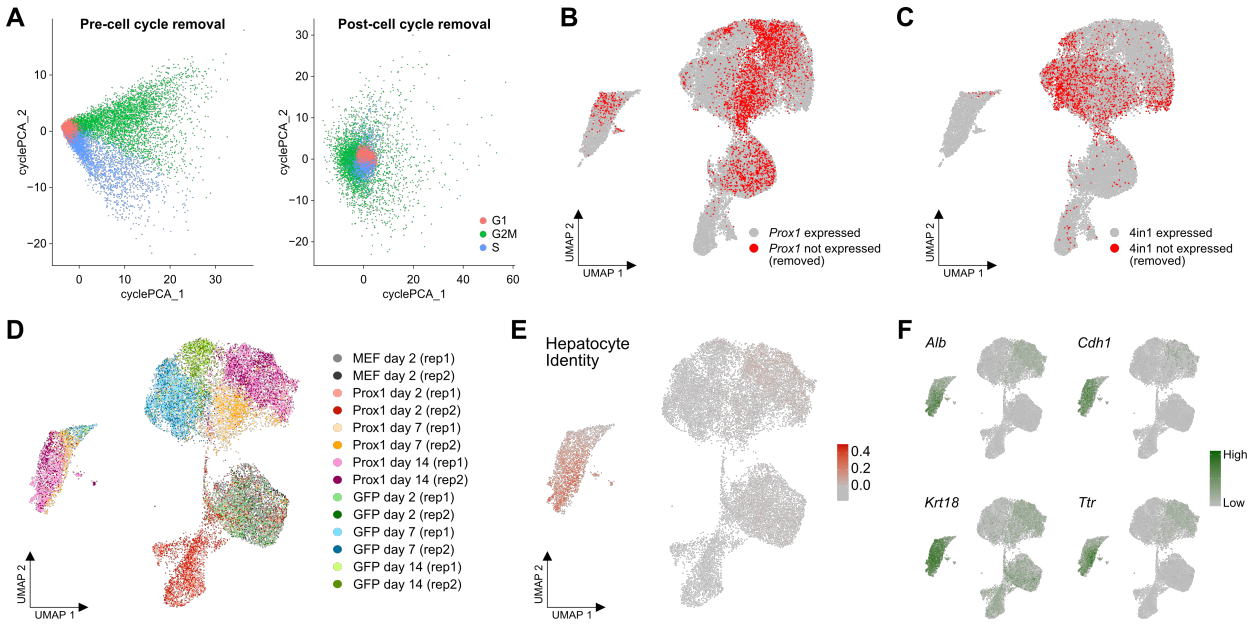

Figure B.1.1. scRNA-seq quality control and marker gene expression during hepatocyte reprogramming time course.
(A) PCA of cells based on cell cycle gene expression before and after adjustment for cell cycle effects using Seurat. (B and C) UMAP of cells labeled by expression of *Prox1* (B) and 4in1 (C), inferred from PROX1 and 4in1 activity. (D) UMAP of all cells labeled by condition and biological replicate following demultiplexing using barcode oligos. (E) Hepatocyte identity score in each cell on UMAP. (F) Expression of selected individual hepatocyte marker genes in all cells.

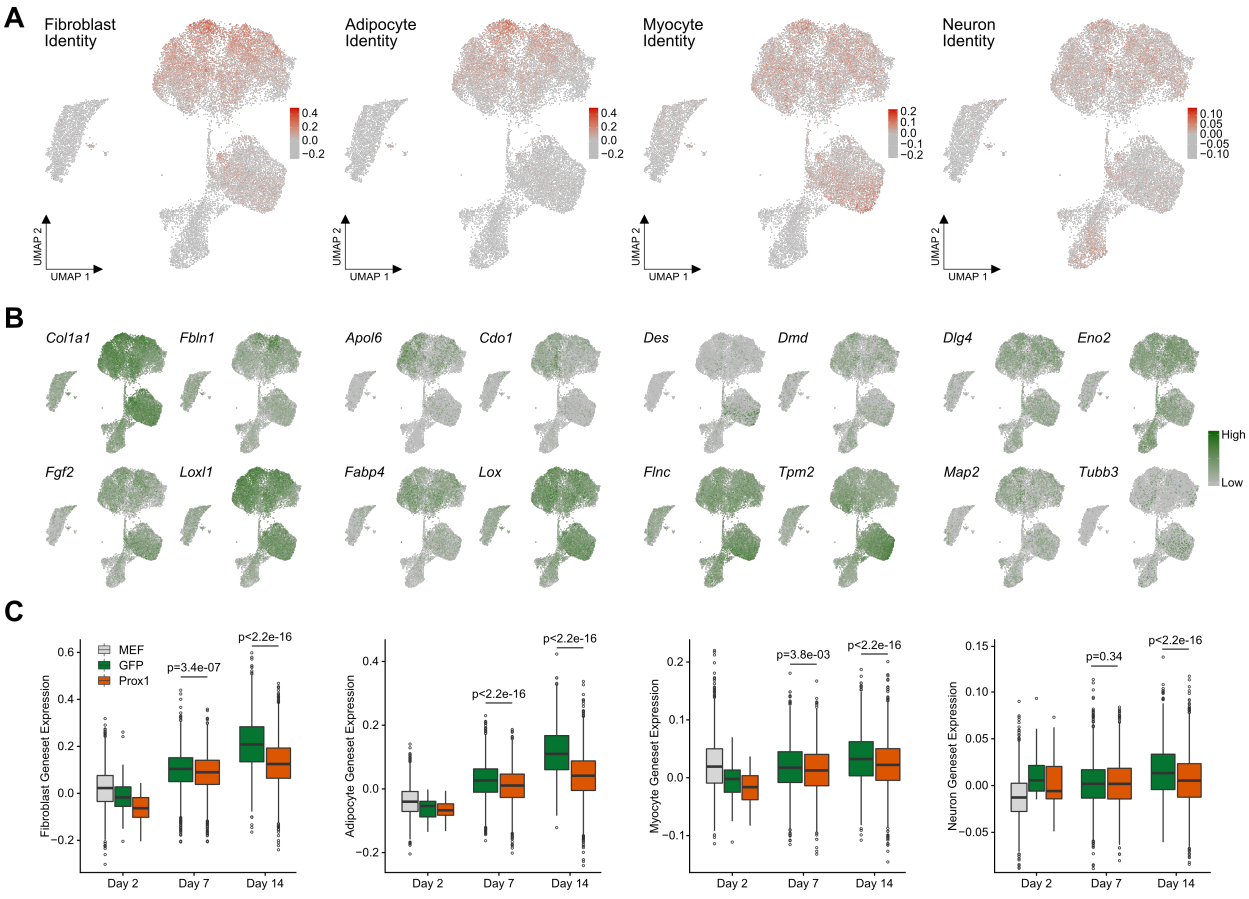

Figure B.1.2. Marker gene expression in scRNA-seq time course for selected alternate cell identities. (A) Fibroblast, adipocyte, myocyte, and neuron identity scores projected onto the UMAP. (B) Expression of selected fibroblast, adipocyte, myocyte, and neuronal marker genes in all cells. (C) Quantification of fibroblast, adipocyte, myocyte, and neuron identity scores shown as box-plots with p-values (two-tailed t-test) for each time point and treatment.

B.2 CUT&RUN and ATAC-seq

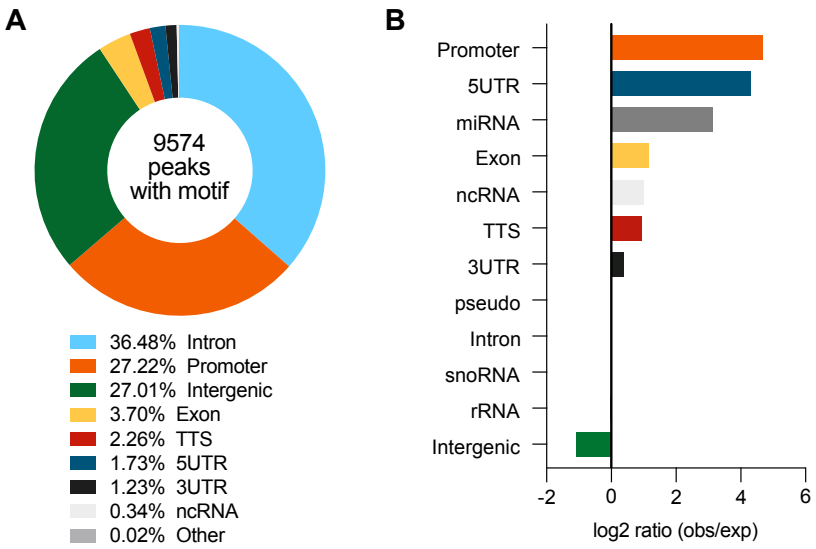

Figure B.2.1. CUT&RUN statistics.
(A) Proportion of PROX1 binding sites located in 9,574 high-confidence PROX1 binding sites. (C) Log2 ratio of observed PROX1 binding sites to expected binding events in indicated genomic regions highlight enrichment at promoters.

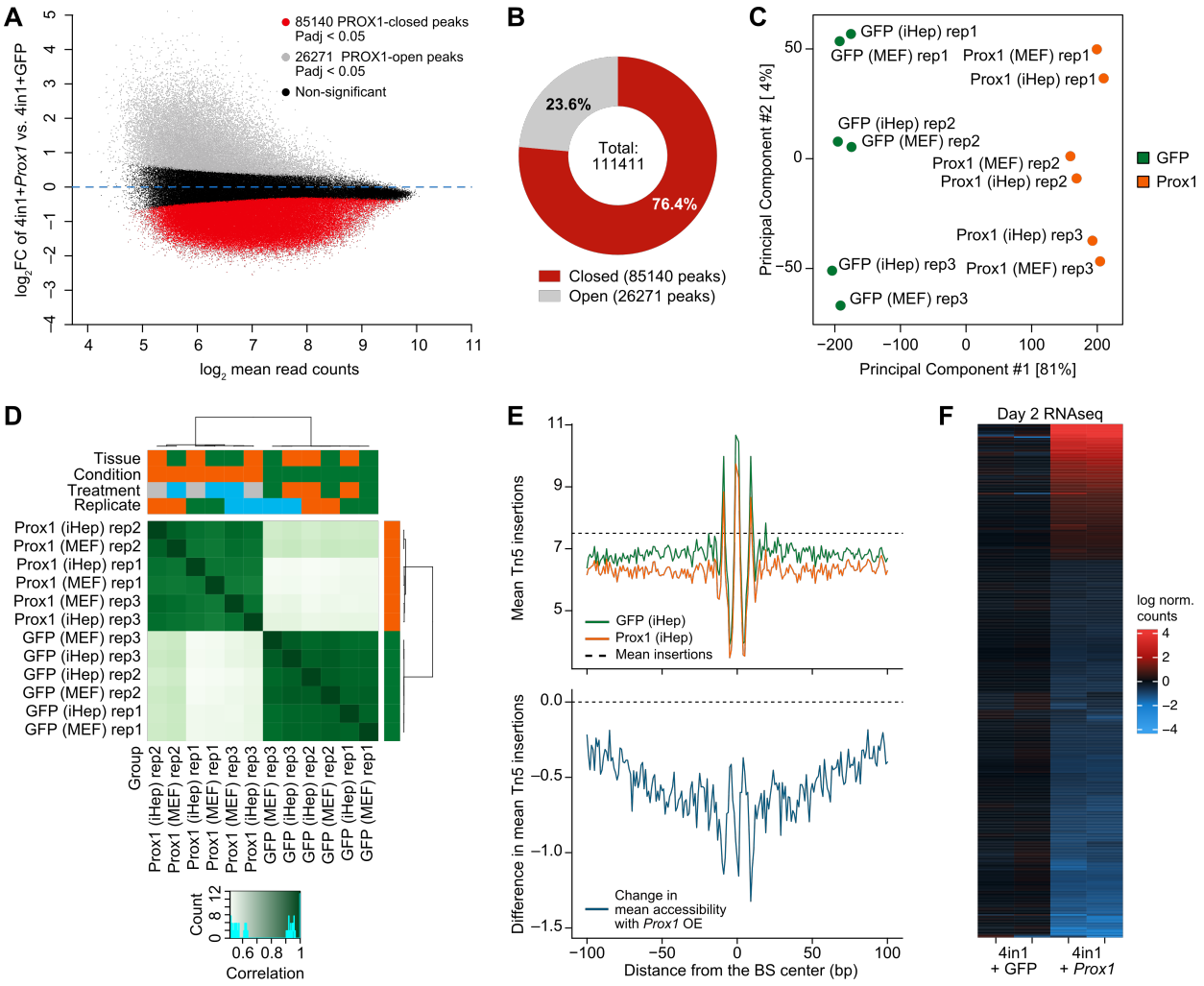

Figure B.2.2. ATAC-seq quality control and data visualisation.
(A) MA-plot of differential accessible regions (DAR) at day 2 of 4in1-induced hepatocyte reprogramming with or without Prox1 overexpression based on ATAC-seq. (B) Percentage of closed and opened regions upon Prox1 overexpression in A. (C) Correlation of DARs between indicated conditions and replicates in A and upon overexpression of GFP or Prox1 in MEFs. (D) PCA of DARs in C labeled by condition and replicate. (E) Top, mean Tn5 transposon adapter insertions at indicated conditions from A centred on PROX1 binding sites from CUT&RUN experiments. Bottom, changes in chromatin accessibility at PROX1 binding sites between 4in1 + GFP and 4in1 + Prox1 indicate decreased accessibility upon Prox1 overexpression. (F) Differential gene expression of cells in A based on RNA-seq (n=2), for genes with differentially-accessible regions (p-adj < 0.05) within 2 kb of their TSS and an overlapping CUT&RUN peak with a PROX1 binding motif, showing global down-regulation of target gene transcription upon Prox1 overexpression.

Appendix C

In vivo supplemental data

C.1 HDTVI experiments

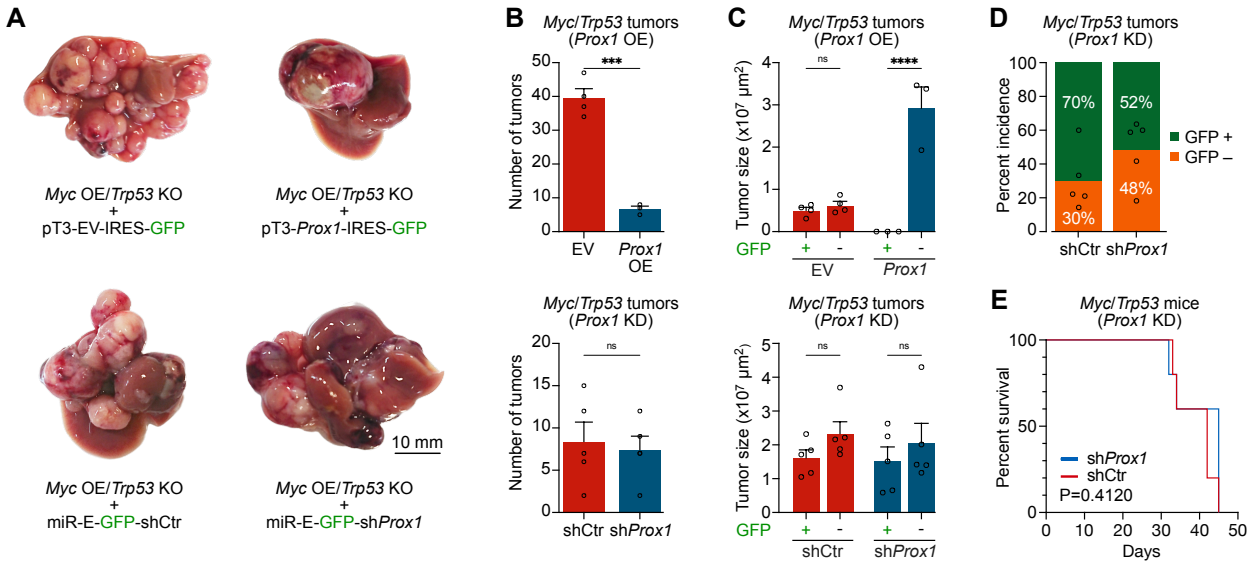

Figure C.1.1. PROX1 influences liver tumour fate in mice.
(A) Representative photographs of mouse livers at the end point following HCC-like tumour induction induced by *Myc* overexpression and *Trp53* knockout using HDTVI, together with *Prox1*-IRES-GFP overexpression or EF1a-GFP-sh*Prox1* knockdown compared to respective controls. (B) Number of tumours induced by HDTVI as in A, following overexpression (OE) of *Prox1*-IRES-GFP (n=3) vs GFP control (n=4), or knockdown (KD) of *Prox1* vs control (n=5). (C) Tumour size (cross-sectional area) following treatments as in B, showing tumours with or without GFP expression. (D) Quantification of GFP protein levels in tumours following *Prox1* KD or control as in B, shown as percentage of total. (E) Survival curve of HCC mice treated as in A comparing knockdown of Prox1 with control (n=5). Log-rank test for survival. Bar graphs show mean values from specified biological replicates. Error bars = SD. Two-sided t-test (B) or two-way ANOVA and Šidák correction (C), * p-adj < 0.05, ** p-adj < 0.01, *** p-adj < 0.001, **** p-adj < 0.0001.