Investigating Liver Regeneration Using Single Cell RNA Sequencing

Jordan Raymond Portman

PhD in Precision Medicine

The University of Edinburgh

2020

Abstract

Liver disease causes over 2 million deaths world-wide each year and is the most common disease related deaths in adults aged between 35-49 years old in the UK. While the liver has a large capacity to regenerate this is overwhelmed during chronic injury. Currently there is no drug-based treatment for end stage liver disease, with transplantation the only option. This is problematic as only around 10% of the global transplantation needs are met and urgent new therapeutics are required to tackling this highly prevalent disease. Despite decades of research elucidating the mechanisms behind the liver's natural capacity to regenerate, there has not yet been a therapeutic agent approved. The prominent feature of liver regeneration is the replication of mature hepatocytes. Research has assumed homogeneity in function among the hepatocyte population during regeneration, however recent studies have shown that there is a more heterogeneous response. More work is required to understand the complex nature of this process at a higher resolution to unpick any heterogeneity and discover any further regenerative signals. Therefore, this project aimed to use a single cell RNA sequencing approach to identify any potentially novel heterogeneous populations of hepatocytes during liver regeneration and further identify any key signalling molecules.

Replicating and non-replicating hepatocytes were sorted from a partial hepatectomies FUCCI2a mouse model which reports G1 cells with a mCherry signal and S/G2/M cells with a mVenus signal. Hepatocytes were sorted directly into 384 well plates for single cell sequencing. Unsupervised cluster was performed to identify several populations of hepatocytes. Non-replicating cells split into cluster denoting there spacial location in the liver lobule. This was identified through the gradient of expression of zonal marker genes. While no novel heterogeneous cluster of replicating hepatocytes were found, several novel signalling molecules such as Il33, Il15, Dll1, Cklf, and Bmp7, were identified to be expressed only by replicating hepatocytes. Furthermore, a population of "primed" mCherry positive hepatocytes appeared to express

several cell cycle marker genes. This population could represent a population of cells captured just before entry into S phase or a priming action of hepatocytes to prepare for cell cycle entry, awaiting an appropriate initiating signal. Interesting this population showed high expression for a circadian rhythm associated gene Timeless. Replication during partial hepatectomy is believed to be highly synchronous and Timeless may be the protein that controls the timing of replication during the model. Further studies of these genes are required to understand their function and importance in liver regeneration. In conclusion, the work demonstrated within this thesis highlights the need to study liver regeneration using high resolution techniques such as single cell RNA sequencing.

Lay Abstract

Liver disease is a major cause of death for many people worldwide. It is caused mainly by excessing alcohol consumption or obesity, although viral infections and drug overdose are also less common factors. There is no treatment for liver disease other than transplantation of the organ and unfortunately the number of people needing a transplant far exceeds the number of liver donors. Therefore, there is an urgent need to find a novel drug therapy to cure this disease. Luckily, the liver has a remarkable capacity to regenerate on its own, when it is still healthy. This occurs through cell division of the livers main cell type, the hepatocyte. Over the past decades, the many signalling pathways that are involved in the process of initiating a hepatocyte to divide, and regenerate the liver, have been identified. However, understanding of these pathways has yet to lead to any potential therapy. Further understanding of the regenerative process is needed. While most previous studies have thought of all hepatocytes as acting the same manner, recent studies have shown that this is not the case. While most hepatocytes will copy their genomic content, the first step of cell division, not all hepatocytes will continue to divide into two cells, the final step of cell division. Instead some hepatocytes stay as one cell. It is unclear why or how this occurs, however, it demonstrates a difference within the hepatocyte populations, a phenomenon called heterogeneity. Therefore, the aim of this project was to identify if there are any heterogeneous populations of regenerating hepatocytes, and if these cells are producing any new signals that will help us understand the process further.

One of the most common methods to force a liver to regenerate is the 2/3rd partial hepatectomy model conducted in mice. Here, 2/3rds of the liver is removed, and over a 7-day period, the hepatocytes will replicate, divide and regenerate the lost mass, resulting in a normal functioning liver. To isolate the replicating cells from the non-replicating cells, we used a transgenic mouse that causes the hepatocyte to produce a red signal when they are not replicating and a green signal when they are. The different hepatocytes can

then be classified by looking for either the red or green signal. To study the difference between the cells, a technique called single cell RNA sequencing was used. This technique allows the mRNA of each cell to be sequenced and compared to the mRNA of the other cells. The mRNA of each cell tells us what proteins the cell is making or intends to make, and these proteins define how the hepatocyte will function. By comparing tens to hundreds of green replicating hepatocytes and red non-replicating hepatocytes, I am able to identify if there are any differences between different types of hepatocytes. Unfortunately, I did not find any novel heterogeneous populations of replicating hepatocyte, however I did identify mRNA molecules expressed by the replicating hepatocytes that were previously unknown. These were Il33, Il15, Cklf, Dll1 and Bmp7. Furthermore, I also identified a population of hepatocytes that appeared to be waiting to replicate. These cells were expressing a mRNA, Timeless, that is associated with telling the cells what time of day it is, such as day or night. This protein made by Timeless could be controlling when each hepatocyte is allowed to enter the cell cycle. I would need to do further work, by first validating that the mRNA molecules are expressed into protein, as sometimes this does not occur. Then I will go on to identify the distinct function of each protein and how it plays a role during liver regeneration. The work shown within this thesis demonstrates the need to look at liver regeneration using these higher resolution techniques.

Acknowledgements

Firstly, I would like to thank my supervisors Prof. Neil Henderson and Prof. Chris Ponting, for their advice, support and guidance throughout this project. To Neil, who welcomed me into his research group, provided endless support and always had his door open for my endless list of questions – I thank you for all your encouragement over the past four years. Many thanks to my secondary supervisor Chris, for the informatics support, as well as having a reassuring smile when asking challenging questions, your advice throughout was invaluable. Also, a special mention to Dr. Prakash Ramachandran for his constant support with experimental design, troubleshooting, and ideas which were instrumental to my progression.

I would like to thank all past and present members of the Henderson research group. I could go through everyone and list all the amazing help they have provided but I fear this may become longer than my thesis. I would therefore like to thank: Andreas Kapourani, Bass Wallace, Beth Henderson, David Wilson, Dyana Markose, Elena Dora, Eleni Papachristoforou, Jasmin Paris, John Wilson Kanamori, Kylie Conroy, Laura Kitto, Madara Brice, Mariana Beltran-Sierra, Prasad Palani Velu, Richard Taylor, Ross Dobie and Stephen Greenhalgh for all the support, technical assistance, comic relief, putting up with my endless chatter, post Ru Paul debriefs, providing a continuous supply of cake, and genuinely being there for me throughout my entire PhD.

To the QMRI Flow Cytometry and Cell Sorting Facility staff Shonna Johnston, Will Ramsay and Mari Pattison, who took the time to both teach and help me use the flow machines – thank you, you not only helped me through the world of flow but were always there for a good old natter. A huge thank you to Jon Henderson – you went above and beyond what was asked of you to care for our animals and provide any technical assistance.

I would like to acknowledge the Medical Research Council for the financial support towards pursuing this PhD and the Precision Medicine PhD

programme directors and organisers for giving me the opportunity to take this incredible journey.

I am also super grateful to all my friends for their moral support. To Katherine Kentistou, Joanna Sweetman, Evan Evangelos and Daniel Simpson, who I started my PhD journey with – you made the move to Edinburgh so easy and made our struggle through the courses much more enjoyable. Thanks to the crab people - David Taggart, the most “modest” person you’d ever meet for his endless “encouragement” and finding Opium (the Bar not the Drug); Dyana Markose for “inviting” me to board games night them many years ago and being the second target of most of the jokes; Ross Mills for bringing the glamour of our home town Rochdale to Edinburgh; Isabel Bravo for all your amazing cooking and motherly love; Lisa Kelly for somehow bringing spring and summer early to Edinburgh; Nikolay Ogryzko for being the big deal that you are; Natasha Gibson who made finding the evil one in any game a lot easier; Lucy Remnant for being the biggest super nerd, although them ducks still give me nightmares; Luke Way for adding that extra bit of entertainment when England beat Wales in Rugby – and to all, you were like family, fighting over board games and bickering but always accompanied by tonnes of laughter and memories. I’d also like to thank everyone who I have met in the CIR, you make it a fantastic place to work. Especially, Danielle Minns, who fed me, put up with my whining, encouraged me … this list could go on but thank you for always being there every step of the way.

To the Happy Uni friends, James Kelly, Kat Davies, Carina Nicu, Nathaniel Ng and Caroline Pearson who kept me sane through all those early university days; unfortunately, I couldn’t fulfil my dominoes wager. To my oldest and closest friends Oliver Chandler, Maja Lorkowska, Adam Noon – you have been there since practically day one. No matter how far we are apart, you were always there for moral support and a good Christmas laugh or two.

Finally, I am extremely grateful to my family, both my mum and dad, sister, Shelley, brother-in-law, Michael, and two fabulous nieces, Skye and Heidi. All my achievements are the result of sacrifices you have made, and I dedicate this thesis to you.

Signed Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. The work presented is entirely my own, some experiments were carried out in collaboration with colleagues in my research group. Where this is the case, I have acknowledged it in text.

Jordan Portman

Table of Content

List of Tables, Figures and Appendices

Abbreviation

APAP	Acetaminophen
ALF	Acute liver failure
Alb	Albumin
Ass1	Argininosuccinate synthase 1
bp	base pairs
BD	Becton Dickinson
BC	Bile canaliculi
Bmp7	Bone morphogenetic protein 7
CCl_4	Carbon tetrachloride
Cdc6	Cell division cycle 6
Cklf	Chemokine like factor
Cdt1	Chromatin licensing and DNA replication factor 1
CLD	Chronic liver disease
Ccne2	Cyclin E2
CDKs	Cyclin-dependent kinases
Cyp1a2	Cytochrome P450 family 1 subfamily A member 2
Cyp2e1	Cytochrome P450 family 2 subfamily E member 1
DAMPS	Damage associated molecular patterns
DLL	Delta Like
DGE	Differential gene expression
Top2a	DNA Topoisomerase II Alpha

Cdh1	E-cadherin
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
Epcam	Epithelial cellular adhesion molecule
ECM	Extra cellular matrix
FGFs	Fibroblast growth factor
FISH	Florescence in-situ hybridisation
FACS	Fluorescence-activated cell sorting
Gmnn	Geminin
Glul	Glutamate-Ammonia Ligase
GS	Glutamine synthetase
Hells	Helicase, Lymphoid Specific
HB EGF	Heparin Binding epidermal growth factor
HPCs	Hepatic progenitor cells
HSCs	Hepatic stellate cells
HCC	Hepatocellular carcinoma
HGF	Hepatocyte growth factor
Hamp	Hepcidin antimicrobial peptide
Hal	Histidine ammonia-lyase
HIF2a	Hypoxia inducible factor 2a
Igfbp2	Insulin like growth factor binding protein 2
Icam2	Intercellular adhesion molecule 2
IFN-γ	Interferon-γ1b

IL	Interleukin
JAG	Jagged
Krt19	Keratin 19
KCs	Kupffer cells
Lgr	Leucine-rich repeat-containing G-protein coupled receptor
LSECs	Liver sinusoidal endothelia
LR	Logistic regression
Mup3	Major urinary protein 3
Mki67	Marker of proliferation Ki-67
NAPQI	N-acetyl-p-benzoquinone imine
Npr2	Natriuretic peptide receptor 2
NPCs	Non-parenchymal cells
NHF1	Normal human fibroblasts
PoD	Paracetamol overdose
PPARγ	Peroxisome proliferator-activated receptor γ
PBS	Phosphate buffered saline
Pck1	Phosphoenolpyruvate carboxykinase 1
Pdgfrβ	Platelet derived growth factor receptor beta
Pecam1	Platelet endothelial cell adhesion molecule 1
PCA	Principle component analysis
Ptprc	Protein tyrosine phosphatase receptor type c
RSPO	R-Spondin
RLE	Relative log expression

RB	Retinoblastoma protein
RT-PCR	Reverse transcription polymerase chain reaction
RT	Room temperature
scRNA-seq	Single cell RNA sequencing
smFISH	Single-molecule fluorescence in situ hybridization
SEM	Standard error of the mean
TF	Transcription factors
Tf	Transferrin
TGF	Transforming growth factor
Ttr	Transthyretin
TNF	Tumour necrosis factor
TACE	Tumour necrosis factor converting enzyme
UMAP	Uniform manifold approximation and projection
UMIs	Unique molecular identifiers
uPA	Urokinase plasminogen activator
VEGF	Vascular endothelial growth factor
WT	Wild type
ZT	Zeitgeber time

1 Introduction

1.1 Liver Disease – The Global Burden

Approximately two million people die each year from liver disease (Asrani *et al.*, 2019). Although the term liver disease can encompass many different aetiologies, they can be roughly categorised into two groups, acute and chronic. Acute liver failure (ALF) is defined by a rapid loss of liver function, usually in a person with no pre-existing liver disease. Chronic liver disease (CLD) is a condition where iterative liver injury results in fibrosis (scarring), and if injury is ongoing, eventually cirrhosis.

Viral hepatitis and drug overdose are two of the leading contributors to ALF depending on geographical location (Bernal and Wendon, 2013). Infections of hepatitis A, B and E are more prevalent in developing countries, where there is less access to vaccination and good sanitation. Conversely, drug overdose is the greater contributor to ALF in developed countries. Acetaminophen accounts for ~60% and ~40% of acute liver failure in the UK and USA respectively.

Principle causes of CLD include viral infections, obesity and alcohol consumption (Asrani et al., 2019). In 2019 acute hepatitis accounted for 145,000 deaths while cirrhosis caused ~1.16 million deaths. Chronic liver failure and liver cancer taken together accounted for 3.5% of all deaths worldwide in 2019 (11th and 16th most common cause of death, respectively), with 90% of liver carcinomas developing in those with cirrhosis (Asrani et al., 2019; Seitz and Stickel, 2006). This figure is likely to be an underestimate; firstly, even though the numbers are small in comparison, this figure does not account for acute cases (Asrani et al., 2013). Furthermore, recording of deaths are known to underestimate liver disease and cirrhosis.

These figures become even more sobering when looking closer to home. Year-on-year mortality due to liver disease has been increasing in the UK (Williams et al., 2014). Figures from 2010 have shown that standardised

mortality due to liver disease has increased by over 400% since 1970 (Figure 1.1). It is now the most common cause of death in those aged between 35-49, and is the third biggest cause of premature deaths behind ischaemic heart disease and self-harm (Marshall et al., 2019; Williams et al., 2014). Furthermore, over 60% of adults are currently classified as obese or overweight which will no doubt lead to an increase in the liver disease burden (Baker, 2019; Williams et al., 2018). The impact of lifestyle goes beyond our health, with statistics from England and Wales showing alcohol misuse and obesity costing the NHS £21 billion and £27 billion, respectively, each year (Williams et al., 2018).

With 90% of all liver disease in the UK being attributed to obesity, excessive alcohol consumption and viral hepatitis, this disease is preventable for many (Verne, 2014). Balanced diet, moderate alcohol consumption and effective vaccination schemes are all simple preventative measures against this disease. In addition, strides have been made to further tackle these problems including the introduction of sugar taxes and minimum unit pricing of alcohol. However, it may be many years before we are able to comment on the full effectiveness of these solutions.

While prevention is the ultimate goal, there will still be a need for effective treatment of liver disease. Currently, liver transplantation is the only treatment for end stage liver disease. Globally liver transplantation is the second most common solid organ transplant performed, and currently only 10% of the global needs are met (Asrani et al., 2019). This highlights a grave need for alternative treatments and therapeutics to cure liver cirrhosis.

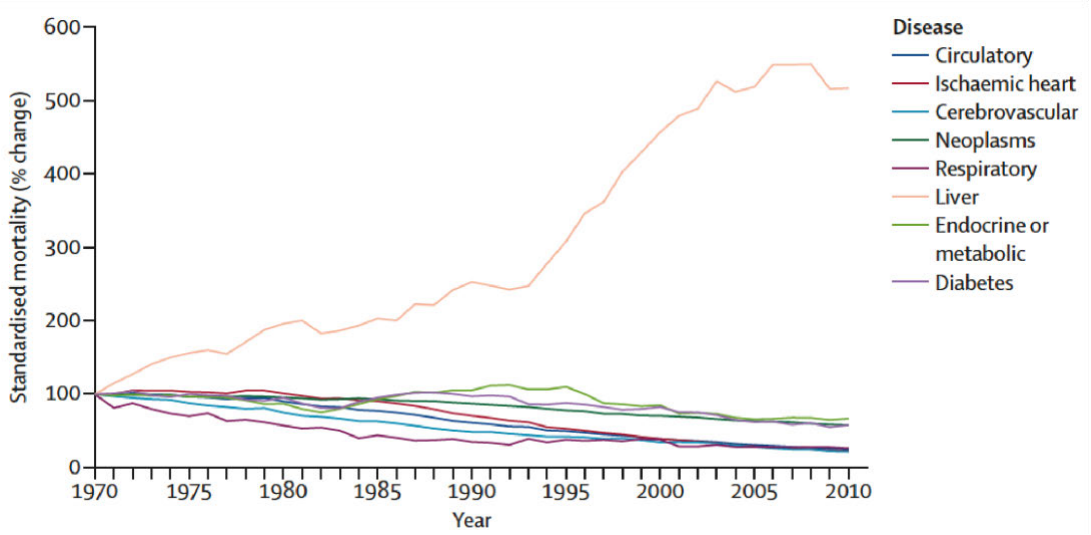


Figure 1.1 – Standardised UK Mortality Rate of Disease in People Under 65 years old

Data shows changes in mortality rate from 1970-2010 for Circulatory, Ischaemic heart, Cerebrovascular, Respiratory, Liver, Endocrine or Metabolic diseases, Neoplasms and Diabetes, normalised to 100% in 1970 . Reprinted from The Lancet (Williams *et al.*, 2014) ©2014, with permission from Elsevier.

1.2 Liver Biology

1.2.1 General Liver Function and Anatomical Structure

The liver performs a myriad of key physiological functions critical for an organism's survival. Such processes include; macronutrient metabolism and storage, bile secretion, detoxification of toxins, cholesterol and lipid homeostasis (Stanger, 2015; Trefts et al., 2017). The functions of the liver can be broadly divided into three subgroups; bile secretion, synthetic function and detoxification.

These functions are generally conserved across all mammals. There are, however, some anatomical changes between species, although these result in only small quantitative differences (Kruepunga et al., 2019). For instance although divided into 4 main lobes: the left and right lobe, the caudate lobe and the quadrate lobe , the human liver has a non-lobulated architecture (Allen, 2009; Kruepunga et al., 2019). On the other hand, the mouse liver is lobulated due to the presence of fissures. It also adopts a 4-lobe nomenclature, with the presence of the left and right lateral, medial (which can be sub divided further into left and right) and caudate lobes (Gross, 2005; Kruepunga et al., 2019). A 1:1 ratio of lobes between mouse and human is not present and the terminology only partially correlates. Lobes of both species can be further divided into a second microscopic functional unit called the hepatic lobule, the structure of which is described in 1.2.3 - Lobular Architecture (Sasse et al., 1992). The microarchitecture of the lobules is largely conserved amongst mammals (Kruepunga et al., 2019).

1.2.2 Cellular Constituents and Function

1.2.2.1 Hepatocytes

Hepatocytes are the main parenchymal cell population, constituting ~60-70% of the liver by number, or ~80% by mass (Stanger, 2015; Taub,

2004). These cells are powerhouses of metabolism and synthesis. Hepatocytes produce large quantities of albumin as well as bile acids, a host of other serum proteins, and can perform glutamine synthesis, urea formation and glycogenesis (Stanger, 2015). Furthermore, hepatocytes contain a wealth of cytochrome P450 (detoxifying) enzymes, used to neutralise a variety of chemicals and sequester them for elimination. These functions are not homogenous across all hepatocytes (Jungermann and Keitzmann, 1996). Hepatocytes in the periportal regions perform different functions than those that reside in the pericentral zone (1.2.4 - Lobular Zonation).

A further source of hepatocyte heterogeneity is their difference in ploidy and nuclear number (Donne et al., 2020). Unlike most cell types, hepatocytes undergo polyploidization during postnatal growth, and polyploid hepatocytes can exist with either one or two nuclei. In rodents, the level of polyploidization is high, with up to 90% of adult hepatocytes characterised as polyploid, while in humans this figure is lower at around 30%. Moreover, this process can also lead to aneuploidy in hepatocytes, which has been suggested to be an adaptive response to stress, aiding liver recovery (Duncan et al., 2012).

1.2.2.2 Cholangiocytes

Cholangiocytes, also known as biliary epithelial cells, are the second parenchymal cell type of the liver. These cells make up between 3-5% of the liver cell number and line the intrahepatic and extrahepatic biliary ducts (Glaser et al., 2009). Although they are less metabolically active in comparison to hepatocytes they still serve key metabolic functions (Stanger, 2015). These include modification of bile composition through absorption of water, electrolytes and other organic solutes (Glaser et al., 2009). Notably, one of its main functions is the production and secretion of bicarbonate into the bile (Maroni et al., 2015).

1.2.2.3 Nonparenchymal cells

The remaining cell populations are termed nonparenchymal cells which consist of mesenchymal cells, endothelial cells and macrophages. Each serve a distinct function. The mesenchymal population can be split into hepatic stellate cells (HSC), portal fibroblasts, vascular smooth muscle cells and mesothelial cells. HSC, during homeostasis, are relatively quiescent, and function as sites of vitamin A storage (Dobie and Henderson, 2016; Wells, 2014). The role of the portal fibroblast is contested but seems mainly to produce and deposit extracellular matrix (ECM) to provide support for the portal triad (1.2.3 - Lobular Architecture). The liver endothelial populations can also be split into several sub-groups; liver sinusoidal endothelial cells (LSECs), vascular endothelial cells (hepatic artery, portal and central veins) and lymphatic endothelial cells. LSECs are a specialised endothelial subtype which have 'fenestrae' and lack basement membrane (Poisson et al., 2017). This produces a highly permeable barrier between the endothelial lumen and the parenchyma, allowing the diffusion of nutrients, macromolecules and metabolites. Lastly, the Kupffer cell (KCs, resident liver macrophage) make up the majority of the liver macrophage population and serve as 'janitors' of the liver, clearing circulating microorganisms, dead cells, endotoxins and debris, thereby preventing infection (Dixon et al., 2013).

1.2.3 Lobular Architecture

As alluded to earlier, the liver can be subdivided into structural/functional subunits. Over decades of study, hepatologists have come up with three main ways that these subunits can be specified: the classical lobule, the portal lobule and the acinus (Rappaport et al., 1954; Sasse et al., 1992). The "hexagonal" classical lobule architecture is derived from focusing on blood supply, whilst the "triangular" portal lobule view is the converse of this, focusing on the counter flowing secreted bile (Adrian, 2020). Lastly, the "elliptical" liver acinus focuses on the zonal metabolism

arrangement of the liver and is the smallest subunit. While all three descriptions are correct, the classical lobule is the most widely used as a structural subunit, while the acinus is considered the more functional subunit. Due to its ease of conceptualising and for consistency, the classical lobule will be used throughout this thesis to describe both the structure and zones of the liver.

The liver lobule is roughly hexagonal in shape centring on the central vein (Figure 1.2 A). Hepatocyte cords radiate out to the periphery with the vertices of each hexagon denoted by the portal triad, consisting of the portal vein, hepatic artery and the bile ducts (Ishibashi et al., 2009; Trefts et al., 2017). The fenestrated endothelia run adjacent to the chords forming sinusoids, with KCs located inside the sinusoidal lumen (Figure 1.2 B). The space between the hepatocyte and the sinusoid is referred to as the Space of Disse and is where the HSC reside. Capillary like structures called bile canaliculi (BC) run parallel to the sinusoids transporting bile from hepatocytes to the bile ducts via the Canals of Hering (Boyer, 2013). BC are formed by the apical domain of adjacent hepatocytes which are contiguous with one another and are sealed by tight junctions. Although the biliary and vascular systems run parallel, flow within them runs in opposite directions. Oxygen rich blood from the hepatic artery (~20% of flow) and nutrient rich blood from the portal vein (~80% of flow) enter the liver from the portal triad and drain through the sinusoids before exiting the liver through the central vein (Jungermann and Keitzmann, 1996). Conversely, bile is secreted into the BC and flows towards the portal triad to then be drained out of the liver by the bile ducts. The arrangement of blood flow through the lobule causes the composition of the blood to change. As oxygen and nutrients are taken up and metabolites and waste secreted, the blood that leaves the liver is more hypoxic and nutrient poor than that which enters. This change in composition results in a partitioning of function in the liver that has become commonly known as liver zonation.

1.2.4 Lobular Zonation

The liver lobule can be divided into three main regions; the periportal region (Zone 1) which surrounds the portal triad, the pericentral/perivenous zone which encompasses the region around the central vein (Zone 3), and the intermediary zone which sits between the pericentral and periportal zones (Zone 2). Many of the liver's functions have been shown to be zonally regulated (Figure 1.2 A&C). For instance, carbohydrate catabolism and anabolism has been one major focus in this area. The periportal regions have been linked to glucose production through gluconeogenesis while the pericentral region utilises glucose through glycolysis and glycogen synthesis (Jungermann and Kietzmann, 1997). The periportal hepatocytes also produce glycogen but use lactate over glucose (Bartels et al., 1987). Similarly, the zonation aspects of fatty acid (Schleicher et al., 2015), amino acid (Brosnan and Brosnan, 2009) and xenobiotic metabolism (Lindros, 1997), plus bile formation (Boyer, 2013), have been delineated.

While several different models have been put forward to explain these zonal patterns, by far the most comprehensive view is the post-differentiation concept. This describes how gradients of morphogens, hormones, growth factors, oxygen and nutrients work in concert to pattern the livers function through both pre- and post-translational modulation (Jungermann and Keitzmann, 1996). For example, the portal to central, high to low, oxygen gradient has been shown to regulate carbohydrate metabolism, with high oxygen levels increasing enzymes involved in gluconeogenesis (Nauck et al., 1981). Wnt/β-catenin signalling is considered as the main driver of central to portal zonation, with central associated functions such as glutamine synthetase and cytochrome P450 enzymes expression dependent upon β-catenin signalling (Gebhardt and Hovhannisyan, 2009; Gebhardt and Matz-Soja, 2014). Conversely, Adenoma Polyposis Coli protein (APC) forms the inverse gradient and is a negative regulator of β-catenin (Benhamouche et al., 2006). Furthermore, preliminary findings from Kietzmann (2017) have indicated increases in mRNA of leucine-rich repeat-containing G-protein

coupled receptor (LGR) 5 due to hypoxia. Taken together with the findings of Planas-Paz *et al.* (2016) which show R-Spondin (RSPO) and LGR4/5 as positive regulators of Wnt signalling in the liver (through Wnt receptor stabilisation), this suggests a possible interaction between the oxygen and Wnt/ β-catenin gradients.

Although the description of metabolic zonation sounds static, it is in fact dynamic, with perturbations of nutrient, oxygen gradients or changes in physiology all having impacts on gene expression and subsequently metabolic zonation (Jungermann and Keitzmann, 1996; Racine-Samson et al., 1996). There are, like most biological systems, caveats to this with some processes that are in fact static. One such example is glutamine synthesis which only occurs in ~3 rows of hepatocytes surrounding the central vein (Gebhardt and Mecke, 1983). However, Planas-Paz *et al.* (2016) showed that injecting RSPO1 into control mice causes a central to portal expansion of glutamine synthetase (GS) expression, suggesting glutamine synthesis may be dynamic under certain conditions.

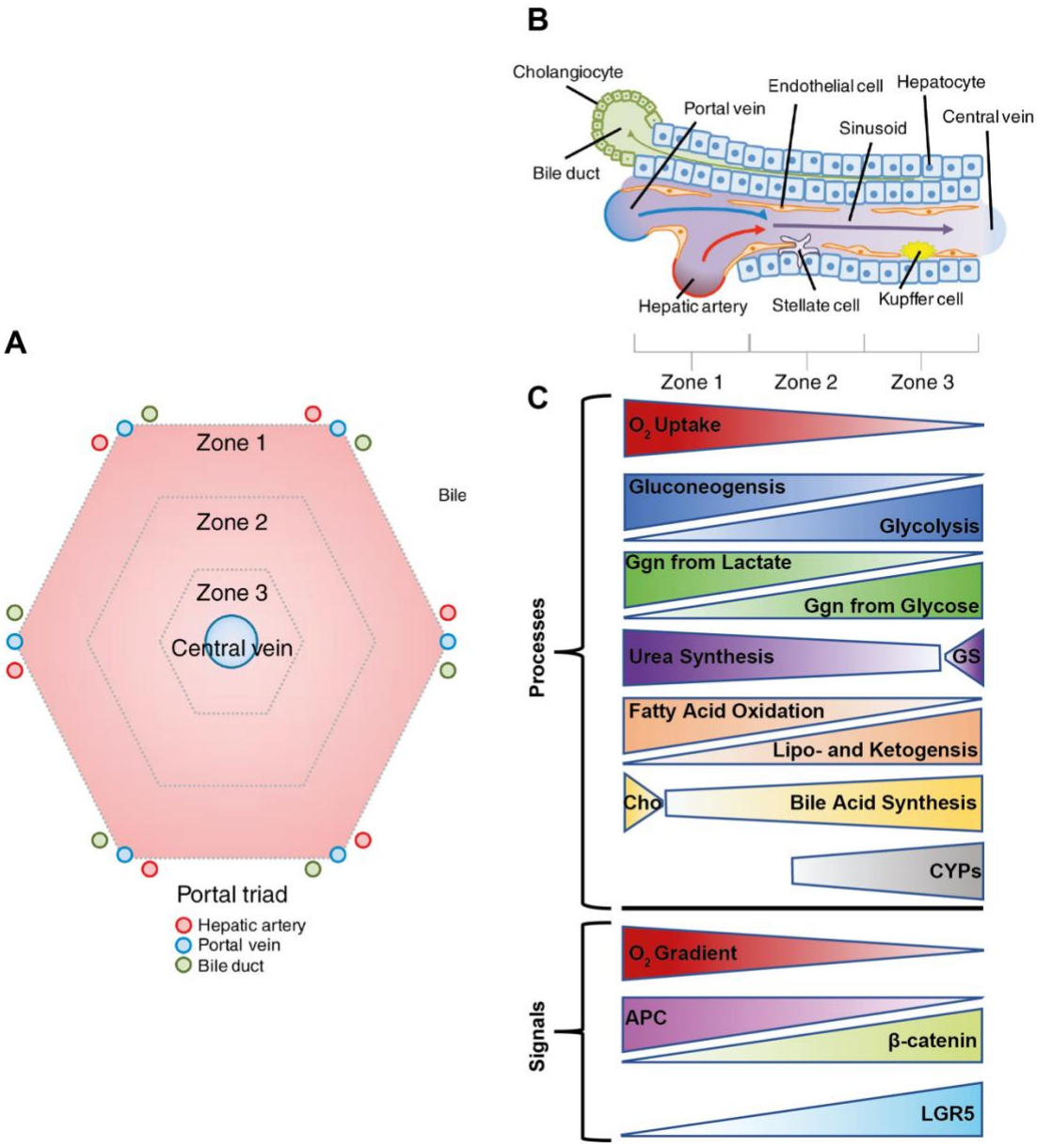


Figure 1.2 – Lobule Organisation and Zonation

(A) Hexagonal representation of the liver lobule centred on the central vein. Zones are depicted radiating out towards the vertices where the portal triad resides. Each triad contains a portal vein, hepatic artery and bile duct.

(B) Diagrammatic cross section showing a single sinusoid and bile canaliculi (not labelled). Nutrient rich venal blood and oxygen rich arterial blood flow from the periportal region (Zone 1) into the sinusoid, eventually draining into the central vein in the pericentral region (Zone

3). All major parenchymal (hepatocytes, cholangiocytes) and non-parenchymal (endothelia, Kupffer cells and HSC) cells are labelled.

(C) Schematic of the distribution of the major metabolic pathways and the signalling pathways that form these distributions. O_2 – Oxygen, Ggn, Glycogen Synthesis, GS – Glutamine Synthesis, Cho, Cholesterol Synthesis, CYPs – Cytochrome P450 enzymes (Xenobiotic Metabolism), APC - Adenoma Polyposis Coli, LGR5 - Leucine-rich repeat-containing G-protein coupled receptor.

Adapted from Trefts, Gannon and Wasserman (2017) ©2017, with permission from Elsevier. (A) & (B) are direct reprint from the publication with (C) being my sole alteration.

1.3 Liver Regeneration

The capacity of cellular regeneration varies across different tissues. The epithelial tissue of the skin and intestines have a high turnover during homeostasis and are dependent on a stem cell pool. On the other hand, the heart and brain exhibit minimal cellular turnover. Liver regenerative capacity is context dependent. In homeostasis, minimal cellular turnover is observed, with less than 1-2% of hepatocytes in cell cycle at any given time, and each cell having a lifespan between 200-400 days (Macdonald, 1961; Stanger, 2015). However, following injury, the liver exhibits a remarkable capacity to regenerate. This is potentially an evolutionary adaptation brought about by the critical nature of hepatic function. Physiologically this regenerative process will occur mainly due to exposure to toxins, however, the extent of the liver's capacity to regenerate is seen more prominently during surgical removal of a portion of the liver (Stanger, 2015). In rodents that have undergone 2/3rd partial hepatectomy (PHx), the liver can regenerate it's mass in less than two weeks, through hepatocyte proliferation and hypertrophy (Mitchell and Willenbring, 2008). The same rapid regenerative capacity is also exhibited in humans, although the process takes slightly longer, between 6-8 weeks (Katoonizadeh, 2017). In certain conditions when hepatocyte proliferative ability is blocked

(senescense), such as experimentally or in chronic liver disease, hepatic progenitor cells (HPCs) located in the canals of Hering, have been shown to proliferate and give rise to hepatocytes and ductular cells (Lu et al., 2015). However, the most common mechanism of liver regeneration, and the main focus of this thesis, is through the proliferation of fully differentiated hepatocytes (Michalopoulos, 2017).

The liver regenerative process is complex with cellular components and signalling changing depending on the insult that initiates the regeneration. There are therefore a multitude of rodent models that are used to study this process *in vivo*.

1.3.1 Models of Liver Regeneration

There many different pre-clinical mouse models of liver regeneration, each trying to recapitulate a different condition which triggers the regenerative process. These models make for ideal systems to study the cellular and molecular mechanisms that occur during liver regeneration.

1.3.1.1 Partial Hepatectomy

Partial hepatectomy (PHx) is one of the most widely used model of liver regeneration in rodents, whereby 2/3^{rds} or ~70% of the liver mass is surgically removed. This model was originally described by Higgins and Anderson (1931) and was later standardised by Mitchell and Willenbring, (2008). Due to the lobulated architecture of the mouse liver, liver lobes can be easily separated and resected. For the standard PHx model, the left lateral and medial lobes are removed. This allows for a “clean” removal of the liver with the regeneration process initiating within the first five minutes after the resection, and full liver mass regeneration occurring over the next 7-10 days. The process involves minimal involvement of any necrotic or acute inflammatory response and is mediated by crosstalk between the parenchymal and non-parenchymal (NPCs) populations. Although this is referred to as a regenerative process, it

is not true regeneration but compensatory hyperplasia. Rather than the resected lobes growing back, the remnant lobes enlarge through hypertrophy and replication of the hepatic cellular populations. During 70% PHx, hypertrophy precedes proliferation of hepatocytes while during a 30% PHx, the regained liver mass is mainly attributed to hypertrophy of the parenchyma alone, with very few cells entering S phase (Miyaoka et al., 2012).

The mechanism of 70% PHx (hereby referred to as PHx) can be broadly split into three main stages: 1) initiation or priming of hepatocytes, 2) the proliferative stages, and 3) the termination of cellular proliferation (Rmilah et al., 2019). The initiation event, occurring between 0-5h post-surgery, is marked by a rapid increase in gene expression of over 100 different genes. These include the production of growth factors and mitogenic signals, such as hepatocyte growth factor (HGF) and epidermal growth factor receptor (EGFR) ligands, that are needed to stimulate the various cell populations to enter the cell cycle from their quiescent state.

During the proliferative phase these signals continue to persist until the liver has regenerated and regained its lost mass. Hepatocytes are the first cell type to undergo proliferation with peak proliferation said to occur between 36 - 48h in mice (Mitchell and Willenbring, 2008; Zou et al., 2012). This variation could be due to differences in mouse strain and method of measurement of hepatocyte proliferation. Peak proliferation of the remaining cell populations is around 1-2 days after that of hepatocytes, for instance, endothelial peak proliferation is around day 3 (Michalopoulos, 2007, 2017). Paracrine signalling between hepatocytes, cholangiocytes and the other NPCs is crucial throughout the proliferative phase to maintain the regenerative response.

The initial kinetic theory of hepatocyte proliferation suggested that almost all hepatocytes enter the cell cycle and complete cell division at least once, with a second round of cell division occurring in a smaller subset of hepatocytes (Rmilah et al., 2019). This equates to each hepatocyte undergoing 1.6 divisions. However, recent evidence suggests that although the majority of hepatocytes enter S phase, only around half undergo cellular division, while the rest do not divide and therefore increase their ploidy

(Miyaoka et al., 2012; Miyaoka and Miyajima, 2013). This gives a reduced figure of only 0.7 divisions per hepatocyte. The mechanisms behind hepatocytes not entering M phase are still not fully understood, neither are the signals that causes the difference between each hepatocyte's potential to complete M phase.

The final stage, termination, is just as vital as the initiation phase, as this prevents excess proliferation of hepatocytes which could result in carcinogenesis (Liu and Chen, 2017; Michalopoulos, 2007). This stage is usually denoted by a small wave of apoptosis to account for "over shooting" of the proliferative response (Sakamoto et al., 1999). Again, as with the first two phases, there are a multitude of signalling pathways activated, however those involved at this stage are less well studied. This is also somewhat true for the proliferative phase. While many of the signalling pathways established during the initiation phase perpetuate during the proliferation phase, the signals that control the temporal and spatial patterns of replications are still not fully understood and require further investigation. Many of the current known signalling pathways are discussed later in section 1.4 – Cellular and Molecular Mechanisms following Partial Hepatectomy.

1.3.1.2 Acetaminophen

Acetaminophen (APAP), or more commonly referred to as paracetamol in the UK, is a well-tolerated drug with minimal side effects when administered at therapeutic dose. However, overdose from APAP (henceforth referred to as paracetamol overdose [PoD]), whether accidental or intentional, can cause severe liver damage, resulting in ALF and potentially death. As discussed in section 1.1, PoD is the leading cause of ALF in the western world.

Similar to PHx, PoD can be divided into three main stages although each stage is not comparable (Bhushan and Apte, 2019). The first stage, known as the initiation phase, is characterised by rapid metabolism of APAP into N-acetyl-p-benzoquinone imine (NAPQI) by cytochrome P450 family 2 subfamily E member 1 (CYP2E1; Zaher *et al.*, 1998). NAPQI is quenched by

glutathione (GSH) however, the rapid production of NAPQI quickly depletes the GSH stores within hepatocytes and exceeds the production rate of new GSH (Jaeschke et al., 2014; Mitchell et al., 1973). The excess NAPQI forms cellular protein adducts within the mitochondria, resulting in dysfunction of the organelle and the release of reactive oxygen species (ROS; Bhushan and Apte, 2019). Furthermore, activation of intracellular signalling molecules such as c-Jun N-terminal kinase, exacerbates the situation by increasing mitochondrial permeability and subsequent release of endonucleases (Hanawa et al., 2008; Henderson et al., 2007; Jaeschke et al., 2012). These events ultimately lead to cell death by necrosis, specifically in the hepatocytes in the pericentral region. This localised cell death is due to the zonation of the CYP2E1 protein which is found at higher levels in the pericentral regions.

The initiation phase is followed by the progression phase, and collectively these two phases are described as the injury phase. During the progression of PoD, necrotic cells release damage associated molecular patterns (DAMPS) such as high-mobility group box 1 protein as well as DNA fragments (Jaeschke et al., 2014). The release of such molecules causes a strong sterile immune response, which is seen throughout the progression of PoD induced liver injury and is vital for the recovery of the liver. Cell death also continues during this stage as a result of the proteolytic enzymes such as calpains released from dying hepatocytes.

Regeneration of the necrotic area begins during the recovery phase through hepatocellular proliferation. However, injury is only resolved if a robust regenerative response is achieved. In cases where the injury is too severe, the regenerative response cannot cope with the amount of cell death and the liver acutely fails, followed by multiorgan failure and death (Mehendale, 2005). For effective recovery of the liver, the necrotic debris needs to be cleared by macrophages followed by the replication of the hepatocytes surrounding the necrotic region. Some of the mitogenic signals that promote hepatocytes to enter into the cell cycle are shared across PHx and PoD, such as EGFR ligands, however their exact mechanism of action can differ (Bhushan et al., 2014). It is not surprising that these signalling mechanisms between PHx and

PoD differ, as the latter manifests a much more extensive immune response (Bhushan and Apte, 2019). Furthermore, the proliferation kinetics of hepatocytes also differ, with all hepatocytes proliferating in a synchronous manner during PHx, while only those hepatocytes surrounding the necrotic region proliferating in PoD in an asynchronous manner.

While PoD is a clinically relevant mouse model, there can be differences in injury extent, temporal dynamics and mechanistic response between mice within an single study and between studies (Mossanen and Tacke, 2015). Age, sex, fasting period and methods of APAP administration can all potentially introduce variability. For instance, female mice are more resistant to APAP hepatoxicity than male mice given the same dose of APAP.

1.4 Cellular and Molecular Mechanisms following Partial Hepatectomy

Study of the mechanisms of liver regeneration following PHx have been ongoing for many decades. A large pool of signalling molecules and pathways have been shown to be involved and important for the regenerative process (Michalopoulos, 2007; Rmilah et al., 2019). Understanding the interactions and identifying the key molecular players in this intricate process is vital for the generation of novel therapeutics. The pool of known signalling molecules that are involved can be split into two main groups based on their ability to cause proliferation of hepatocytes. These are the complete mitogens and the auxiliary mitogens (Michalopoulos, 2010). Complete mitogens are the growth factors that can initiate replication in cultured hepatocytes with chemically defined serum free media. Such growth factors will also cause liver enlargement and hepatocyte DNA synthesis when injected *in vivo* in sufficient quantities. Currently the only two known complete mitogens are hepatocyte growth factor (HGF)/c-Met receptor, and ligands of the epidermal growth factor (EGF) receptor (EGFR). Auxiliary mitogens cause delays in liver regeneration when removed, but do not abolish the whole regenerative process. They also fail to cause DNA synthesis in hepatocytes when administered alone, either to

hepatocytes in culture or when injected *in vivo*. These mitogens are seen to enhance the effects of the compete mitogens. The list of these molecules is much larger than the complete mitogens, and include; interleukin 6 (IL-6; Cressman *et al.*, 1996), Notch and Jagged (Köhler et al., 2004), norepinephrine (Cruise et al., 1987), tumour necrosis factor (TNF; Yamada *et al.*, 1998), insulin (Cruise et al., 1985), bile acids (Huang et al., 2006), serotonin (Lesurtel et al., 2006), vascular endothelial growth factor (VEGF; LeCouter *et al.*, 2003) and fibroblast growth factor (FGFs; Steiling *et al.*, 2003). FGFs and recombinant Jagged 1 are somewhat unique by the fact they can induce some degrees of proliferation in cultured hepatocytes however, cannot *in vivo* (Michalopoulos, 2010). Further to these two groups, urokinase plasminogen activator (uPA), circulation dynamics and the Wnt/ β-catenin also play an important role in hepatocyte proliferation.

While all molecules mentioned so far initiate and perpetuate regeneration, only TGF-β1 has been shown to be vital in the termination of regeneration. Outlined below is an overview of how some of these signals interact to control hepatocyte proliferation/regeneration and will be evaluated temporally. A generalised overview can be found in Figure 1.3.

1.4.1 Circulation and Urokinase Plasminogen Activator

During 2/3rds PHx there are large changes in hepatic blood flow. Arterial blood supply per unit of liver is not altered, however the portal blood supply per unit triples (Michalopoulos, 2007). This is because the entire blood supply from the portal vein now traverses a liver that is 1/3rd of the size it was pre-PHx. Shunting of the portal blood supply coincides with a reduction in HGF activation and increased apoptosis of hepatocytes, suggesting that increased portal pressure is a key initiating response for liver regeneration following PHx (Marubashi et al., 2004).

Alongside the initial increase in hemodynamic pressure, uPA is one of the first molecules known to increase in activity during PHx, starting as early as five minutes post-surgery (Mars et al., 1995). uPA is a known activator of

matrix remodelling and is seen in most tissues during wound healing (Michalopoulos, 2007). uPA acts to convert inactive plasminogen to active plasmin, which can facilitate the breakdown of fibrinogen into fibrinogen degradation products (Kim et al., 1997; Tanaka et al., 2001). Furthermore, a second matrix altering protein metalloproteinase 9 (MMP9) is also increase early in PHx (Olle et al., 2006). In homeostatic liver, HGF is inactive and bound to the ECM, but the increase in these matrix altering proteins leads to the remodelling of the ECM early in PHx and the subsequent release of HGF (Mars et al., 1995). uPA now functions to activate the single chain HGF to its heterodimeric form (Mars et al., 1993). HGF then acts locally and systemically as one of the main mitogens for liver regeneration.

Data in human coronary artery endothelial cells also suggests that circulatory changes could increase uPA activity. Sokabe *et al.* (2004) showed that increases in turbulent shear stress results in an increase in uPA mRNA half-life and stability, as well as an increase in protein secretion in human coronary artery endothelial cells. Further work is required to assess to what extent this may be a relevant mechanism regulating liver regeneration following PHx. Furthermore, changes in oxygen and nutrient supply also drastically change during PHx and this is currently an understudied area of the model.

1.4.2 Complete Mitogens

1.4.2.1 Hepatocyte Growth Factor and c-Met

HGF during homeostasis is stored in relatively large quantities in the ECM of the liver. During PHx these stores are mobilised through ECM remodelling causing a significant increase in HGF plasma levels (Lindroos et al., 1991). In the rat model, these intrahepatic stores are consumed within the first 3h followed by new HGF synthesis (Pediaditakis et al., 2001). In mice, which lack HGF expression through conditional knock out (KO), the HGF protein and mRNA stores prior to KO were sufficient to sub serve at least one

regenerative response, be that either PHx or a single CCl_4 injection (Nejak-Bowen et al., 2013). HSC have been shown to sense HGF levels and it is speculated that HGF mRNA is long-lived in homeostasis. However, following injury, HGF protein in the ECM is rapid depleted. HSC will therefore utilise pre-existing mRNA to produce new HGF, resulting in degradation of HGF mRNA. Subsequent insult to the liver would then require *de novo* synthesis of HGF which would is not possible in the KO mice, leading to compromised liver regeneration. Production of new HGF is also carried out by liver endothelial cells (LeCouter et al., 2003), however there is some controversy regarding whether mature LSECs are the main HGF producing endothelial cell or if it is in fact the endothelial progenitors (Wang et al., 2012). Reciprocally, VEGF produced by hepatocytes causes an increase in endothelial cell proliferation (Shimizu et al., 2001). Recent evidence has identified hypoxia, acting through hypoxia inducible factor 2a (HIF2a), as a driver of hepatocyte VEGF production (Kron et al., 2016). This would suggest non-vascular regions of proliferating hepatocytes are induced to produce VEGF through HIF2a causing replication of the endothelial cells and the vascularisation of the new region. In turn the endothelial cells then promote proliferation of hepatocytes. In addition to these two intrinsic liver sources of HGF, distal organs, such as lung, kidney and spleen, induce HGF production in response to PHx (Kono et al., 1992; Yanagita et al., 1992).

HGF acts through its receptor c-Met to induce proliferation in hepatocytes. Upon binding to HGF, c-Met will dimerise resulting in its activation through tyrosine phosphorylation (Wang et al., 2020). c-Met activation has been seen within the first 1h post-PHx and results in a myriad of signalling cascades (Stolz et al., 1999). Studies in hepatocytes and other cells types have shown that pathways such as the PI3K, MAPK, NF-κb, and STAT3 are activated by c-Met and in turn lead to proliferation (Borowiak et al., 2004; Rmilah et al., 2019). Deletion of the c-Met receptor causes a delay in liver regeneration, but when combined with EGFR inhibition this completely abolishes the regenerative process (Paranjpe et al., 2016). While promoting

proliferation, the c-Met receptor can also bind to the Fas apoptotic receptor and prevent its trimerization and activation of apoptosis (Wang et al., 2002).

1.4.2.2 Epidermal Growth Factor Receptor and Ligands

Similar to c-Met, elimination of EGFR signalling results in delay but not complete abolishment of liver regeneration unless c-Met is also removed (López-Luque et al., 2016). EGFR signals through many of the same mechanisms as c-Met and becomes activated at a similar time during PHx (Collin De L'hortet et al., 2012; Stolz et al., 1999). The lack of complete regeneration abolishment through removal of only c-Met or EGFR signalling would suggest a compensatory mechanism occurring between these two receptors.

There are four ligands of EGFR that are currently known to play a role in PHx; EGF, TGFα, Heparin Binding EGF (HB EGF) and amphiregulin (Collin De L'hortet et al., 2012). While there is redundancy in this system, each ligand is not acting in a completely similar manner. TGFα deficient mice appear to have normal liver regeneration while those deficient in HB EGF and amphiregulin do have deficient regeneration (Berasain et al., 2005; Mitchell et al., 2005; Russell et al., 1996). Currently there has been no genetic KO of EGF, however removal of the known sources of EGF does result in reduced regeneration (Lambotte et al., 1997; Olsen et al., 1988).

EGF is constantly available to the liver through the portal circulation. The Brunner's and salivary glands produce EGF, with removal of either gland resulting in either decreased or delayed regeneration (Lambotte et al., 1997; Olsen et al., 1988). Injection of EGF *in vivo* results in initiation of DNA synthesis in hepatocytes, demonstrating its mitogenic potential (Bucher et al., 1977). However, with the constant supply of EGF to the liver it is not fully understood how EGF can become more mitogenic during injury than homeostasis. One potential answer is due to the reduction of the liver mass. EGF now targets fewer hepatocytes and therefore has enhanced effects. Also norepinephrine, which increases during PHx, enhances production of EGF from the Brunner's

gland (Cruise et al., 1987; Olsen et al., 1985). A second theory involves the increased sensitivity of EGFR to EGF. Cultured hepatocytes have increased DNA synthesis when exposed to norepinephrine, but only in the presence of EGF (Cruise et al., 1985). It is suggested that norepinephrine, acting through the α-adrenoreceptor, increase EGFR sensitivity to EGF. Furthermore, cross talk between c-Met and EGFR could also be a second source of increased sensitivity for EGFR (Jo et al., 2000).

Inactive TGFα is produced by hepatocytes during PHx and is activated through cleavage of the extracellular domain by tumour necrosis factor converting enzyme (TACE; Le *et al.*, 2003; Mead and Fausto, 1989). As hepatocytes produce TGFα, it is speculated that there is an autocrine loop of cellular proliferation. While TGFα deficient mice do not have altered liver regeneration, over expression results in liver enlargement and hepatocyte replication, demonstrating its mitogenic potential for hepatocytes (Webber et al., 1994).

HB EGF mRNA expression is increased in endothelia and KCs in rats 1.5h after PHx (Kiso et al., 1995). When measured, protein expression was also increased 2.8-fold over normal, 10h after PHx. KO of HB EGF in mice causes a delay in regeneration while overexpression accelerates the proliferative response (Kiso et al., 2003; Mitchell et al., 2005).

1.4.3 Auxiliary Mitogens

1.4.3.1 Interleukin 6

Il-6 is an important signalling molecule that has been linked to the increased expression of around 36% of the genes that are activated early in PHx (Li et al., 2001). Il-6 is mainly produced by KC and is believed to be under the control of TNF-α (Aldeguer et al., 2002; Yamada et al., 1997). Il-6 binds to the Il-6 receptor on hepatocytes and forms a complex with gp130 to activate STAT3 signalling (Taub, 2004). This ultimately leads to enhanced proliferation with increases in cyclin D1 expression (Cressman et al., 1996). This was

evident in Il-6 deficient mice which had a defective or delayed regenerative response after PHx, with some mice characterised with liver necrosis and liver failure. Il-6 deficiency also resulted in decreased STAT3 signalling and cyclin D1 expression during PHx. It should be reiterated however that Il-6 is not considered a complete mitogen for hepatocytes, with injected Il-6 having no impact outside of an injury context. Il-6 is considered to prime hepatocytes for replication. However, other studies have shown that regeneration in Il-6 deficient mice is somewhat normal, with deficient mice having increased hepatocyte replication at 24h over wild type (WT) but the converse at later timepoints (Sakamoto et al., 1999). Further studies have shown that Il-6 may act to cause cell cycle arrest and increase DNA repair proteins at early timepoints in PHx (Tachibana et al., 2014). This suggests that Il-6 could be acting as a cell cycle checkpoint protein, ensuring accurate DNA replication.

More recent studies have also shown a link between Yap and Notch signalling with the gp130 receptor in a STAT3 independent manner (Taniguchi et al., 2015). This suggests cross talk between the many signalling pathways involved during PHx. Clearly, Il-6 plays a positive important role in PHx but similar to many of the signalling molecules and pathways, more work is required to fully understand these processes.

1.4.3.2 Wnt/β-catenin

Wnt/β-catenin signalling is activated early during PHx, with a 2.5-fold increase in β-catenin and its translocation to the nucleus (Monga et al., 2001). This subsequently causes the increase in expression of its target genes which include cyclin D1 (Tan et al., 2006). Finding the source of Wnt has been difficult, but it is now believed that LSECs and KCs are at least two sources of Wnt (Ding et al., 2010; Yang et al., 2014). Conditional KO of β-catenin caused a delayed regenerative response (Tan et al., 2006). There was a 2-fold decrease in the number of proliferative hepatocytes at 40h post PHx, which is usually around the time of peak hepatocyte proliferation. Instead there was an increase in proliferation 3 days post PHx, which would suggest that

compensatory pathways were eventually able to account for the loss in β-catenin signalling. A similar response is also observed in mice that lack Wnt co-receptors LRP5/6 or LGR4/5 (Planas-Paz et al., 2016; Yang et al., 2014). Interestingly while Wnt signalling can be pro-regenerative, Wnt5a has been shown to be inhibitory, suppressing β-catenin signalling, in primary hepatocyte cultures (Yang et al., 2015). A Wntless KO (the protein which transport Wnt out of the cell for secretion), showed prolonged proliferation of hepatocytes compared to controls, which the study theorises was a result of decreased Wnt5a secretion.

1.4.3.3 Notch signalling

Notch signalling is comprised of several membrane bound ligands and receptors which are able to co-ordinate the differentiation and proliferation of many different cell types and tissues (Alves-Guerra et al., 2011; Bray, 2016). Mammals have 4 Notch paralogues (Notch 1-4) and two ligand families, delta like (DLL1, 3 & 4) and Jagged (JAG1 & 2). Notch is considered the receptor of this signalling network, binding to either the DLL or JAG ligands, however, increasing evidence is showing that this signalling can be bi-directional with each having both a receptor and ligand role (Alves-Guerra et al., 2011; Metrich et al., 2015; Six et al., 2003). Nevertheless, in the classical view, binding of either DLL or JAG caused the cleavage of the Notch intracellular domain (NICD) by ADAM proteases and γ-secretase (Alves-Guerra et al., 2011; Bray, 2016). This then allows NICD to translocate to the nucleus where it functions as a transcription factor/co-activator to mediate the expression of HES and HEY proteins, and is also believed to control cell cycle genes such as cyclin D1 (Ronchini and Capobianco, 2001; Zhang et al., 2018). Notch-1 and JAG-1 are known to be upregulated during PHx, with NICD seen to translocate to the nucleus of hepatocytes after just 15 minutes post-surgery in rats (Köhler et al., 2004). Increased expression of HES-1 was seen after 30-60 minutes post PHx. Mutations in JAG-1 in humans is known to cause Alagille syndrome, a disorder that effects the liver, heart and other parts of the body (Li et al., 1997). Most

notable however, is the major liver damage and abnormalities in the bile ducts within patients of Alagille syndrome. Deletion of Notch 1 in mice showed a retarded liver regenerative response, with livers of KO mice taking 4 weeks to full regain liver mass as opposed to the 9 days of the control mice (Croquelois et al., 2005). As Notch 1 is embryonically lethal, Notch 1 was deleted after birth using a condition KO system. Interesting, outside the context of injury, mice with Notch 1 deletion showed nodular regenerative hyperplasia, and have an increase liver weight to body ratio. This would suggest a cell cycle inhibitory role of Notch 1. The authors believe that Notch 1 is more of an inhibitory regulator of regeneration and defend this by first showing that while hepatocyte proliferation was reduced in Notch 1 KO mice, the liver to body weight ratio was not significantly different between the control and KO mice. They also show that the remnant liver of the Notch 1 mice makes up a larger percentage of the body weight over the control mice (due to their enlarged livers). They state, "the regenerative response is inversely correlated with the amount of remaining liver after" PHx, and the Notch 1 KO mice have a less vigorous proliferative response because of their larger percentage of remnant liver and not their inability to signal though Notch 1. It appears there is a reasonable argument to suggest that Notch 1 acts as an inhibitor of hepatocyte proliferation rather than a promoter, but further work is required to fully confirm this.

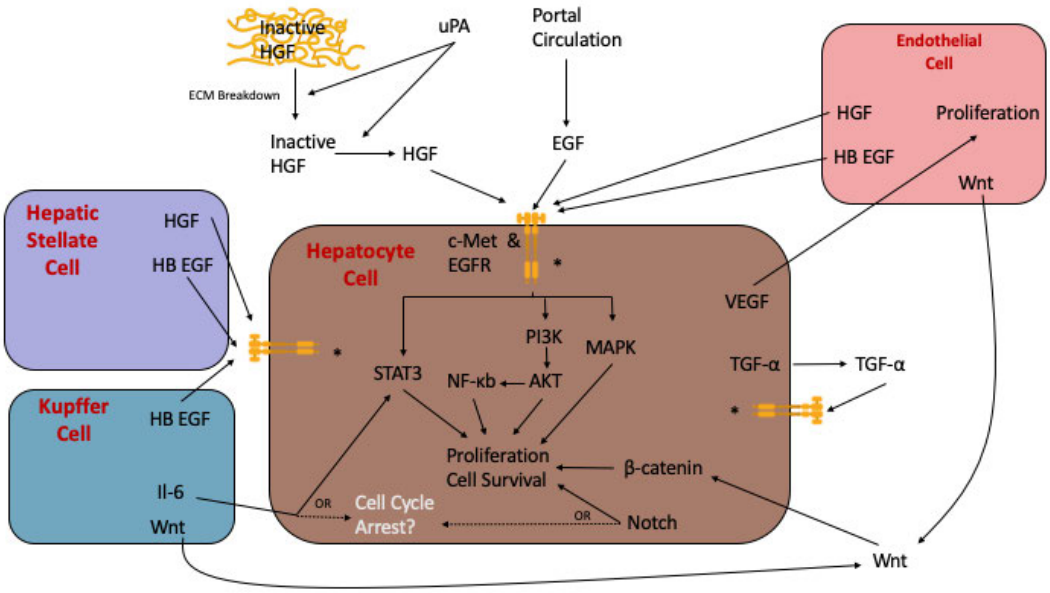


Figure 1.3 – Schematic of Signalling Pathways following Partial Hepatectomy

Summary of signalling pathways following PHx with arrows depicting signalling direction. In some situations, an alternative hypothesis of the function of a signalling molecule is depicted by a dashed arrow. * represents the continuation of the main c-Met and EGFR signalling cascade. Brown cell – Hepatocyte, Pink cell – Endothelial cell, Turquoise cell – Kupffer cell, Lilac cell – Hepatic Stellate cell.

1.4.4 Cell Cycle

The purpose of these pro-regenerative signals is to guide the hepatocyte through the cell cycle. The cell cycle is made up of 4 predominant phases, G_1, in which cells grow in preparation for DNA replication, S phase, in which cells copy their DNA, G_2, which is a second growth phase, and finally, M phase, the phase in which mitosis occurs. During homeostasis, hepatocytes reside in G_0, which is the quiescent phase, or G_1. The transition between each phase is a highly controlled and well-ordered system. This is to ensure that the cell has enough intracellular material to allow replication, and that the DNA of each cell has correctly copied without any errors. The key proteins which are the 'gate keepers' of these phases are the cyclin-dependent kinases (CDKs) (Nevzorova and Trautwein, 2015; Vermeulen et al., 2003). These are a family of serine/threonine protein kinases. These proteins remain stable throughout the cell cycle, whereas their activator proteins, cyclins, do not. Cyclin levels rise and fall at different stages of the cell cycle to activate the correct CDK at the correct time. In mammalian cells, there is sequential activity of at least four different CDKs. These are CDK1, 2, 4, and 6. There are of course various families of cyclins which activate each or multiple CDKs. Cyclin D (D1, D2 ,D3) activates CDK4 & 6, cyclin E (E1, E2) activates CDK2, cyclin A (A1, A4) activates CDK1 & 2, and finally cyclin B (B1, B2, B3) activates CDK1.

During G1, type D cyclins initiate phosphorylation of retinoblastoma protein (RB) through activation of CDK 4/6 (Albrecht, Hu and Cerra, 1995; Borowiak *et al.*, 2004; Jaumot *et al.*, 1999; Figure 1.4). When unphosphorylated, RB acts to repress gene transcription through binding to E2F transcription factors (TFs; Giacinti and Giordano, 2006). These are a family of transcription factors which drive entry into S phase. Now with RB phosphorylated, E2F TFs are able to initiate E2F early responsive cell cycle genes such as cyclin E. CDK2 is then activates by Cyclin E in late G1 and completes the phosphorylation of RB, causing further activation of E2F TFs (Hu *et al.*, 2014; Pujol *et al.*, 2000; Figure 1.4). These cells are then able to pass through the G1/S restriction point into S phase.

Here Cyclin A regulates the DNA synthesis process along with CDK2 (Castro et al., 1994). Usually, once DNA replication has successfully finished, cells enter into G2 where, CDK2 is swapped for CDK1 to prepare the cell for mitosis (Corlu and Loyer, 2012; Nevzorova and Trautwein, 2015). Finally, cyclin B and CDK1 complexes are active through mitosis to complete the process (Figure 1.4). However, evidence has shown that CDK1 also regulates the S phase in hepatocytes, with CDK1 and cyclin B or A complexes shown to be active during this phase (Corlu and Loyer, 2012; Garnier et al., 2009). There is also evidence that CDK1 can compensate for CDK2 in CDK2 KO mice.

The process is also fine-tuned through a group of negative regulator proteins (Vermeulen et al., 2003). As well as RB there is; the INK4 family (p16, p15, p18, p19) which binds to CDK 4/6 and prevents cyclin D activity, and the cyclin dependent kinase inhibitors family (p21, p27, p57) which inhibits CDK2/1 and cyclin E/A/B complex activity (Figure 1.4). p53 specifically is known for its pausing of the cell cycle and recruitment of DNA repair proteins when there is DNA damage. This protein also triggers apoptosis when the damage cannot be repaired, preventing propagation of the incorrect genomic content and therefore preventing possible cancer development. There are of course many more proteins that are layered on top of these mentioned here that control the activity of these proteins. This allows for tight control of the cell cycle through sensing of various problems such as DNA damage or spindle malformations.

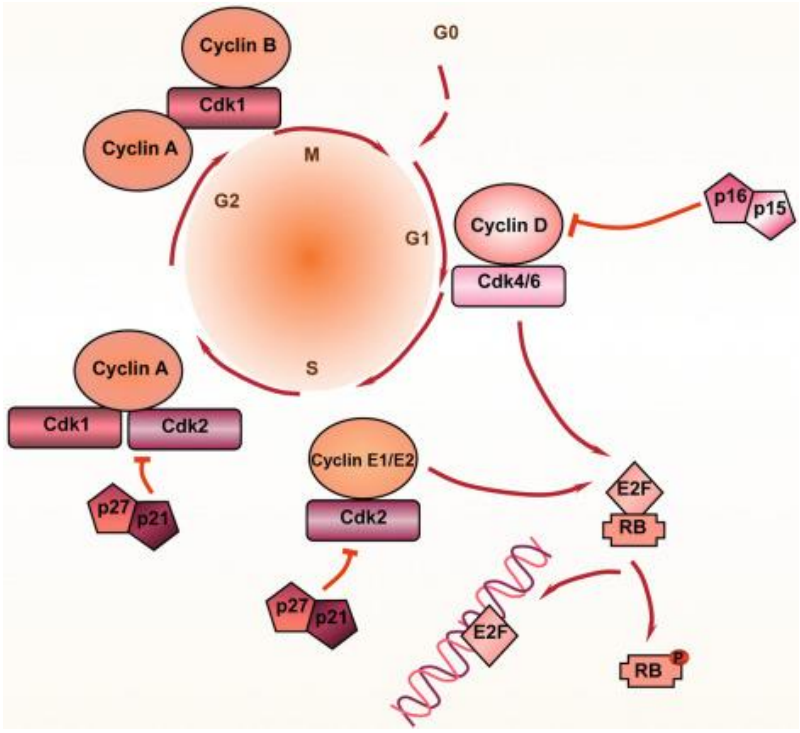


Figure 1.4 – Cell Cycle Control

Specific cyclin and cyclin dependent kinases (CDKs) interact at various points around the cell cycle. Early cyclin and CDK complexes interact with retinoblastoma protein (RB) to causes phosphorylation of RB and release E2F transcription factors which bind to DNA. Examples of the INK4 cyclin/CDK inhibitor family (p15, p16) and cyclin dependent kinase inhibitors family (p27, p21) specifically regulate certain cyclin/CDK complexes. Adapted from (Nevzorova and Trautwein, 2015) ©2015, with permission from Elsevier.

1.4.5 Termination and TGF-β

Whilst there is a deep understanding of the events and molecules that contribute to the initiation and replication of hepatocytes, the mechanisms regulating the termination of regeneration are much less understood. By far the most studied in this area is TGF-β (Michalopoulos, 2007). This molecule is produced by mesenchymal cells and inhibits the epithelia of most tissues. The liver is no different, with HSC being at least one source of TGF-β. TGF-β is a mito-inhibitor for hepatocytes (Houck et al., 1988; Ikeda et al., 1998). Similar to HGF, TGF-β is bound to the ECM and is believed to exert a competing 'tonic' effect against the growth factors that are also bound to the ECM, therefore keeping hepatocytes in a quiescent state (Michalopoulos, 2007). If TGF-β receptor 1 is inactivated by dominant negative DNA constructs, there is a noticeable increase in DNA synthesis, suggesting the 'tonic theory' is a reasonable mechanism of action (Ichikawa et al., 2001). Cleary, during PHx, hepatocytes become resistant to this inhibition in some way. As the ECM becomes remodelled during PHx, the release of the various mitogens shifts the balance in favour of a more proliferative environment. Furthermore, the release of TGF-β1 into the circulation allows it then to be inactivated by alpha-1-macrogobulin and norepinephrine (Houck et al., 1988; Houck and Michalopoulos, 1989; LaMarre et al., 1991, 1990). It should be noted however that the roles of alpha-1-macrogobulin and norepinephrine have only been confirmed in hepatocyte cultures and currently not directly in the context of PHx. In addition, hepatocytes also show decreased expression of TGF-β receptors during the first few days of PHx, therefore conferring a further resistance to its effects (Chart et al., 1995). As the regenerative model progresses, new TGF-β is synthesised, with synthesis starting at around 2-3h and remaining elevated until 72h in rats (Jakowlew et al., 1991). TGF-β1 is known to stimulate the producing of ECM proteins by mesenchymal cells (Roberts et al., 1992). Evidence has shown that mouse Englebreth-Holm-Swarm sarcoma matrix extracts inhibit DNA synthesis in hepatocytes (Rana et al., 1994). Furthermore, removal of integrin-linked kinases have also lead to

increased cell proliferation and improper termination of PHx (Apte et al., 2009; Gkretsi et al., 2008). This suggest that ECM also provides a mito-inhibitory effect. This could be through direct binding to the ECM, but also the newly synthesised ECM can rebind HGF, preventing its activation and mitogenic effects, and also rebind TGF-β, returning to the previous paradigm in which TGF-β exerts a 'tonic' mito-inhibitory effect (Michalopoulos, 2007).

While this makes for a very clean model with TGF-β1 at the centre, almost normal regeneration with effective termination occurs in TGF-β1 receptor KO mice, unless activin, another known mito-inhibitor of hepatocytes, is also removed through follistatin administration (Oe et al., 2004). This suggests that termination requires at least one of these molecules and that there are again compensatory effects by multiple pathways to ensure proper regeneration occurs. Moreover, the importance of integrins in the ECM cannot be overlooked as the key terminator of PHx.

1.4.6 Partial Hepatectomy Mechanisms Summary

The mechanisms that control PHx are extensive and can overlap in many ways. Due to the large number of signalling pathways that are triggered during this model, identifying the individual function of any one molecule becomes difficult and can lead to contradicting theories. Many studies have used techniques that look at global changes and try to assign function based from these global changes. However, this is likely a source of confusion in the field and higher resolution techniques are required to resolve many of these complex mechanisms and pathways.

1.5 Single-Cell RNA Sequencing

1.5.1 RNA Sequencing

Gene expression studies have become one of the major ways in which researchers compare the differences between samples such as between conditions or time points. They allow a researcher to gain insight into the transcriptomic prolife of the cell by measuring the relative amount of each mRNA molecule from a sample. This is then commonly used as a pseudo-representation of the protein level within the cell. These types of studies only became possible with the advent of next generation sequencing (NGS). NGS was a major advancement over sanger sequencing, and allowed for sequencing of millions of fragments of DNA at once, giving NGS sequencers a much higher throughput (Sanger et al., 1977; Shendure and Ji, 2008). However, there is not as yet a commercially available sequencer that can sequence RNA molecules directly. Therefore, mRNA molecules are sequenced by first converting the mRNA into cDNA using reverse transcription polymerase chain reaction (RT-PCR; Haque *et al.*, 2017). This is then followed by standard PCR to amplify the small amount of cDNA created from the RT-PCR reaction, to gain enough material for sequencing. This is even more vital for single cell RNA sequencing (scRNA-seq) where the starting material is very low.

The main difference between bulk RNA sequencing and scRNA-seq is within the name. Bulk RNA sequencing takes the RNA from a large sample of many different cells, usually from a particular organ and/or particular cell type. While there are many methods allowing separation of cells into specific populations to allow bulk RNA sequencing, we are limited by the current biological knowledge about these cell types. scRNA-seq however considers each cell as an individual sample and allows the transcriptome of each individual cell to be measured. Therefore, it is able to distinguish unknown heterogeneous populations without any prior knowledge of these populations that would have otherwise been masked by the bulk samples.

scRNA-seq is a multistep process, which follows many of the same processes as bulk RNA seq (G. Chen et al., 2019; Haque et al., 2017). Cells first need to be isolated and captured, lysed and their mRNA reverse transcribed. The cDNA is then amplified before being sequenced.

1.5.2 Cell Isolation and Capture

There are multiple ways to isolate the cells of interest for scRNA-seq and these can be spilt into low-throughput and high-throughput methods (Hu *et al.*, 2016; Figure 1.5 A). Low throughput methods include laser capture microdissection or micropipetting. Both these methods allow for very precise selection of cells or a region of cells however are very labour and time intensive. High throughput methods on the other hand are very fast and include the use of techniques such as fluorescence activated cell sorting (FACS) or microfluidic based systems.

Theses isolation techniques usually go hand in hand with the style of scRNA-seq that is being carried out (G. Chen et al., 2019). The different types of scRNA-seq can be split into many categories but for a capture orientated view, they can be broadly split as plate-based or droplet-based. Plate-based techniques involve sorting/pipetting the isolated cells into either 94 or 384 well plate. This can be achieved by use of the low throughput methods, but prominently is achieved by FACS. The main reason for this, is a FACS is able to process hundreds of cells in minutes as opposed to the hours of individually picking cells. The technique however relies on the ability to be able to identify the cells of interest using intrinsic properties of the technique or fluorescently labelled antibodies. Once the cells are sorted into a plate, there are many different types of library preparation protocols including; SMART-seq, MARS-seq or NEBNext® Single cell sequencing, to name but a few (Haque et al., 2017).

Droplet-based sequencing, such as Drop-seq or 10x Genomics Chromium, feed cells through a microfluidic device and use oil droplet emulsions to capture the cells along with the necessary reagents (Salomon et

al., 2019). This process allows for thousands of cells to be captured in one run. This is by far the highest throughput technique available. Each droplet-based platform has its own distinct library preparation protocol, however the fundamentals remain the same. While a user can feed in a freshly isolated population of cells without any prior enrichment, most researchers again utilise FACS.

1.5.3 Library Preparation

The overall steps in all scRNA-seq library preparation, as with bulk RNA sequencing, are the same (Haque et al., 2017). These involve, lysis of the cell, followed by first and second strand synthesis. Final steps include cDNA amplifications and the attachment of sequencing primers. While each different library preparation achieves these steps in different ways, the main two differences that are seen between library preparation are; the use of unique molecular identifiers (UMIs) and transcript coverage (Figure 1.5 B).

Protocols such as SMART-seq 2 and NEBNext® single cell are able to produce full length transcript coverage which can provide extra information such as alternative-splicing events and allele-specific expression (G. Chen et al., 2019). Other plate-based techniques produce either 5' and 3' transcripts, such as CEL-seq. This still allows for effective counting of transcripts, but you lose the ability to find isoform variants. However, a decrease in sensitivity, reduces cost (Ding et al., 2019). These types of protocol are therefore used when high numbers of cells are desired. This type of coverage is the only type that is offered by current droplet-based techniques.

There is a correlation with the use of UMI and the transcript coverage generated for each library preparation technique. Almost all full-length transcript protocols do not allow for the use of UMIs, except MATQ-seq (G. Chen et al., 2019). The use of UMIs is predominantly seen with droplet-based sequencing approaches such as 10x Chromium single cell RNA sequencing, or any of the 3' and 5' coverage plate-based techniques. UMIs are small barcodes of DNA that are attached to each piece of mRNA during the initial

RT-PCR (Hwang et al., 2018). The barcodes are different for each mRNA molecule and therefore allow the detection of PCR duplicates when comparing the sequences during the alignment and analysis. Removal of PCR duplicates is a great advantage as a more realistic estimation of the cell's transcriptome is recovered with the removal of duplicate bias.

One final choice that can be made during the library preparation step is the inclusion of spike-ins. Spike-ins, such as the External RNA Control Consortium (ERCC), are RNA transcripts with known sequences that are added to each sample (cell) in a known quantity (Lun et al., 2017). In theory, the use of spike-ins would allow the estimation of technical noise. Like UMIs, these can only be used with certain protocols, in general, plate-based approaches implement spike-ins. However, the use of spike-ins can be problematic within themselves. The effective use of spike-ins relies heavily on the accurate concentration being added to each sample. They are also found to be quickly degraded and are not as efficiently captured as endogenous RNA. For these reasons, the use of spike-ins has fallen out of favour with many research groups.

1.5.4 Sequencing

Broadly speaking, the lower number of cells captured in plate-based approaches is usually associated with a deeper sequencing readout per cell. On the other hand, when higher cell numbers are captured, for example in droplet-based protocols, a lower sequencing depth is obtained (Figure 1.5 B). One main reason for this is cost (Ziegenhain et al., 2017). When considering how many reads that can be afforded for a particular experiment, spreading these reads across fewer cells will give a higher read depth per cell, while spreading them across a larger number of cells will reduce the depth per cell. That being said multiple studies have shown that sequencing any library at one million reads per cell reaches a reasonable level of saturation (Wu et al., 2014; Ziegenhain et al., 2017). In practise it is usually more efficient to sequence the cells from a droplet-based approach at a factor of 10 lower than those of plate-

based full length approaches (Haque et al., 2017). It should be noted that sequencing cells from one protocol to the same degree as another will not result in discovering the same number of genes, nor will doubling the sequencing depth necessarily give a 2x increase in the number of genes detected (Ziegenhain et al., 2017). The reason for this is that each protocol will have its own efficiency of capturing and reverse transcribing the RNA for each cell. Furthermore, regardless of platform many genes that are expressed at a moderate or low level may not be captured and detected. Therefore, after a certain point, increasing sequencing depth will give diminishing returns as more PCR duplicates are sequenced.

The sequencing itself for most studies is likely to be conducted on an Illumina platform (Buermans and den Dunnen, 2014). Illumina are by far the largest company in the NGS sector and it has become the norm to use their machines in almost all sequencing experiments. Their process of sequencing DNA is known as sequence by synthesis. First, cDNA from a sample is immobilised on a flow cell through binding of the adapter sequences to oligonucleotides attached to the flow cell. The cDNA fragments are then amplified through bridge amplifications to form colonies of the same cDNA fragment. A cluster of cDNA molecules allows for a stronger signal to be produced during sequencing. This process was greatly improved through the introduction of nano-wells into the flow cell. These pre-formatted grids of nano-wells allow for specific amplification of a colony originating from a single cDNA molecule. This is achieved through the difference in speed between initial capture and amplification of the cDNA. Amplification of cDNA is faster than the initial capture, therefore, once a cDNA has bound into a nano-well, the cDNA fragment is amplified faster than a second cDNA fragment can bind into the same well. Following on from cluster generation, the reverse strands of cDNA are cleaved and removed to allow the forward strand to be sequenced first. Sequencing is achieved through the addition of fluorescently tagged nucleotides by a DNA polymerase. As each nucleotide binds to its complementary base, the fluorescent signal is measured, with each nucleotide corresponding to a particular fluorescent signal. Once the forward strand has

been sequenced, the same occurs for the reverse strand (which is generated through a single round of amplification) to generate paired end reads. This allows for millions of bases to be sequenced in one run which are collated and aligned to a reference genome to generate an expression or counts matrix of genes and cells.

1.5.5 Data analysis

There are many well established pipelines to conduct bulk RNA sequence analysis (Love et al., 2015; Robinson et al., 2010). scRNA-seq is however a newer technique and while there are many different papers describing the “best” way to analyse the data, currently, there is no “gold standard” pipeline (Luecken and Theis, 2019). One of the main differences that was initially considered between the two techniques, was the large number of zeros that is found within scRNA-seq data (Qiu, 2020). Commonly referred to “dropouts”, these are transcripts which are seen to be expressed at a moderate or low level in one cell but are completely absent from another cell of the same type. These arise due to the low starting material, mRNA capture efficiency, the non-linear nature of the amplification steps, and the stochastic nature of sampling. These technical zeros cannot be distinguished from biological zeros which represent the potentially true “on- off” nature of a gene (Andrews and Hemberg, 2019; Silverman et al., 2020). Nevertheless, it is believed that “dropouts” cause a zero inflation within the data, and therefore many analytical techniques have been developed to deal with this. One suggestion is to use zero-inflated statistical models such as the zero-inflated negative binomial model (Kharchenko et al., 2014; Qiu, 2020). These have quickly become popular, however recent evidence has shown that the use of zero-inflated models may not be appropriate for UMI based data (Svensson, 2020). While many methods attempt to model the count data directly, it is still common to find many workflows using logged normalised counts (Luecken and Theis, 2019). Studies evaluating various single cell analysis pipelines have shown that good normalisation removed the need for specialised tools such as those

used for differential expression. The analysis of scRNA-seq data is explained in greater detail in chapters 4 and 5.

1.5.6 Novel Findings

The number of publications and studies that use scRNA-seq approaches has increased dramatically over the past couple of years. scRNA-seq allowed for the identification and interrogation of previously unknown population of cells.

One area of liver biology that has expanded rapidly with the advent of scRNA-seq is that of zonation. Halpern *et al.* (2017) used a combination of single-molecule fluorescence in situ hybridization (smFISH) and scRNA-seq to uncover the extent to which the hepatocyte transcriptome is zonated. This was achieved by reconstructing the spatial zonation in the scRNA-seq data based on their expression of landmark genes. The liver lobule was divided into nine regions and they found that more than 50% of expressed genes are zonated as well as the presence of non-monotonic gene zonated profiles which peak in the mid-zonal regions. Furthermore, they found the genes associated with bile acid biosynthesis were zonated in the order of their function along the direction of bile flow. Enzymes early in the pathway were expressed pericentrally, and those later in the pathway expressed more periportally. This work challenged the binary/tertiary classification of liver zones. Therefore, if zones are to be described as distinct regions, it may be more appropriate to distinguish between a more fixed spacial classification, i.e. periportal or pericentral, with fixed division based on distance from the central vein and portal triad, and a more dynamic metabolic definition, i.e. gene x is expressed in metabolic zone x. The latter having a dynamic boundary that is linked to well established zonated genes and processes.

scRNA-seq has also led to the discovery of spatial and functional zonation of other cell types in the liver (Dobie et al., 2019; Halpern et al., 2018). A recent study by the Henderson lab (Dobie et al., 2019) revealed zonation of HSC, identifying portal vein associated HSC (PaHSC) and central vein

associated HCS (CaHSC). Importantly, CaHSC were shown to be responsible for the majority of pathogenic fibrillar collagen production in a model of centrilobular fibrosis. This study clearly shows the importance of identifying sub-populations, as such resolution allows for highly targeted therapeutic approaches to be developed.

Another exciting area that has developed with the advent of scRNA-seq, is the generation of a cell atlas. Large efforts are underway to combine studies to produce a “Human Cell Atlas” to identify all cell types and sub-populations of cells that are found within the human body (Regev et al., 2017). This will generate an exceedingly high resolution map of the cells of the human body, with a view to helping understand the biology of the various organ systems. This will also have a significant impact on the field of biomedicine and aid the identification of targeted treatments for a broad range of diseases. Two such examples exist already for human liver in which both the parenchymal and NPCs were classified by scRNA-seq (Aizarani et al., 2019; MacParland et al., 2018). Within both studies, hepatocyte and endothelial zonation was identified, as well as identifying sub-populations of cell types that were once believed to be homogenous. MacParland *et al.* (2018) were able to identify two sub-populations of KCs that showed functional differences, while Aizarani *et al.* (2019) discovered a novel bipotent epithelial progenitor.

As well as uncovering homeostatic cell populations, scRNA-seq has been vital to expand our knowledge of disease specific populations. Ramachandran *et al.* (2019) identified a scar associated $TREM2^{+}CD9^{+}$ subpopulation of macrophages in human liver cirrhosis. It was shown that this sub-population expands in cirrhosis through the differentiation of circulating monocytes and established a pro-fibrotic phenotype. Such work will aid the identification of potential therapeutic targets for liver cirrhosis.

As scRNA-seq as a technology advances further, researchers will be able to unpick more subtle differences in biology. In particular with the advent of spatial transcriptomics, these novel cell types will be able to be mapped spatially and more precise hypotheses of cellular interactions will be able to be derived from the data.

1.5.7 Choice of Single Cell RNA Sequencing Approach

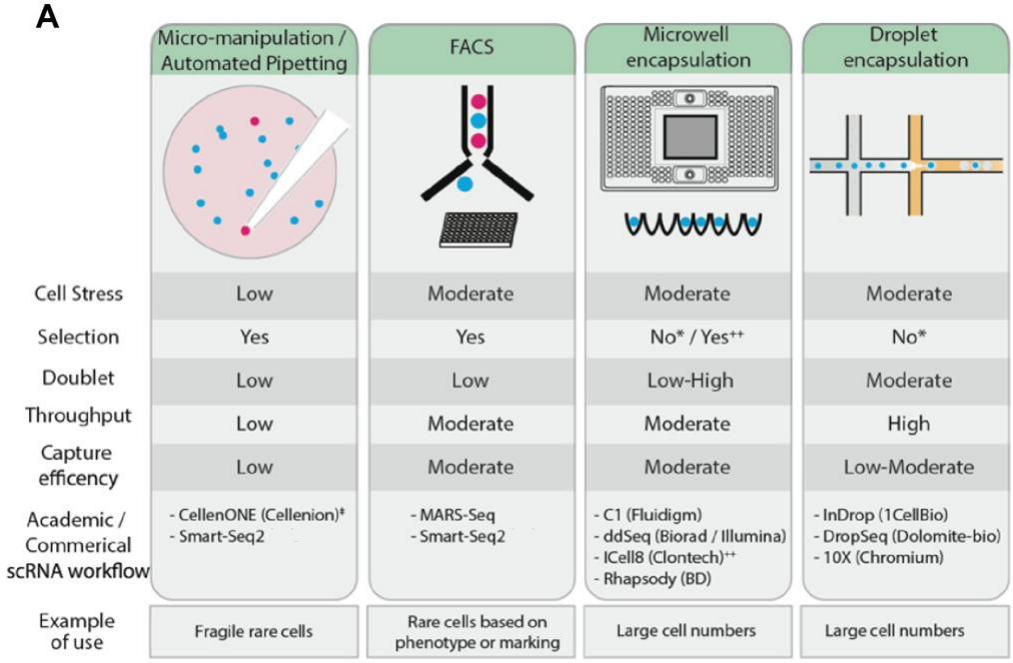
As discussed, there are many different methodologies for scRNA-seq. The choice of method is dependent on the biological question and/or what is expected to be found (Haque et al., 2017). For instance, when trying to identify a rare subpopulation, a large number of cells is most likely required, however, this usually will come at a cost of number of gene detected per cell. Generally speaking, most plate based full length transcript methods (SMART-seq, NEBNext®) will give a greater depth into the transcriptome of each cell and therefore yield a greater number of genes per cell as well as also allowing gene isoform discovery. However, these methods are only able to sequence a few hundred cells in any one run, and to sequence thousands of cells would be very expensive. Molecular tag-based methods (MARS-seq) do not give full transcript coverage or as many genes per cell, but they allow for UMI incorporation which helps to reduce many downstream technical artefacts. UMIs are also used in conjunction with droplet-based isolation methods (10x Genomics) that allow for many thousands of cells to be sequenced at once.

It is clear that the regenerative process during PHx is complex and researchers are still uncovering new key signals pathways involved within the process. To date, most studies have looked broadly at hepatocyte signalling, however it has been shown hepatocytes contribute to the process in a heterogenous manner, with only a proportion of hepatocytes that enter the cell complete mitosis (Miyaoka and Miyajima, 2013). Therefore, scRNA-seq would be the ideal technique to try and unpick the heterogenous response of hepatocytes during injury and attempt to identify the signals that may be specific regulators of certain populations of hepatocytes. Such novel signalling may provide avenues for new and novel therapeutic targets to tackle liver diseases.

For this project I initially felt that a droplet based approached would be the most optimal for my research questions, as the larger number of cells would allow for the greatest potential to identify subpopulations and therefore

any key signals from these populations. However, due to technical reasons I had to turn to a plate-based approach. Although this might make identification of sub-populations more challenging, the greater number of genes detected allowed for better identification of key signals from the data. Due to the lower number of cells captured by a plate-based approach, I selectively enriched for replicating hepatocytes using a transgenic mouse (Fucci). I felt the analysis of replicating hepatocytes would be key in the identification of the signalling molecules that regulate the regenerative niche, helping to uncover the molecular mechanisms that regulate hepatocyte proliferation.

A



B

Protocol example	*C1 (SMARTer)*	*Smart-seq2*	*MATQ-seq*	*MARS-seq*	*CEL-seq*	*Drop-seq*	*InDrop*	*Chromium*
Transcript data	Full length	Full length	Full length	3′-end counting	3′-end counting	3′-end counting	3′-end counting	3′-end counting
Platform	Microfluidics	Plate-based	Plate-based	Plate-based	Plate-based	Droplet	Droplet	Droplet
Throughput (number of cells)	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^2–10^3	10^3–10^4	10^3–10^4	10^3–10^4
Typical read depth (per cell)	10^6	10^6	10^6	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5	10^4–10^5

‡Automated pipetting system
*Preselection or enrichment can be performed prior
++Only reagents added to wells containing singlets, determined by system
FWP: Fluidigm white paper
PB: Product brochure / manual

Figure 1.5 – Characterisation of Common Cell Isolation and Single Cell RNA Sequence Methods

A) Shown are difference methodologies of cell isolation for single cell sequencing, along with associated properties, such as ability to select cells, potential of doublets, throughput and capture efficiency. Commonly associated single cell methodologies are shown. Image was adapted from Nguyen *et al.* (2018)

B) Various library preparations with the transcript coverage, associated platform, usually cell throughput and typical read depth shown for each one. Image was adapted from Haque *et al.* (2017).

1.6 Hypothesis and Aims

Hypothesis - Subpopulations of replicating hepatocytes produce key signals that regulate the hepatic regenerative niche.

Aim – Use scRNA-seq to investigate the transcriptomic profile of both replicating and non-replicating hepatocytes following PHx in mice.

The experimental objectives to test this hypothesis are:

1. Optimise the isolation of replicating and non-replicating hepatocytes for single cell RNA sequencing.
2. Perform single cell RNA sequencing on both replicating and non-replicating hepatocytes to study the transcriptional changes of these cells following PHx.
3. Use cutting-edge bioinformatics techniques to analyse transcriptomic data to identify putative pro-regenerative signals produced by replicating hepatocytes.

2 Material and Methods

2.1 Mice

C57BL/6J wild type (WT) mice were purchased from Charles River or Bioresearch & Veterinary Services (University of Edinburgh). FUCGLC$^{+/-}$ mice (hence referred to as Fucci^{+}) were obtained from Dr. Kylie Matchett. All mice were maintained on a C57BL/6J background and bred at the University of Edinburgh under specific pathogen-free conditions. All experiments were conducted using 8 to 20-week-old male mice following approval from Bioresearch & Veterinary Services under UK Home Office Legislation.

2.2 Partial Hepatectomy

All two-thirds partial hepatectomy surgeries (PHx) were performed by Dr. Kylie Matchett with assistance from myself. Mice were initially anesthetised by inhalation of 3% isoflurane (Henry Schein) mixed with 3L/min oxygen flow and kept on a heated mat while they were prepped for surgery. Warm sterile 0.9% saline (25mL/kg, Braun, Sodium Chloride 0.9% w/v) and buprenorphine (0.1mg/kg, Ceva, Vetergesic) diluted to 0.03 mg/ml in sterile water was injected subcutaneously. Eye lubricate (Viscotears Liquid Gel, Alcon Laboratories UK Ltd.) was applied and mice were shaved, and their skin sterilised with chlorhexidine. Mice were then transferred to a sterile heatmap and anaesthesia inhalation was reduced to 2% isoflurane with 2L/min oxygen flow.

PHx was performed as described in Mitchell and Willenbring (2008). A small abdominal skin and muscle incision was made along the midline close to the sternum. The medial and left lateral lobes were exteriorised, individually ligated with 1.5M braided silk (SMI) and then excised. The small abdominal muscular incision was sutured closed via a continuous stick with 1.5M polyglactin 910 (Ethicon, Vicryl) thread. The outer skin incision was close with

surgical clips (Biochrom, 9mm Autoclips). Animals were humanely killed by either CO_2 induction or perfusion termination (Methods 2.4) at stated timepoints post PHx surgery.

2.3 5-Ethynyl-2'-deoxyuridine Administration

2.5mg/mL of 5-Ethynyl-2'-deoxyuridine (EdU, Sigma Aldrich) was dissolved in sterile Phosphate Buffered Saline (PBS; Thermo Fisher Scientific) and injected intra-peritoneally at 50mg/kg into mice, 3 hours prior to humane termination.

2.4 Liver Digestion via Portal Vein Perfusion

Liver digestion protocol was kindly supplied by Dr Sofia Ferreira-Gonzalez. Mice were anaesthetised by inhalation of 3% isoflurane mixed with 3L/min oxygen flow and kept on a heated mat during the procedure. A long incision was made along the midline of the mouse's abdomen. The portal vein (PV) was cannulated with a 24G cannula (VWR), tied in place with a suture, and the intra vena cava (IVC) was cut. A peristaltic pump (Minipuls 2, Gilson) was used to perfuse 15ml of liver perfusion media (LPM, Thermo Fisher Scientific) at a rate of 5ml/min through the PV into the liver and out of the IVC. Liver perfusion media was pre-warmed at 45^{o}C and kept at this temperature in a water bath throughout the procedure. Following this, the IVC was secured shut with a second suture and the heart was cut to allow a further 15ml of LPM to pass through the liver and out through the superior vena cava (SVC). Finally, the SVC was secured with a third suture and the liver was perfused with 30ml of pre-warmed (45^{o}C in water bath) liver digest media (LDM, Thermo Fisher Scientific).

The digested liver was dissected out of the mouse with care not to cut the oesophagus or the liver capsule and the gall bladder was removed. The liver was placed in 25ml Hepatocyte media (HM, Williams E medium, no

glutamine [Thermo Fisher Scientific, 22551022], 1% v/v Penicillin/Streptomycin [10,000 U/ml, Thermo Fisher Scientific], 1% v/v L-Glutamine [200mM, Thermo Fisher Scientific], 1% v/v MEM non-essential amino acids [100x, Thermo Fisher Scientific]) on ice if the shortened hepatocyte isolation protocol was followed (Methods 2.5) or at room temperature (RT) if the longer protocol was being followed (Methods 2.6).

2.5 Hepatocyte Isolation – Long (Original)

The digested liver from Methods 2.4 was gently pulled apart with forceps in 25ml of HM. The resulting cell suspension was filtered through a 70µm filter with addition 25ml of HM followed by a 135xg centrifugation (acceleration = 9 / deceleration = 9, Thermo Heraeus Multifuge, 1S-R, D-37520) for 1 minute (min[s]). The supernatant was discarded and 50ml of a 50:50 ratio of HM and Percoll solution (9:1 v/v Percoll [Sigma Aldrich, P1644-1L], Hanks Balanced Salt Solution [HBSS,10x, Thermo Fisher Scientific]) was added to the cell pellet. The cell suspension was mixed gently by inversion. The cell suspension was centrifuged (acceleration = 1 / deceleration = 1) at 50xg for 10 mins. The supernatant was carefully removed via aspiration, discarded and the pellet was resuspended in 25ml of HM. The cell suspension was centrifuged at 135xg for 1 min, supernatant discarded, and washed again with 25ml of HM. Once centrifuged a second time the pellet was resuspended in 10-25ml of HM. Cells were subsequently counted using the TC20™ Automated Cell Counter (Bio-Rad). Individual samples of $1x10^6$ cells were created and either washed twice with wash buffer (WB, Phosphate Buffered Saline – Calcium – Magnesium [PBS, Sigma-Aldrich], 0.38mg/ml Ethylenediaminetetraacetic acid [EDTA, Sigma]) or were stained with 1.5µl/ml Hoechst 33342 (10mg/ml stock, Life Technologies) plus 1µl/ml Reserpine (5mM stock, Sigma-Aldrich) for 35 mins at 37˚C prior to the two washes. Cells were finally resuspended in 500 ul of WB.

2.6 Hepatocyte Isolation – Short (Optimised)

The shortened hepatocyte isolation protocol followed the same steps as Methods 2.5; however, the percoll purification and 2 subsequent wash steps were removed. Following the initial centrifugation prior to the percoll purification, the cells were resuspended in 10-25ml of HM before being counted.

2.7 Liver Fixation and Paraffin-Embedding

The liver was perfused via the portal vein with 10ml of PBS before being harvested into 4% formaldehyde overnight. The tissue was transferred into 70% Ethanol prior to paraffin-embedding and sectioning (5µm) by the Queen Medical Research Institute (QMRI) Histology service.

2.8 Immunofluorescence and Immunohistochemistry

2.8.1 Paraffin-Embedded tissue preparation

Tissue sections were de-waxed and rehydrated through two incubations in xylene for 5mins, followed by 2 min incubations in each of 100%, 75% and 65% ethanol. Sections were subsequently washed for 5 mins in dH_20 before 20 min post fixation in 4% formaldehyde. Sections were washed for 5 mins in 0.5 % PBS-Tx (v/v Triton-x [Sigma-Aldrich] in PBS) prior to stated heat mediated antigen retrieval by microwaving in 10mM Sodium Citrate buffer, pH6, for 15 mins. Endogenous peroxidase activity was inhibited through 10 min incubation in 3% v/v hydrogen peroxidase. Finally tissue sections were permeabilised for 20 mins in 0.5% PBS-Tx at RT. All subsequent wash steps were performed for 5 mins using 0.5% PBS-Tx at RT and blocking was performed with serum block (Biolegend) at RT, unless otherwise stated. All primary, secondary and isotype controls can be found in Table 2.1 and were diluted in antibody diluent (Abcam). Appropriate isotype controls were used to

asses non-specific binding and act as negative controls to confirm positive staining, unless otherwise stated.

2.8.2 Hnf4α/Edu/E-Cadherin

Paraffin-embedded tissue was washed two times in PBS prior to the addition of EdU detection cocktail for 30 mins at RT, made per the manufacturer's guidelines (Click-iT Plus EdU Imaging Kit Alexa Fluor 647 [Thermo Fisher Scientific]), to visualise incorporated EdU. Tissue from mice that had not been subjected to EdU incorporation was used as a negative control, alongside tissue with EdU incorporation incubated with Antibody diluent. Sections were washed thrice and blocked for 30 mins followed by staining with E-cadherin for 1h at RT. Tissue sections were washed thrice prior to a 30 min incubation with anti-mouse polymer ImPRESS (Vector Laboratories) at RT processed by a further three washes. Tyramide Signal Amplification (TSA) plus Fluorescein (Perkin Elmer) was applied for 10 mins at 1:1000 and RT. Sections were washed three times prior to a second round of antigen retrieval and permeabilisation as described in Methods 2.8.1. Tissue sections were blocked for 30 mins followed by overnight incubation of HNF4α at 4˚C. The same wash, polymer incubation, and TSA steps as for E-Cadherin were repeated with the substitution of Fluorescein for Cyanine 3 (Perkin Elmer). Following TSA incubation sections were washed a final three times in PBS-Tx, counterstained with DAPI (Sigma Aldrich) diluted at 1:1000 in PBS for 10 mins at RT, washed once in PBS before mounting in ProLong Gold (Thermo Fisher Scientific).

Table 2.1 - Antibodies used in Immunohistochemistry

Description	Application	Manufacturer	Catalogue Number	Dilution
Mouse Anti-HNF4α	1°	Bio-Techne Ltd	PP-H1415-00	1:200
Mouse Anti-E-Cadherin	1°	BD Biosciences	610181	1:1000
Mouse IgG2a	Isotype	Sigma-Aldrich	M5409	1° dependent

2.9 Image Capture

Fluorescent images of whole tissue sections were captured on the AxioScan.Z1 (Carl Zeiss) at 20x magnifications. Cells sorted from Fucci^{+} mice were imaged using the Leica Confocal (SP8, Leica) at 20x on a POC mini (Pecon). Evaluation of single hepatocytes in 384 well plates was achieved using the Axio Vert.A1 Bio (Carl Zeiss) inverted microscope. Fluorescence intensity was set using negative controls and kept constant throughout a single experiment.

2.10 Hepatocyte Replication Quantification

Slide scanned images were loaded into QuPath (Bankhead *et al.*, 2017) for quantification analysis and visualisation. Random fields were chosen from one repeat per timepoint to generate a training image. The QuPath Cell Detection algorithm was optimised using the DAPI signal on the training image. HNF4α^{+} and HNF4α^{-} cells were manually selected within the training image to optimise the QuPath Object Classifier. This allowed automated classification of Hepatocytes (HNF4α^{+}) and other cell types (HNF4α^{-}). A threshold was built into the classifier to further assign cells as EdU^{+} or EdU^{-} based on the intensity of the EdU signal. The threshold value was chosen based on the training

image. E-Cadherin^{+} and E-Cadherin^{-} regions were manually curated within the training image to optimise the QuPath Pixel classifier. The Pixel classifier allowed for automated classification of periportal (E-Cadherin^{+}) and pericentral (E-Cadherin^{-}) regions. Whole slide sections were analysed using the three optimise classifiers to generate the number of replicating hepatocytes (HNF4α^{+}/EdU^{+}) stratified by their zonal position (periportal or pericentral).

2.11 Flow Cytometry and Fluorescence-Activated Cell Sorting

For flow cytometry and fluorescence-activated cell sorting (FACS), cells were isolated as described in section 2.5 or 2.6. WT mice were used as a negative control for the Fucci^{+} transgenic mouse. Cell viability was assessed using Draq7 (1:500; Biolegend) added 10 minutes prior to cell sorting. Cell sorting was performed on FACSAria II (BD Biosciences) and data was analysed using FCSExpress 7 Research software.

2.11.1 Gating Strategy

Hepatocytes were detected by gating on the larger cells using forward (FSC-A) and side (SSC-A) scatter (Figure 2.1). Doublets were removed via FSC-A vs FSC-H followed by SSC-A vs SSC-H. Live cells were isolated based on exclusion of Draq7. For identification of proliferation and non-proliferating cell the gating strategy is described in section 3.2.4. Briefly, replicating hepatocytes were isolated as mVenus^{+}/mCherry^{-}/Auto-fluorescent^{-} and non-replicating hepatocytes were defined as mVenus^{-}/mCherry^{+}. All quantification was calculated as a percentage of live single hepatocytes.

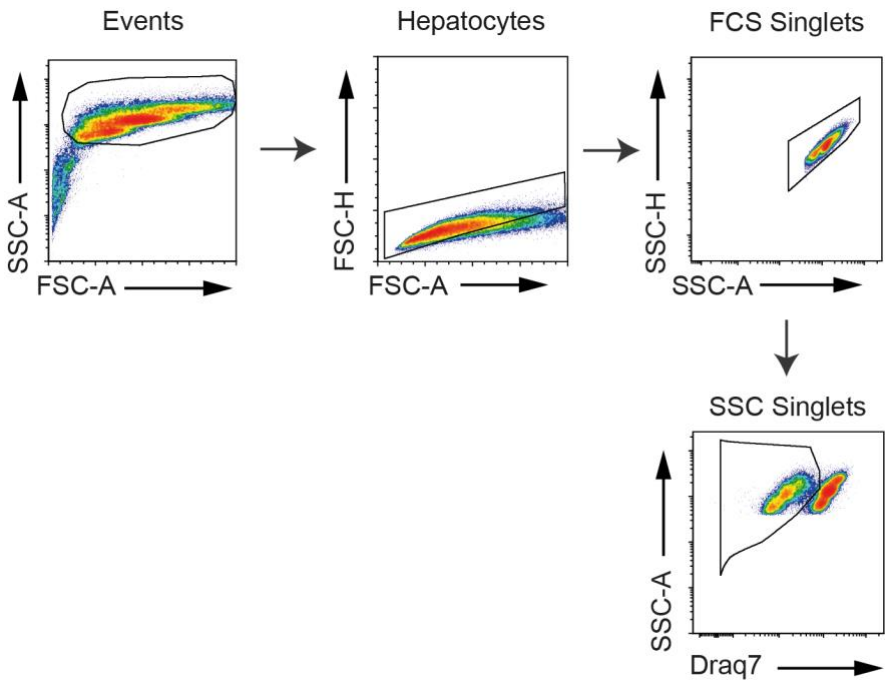


Figure 2.1 – Live Single Hepatocyte Gating Strategy

Gating strategy for live single hepatocytes. Hepatocytes were gated using forward scatter (FCS-A) and side scater (SSC-A). Single hepatocytes were gated using FCS-A vs FCS-H followed by SSC-A vs SSC-H. Live hepatocytes were isolated by the exclusion of Draq7. Title above each plot refers to parent gate.

2.12 Droplet-Based Single Cell RNA Sequencing

Single live hepatocytes were first isolated using the gating strategy described (Figure 2.1). The resulting hepatocyte population was processed through the 10x Genomics Chromium™ Single Cell Platform using the Chromium™ Single Cell 3' Library and Gel Bead Kit v2 (10x Genomics) and the Chromium™ Single Cell A Chip Kit (10x Genomics) as per the manufacture's protocol. The hepatocytes were sorted into sterile PBS, washed twice, filtered using a 40µm Flowmi® filter (Sigma) and counted using the TC20 (Bio-Rad). 7000 cells and 10x reagents were added to one lane of the 10x microfluidics chip. 10x gel beads were added into the second lane and oil into the third. The chip was inserted into the 10x Chromium™ controller where the cells were partitioned into a gel bead emulsion. Within the emulsion, the cells where lysed, the RNA barcoded and reverse transcribed. The resulting cDNA was amplified, fragmented and followed by attachment of the 5' adaptors and sample index. Libraries were sequenced on an Illumina HiSeq 4000 at Edinburgh Genomics.

2.13 Plate-Based Single Cell RNA Sequencing

Each well of a 384 well plate (Eppendorf TwinTec 384 Well PCR Plate) was first filled with 2µl of the lysis buffer from the NEBNext® Single Cell/Low Input RNA Library Prep Kit for Illumina® using a E1-ClipTip™ Bluetooth™ Electronic Multichannel Pipette. Plates were sealed with Adhesive PCR plate foils (Thermo Fisher Scientific) and stored at -80^{o}C prior to use. Plates were defrosted one ice and the seal removed before being placed in the FACSAria II. One live single cell (hepatocyte) were index sorted into each well. The 384 well plates were re-sealed and spun at 1200g for 5 seconds before being frozen on dry ice and stored at -80^{o}C prior to shipping to the Welcome Sanger Institute for cDNA creation and library preparation using NEBNext® Ultra™ FS DNA Library Prep Kit for Illumina. The quality of cDNA was assessed using 2100 Bioanalyser (Agilent) for 11 randomly selected wells. cDNA

concentration was measured for all wells using Omega FLUOstar plate reader (BMG Labtech). Sequencing was performed using a NovaSeq 6000 at a depth of 600,000 reads per cell.

2.14 Bioinformatics Analysis of Single Cell RNA sequencing Data

2.14.1 Droplet-Based Sequencing Data

FASTQ files were downloaded from Edinburgh Genomics and aligned to the mm10 (Ensembl84) mouse reference genome using the Cell Ranger v2.1.0 Single-Cell Software Suite (10x Genomics). The web summary output from this pipeline was inspected to ascertain the quality of the data. As discussed in section 3.2.3 the data was below the quality expected of publishable material, and analysis was stopped at this point.

2.14.2 Plate-Based Sequencing Data

2.14.2.1 Quality Control

FASTQ files were downloaded from the Welcome Sanger Institute and trimmed using Flexbar v3.5.0. Trimmed FASTQ files were aligned to the (GRCm38.p4; Ensembl release 81) mouse gene and an expression matrix generated using the nf-core rnaseq pipeline (version 1.4.2; Ewels *et al.* [2019]). All subsequent analysis was performed using the R statistical programming language (version 3.6.3; R Core Team [2020]). Cells were removed if they expressed fewer than 5000 genes or had a mitochondrial read count percentage above 10%. Genes were excluded if they did not show a count level of five or more in at least three cells. Contaminating cells were removed based on their marker expression (Table 4.2). All cells expressing at least one count of any contaminating marker were removed. Associated fluorescence metadata from the Aria II index sort was linked to corresponding hepatocytes to allow for classification of each cell as $mVenus^+$ only, $mCherry^+$ only or

mCherry^{+}/mVenus^{+} (Dual^{+}; Figure 4.7). Cell variability prior to normalisation was assessed by first performing principle component analysis (PCA) on log gene expression of all genes and then fitting a linear regression for PC 1-10 against individual associated quality control (QC) metrics. All logs are natural log unless otherwise stated. The R-squared value from the linear regression was used to determine the influence each QC metric had on the data. Gene variability prior to normalisation was determined using a linear regression model of log gene expression of all genes against individual associated QC metrics. The *plotExplanatoryVariables* function from the *scater* R package (version 1.14.6) was used to perform this analysis (McCarthy et al., 2017). Again, the R-squared value functioned as a measure of how much variability was explained by each QC metric.

2.14.2.2 Normalisation, Variable Genes and Scaling Data

Various normalisation approaches were assessed using the *scone* R package framework (Cole et al., 2019). Briefly, *scone* was used to test: *scran* (version 1.14.6; Lun, McCarthy and Marioni [2016]), *SCnorm* (version 1.8.2, Bacher *et al.* [2017]), *TMM* (trimmed mean of M), *sum* (counts corrected by the summed library size of each cell), and *batchelor* (version 1.2.4; Haghverdi *et al.* [2018]). For *scran* normalisation, *quickCluster* (*scran* package) was first used to cluster cells into broad groupings, *computeSumFactors* (*scran* package) was used to generate the size factors for each cell and finally *normalize* (*scran* package) was used to apply the scale factor normalisation and log the resulting values. For *SCnorm* normalisation, the *SCnorm* function (*SCnorm* package) was used specifying batch as the different conditions. The *TMM_FN and SUM_FN* functions from the *scone* package were used to perform TMM and Sum normalisation respectively. Finally, for *batchelor* normalisation, *quickCluster* and *computeSumFactors* were used to cluster cells and generate scale factors for each batch independently. The *multiBatchNorm* function from the *batchelor* package was used to re-scale the

batch specific scale factors and apply them to the data to normalise the read count values, again these values were logged.

The QC metrics, number of read counts per cell, mitochondrial read percentage and batch, as well as the biological classification (mVenus^{+}, mCherry^{+}, Dual^{+}) were used within the scone analysis. The resulting metrics (Table 4.3) were visualised using a PCA biplot. Relative log expression per normalisation strategy was calculated and visualised using a modified version of *plotRLE* from the *scater* R package, that allowed for RLE to be calculated per normalisation strategy and plotted on the same graph.

Following normalisation, the top 2000 highly variable genes were selected using the *modelGeneVar* and *getTopHVGs* functions from the *scater* package. Gene variation was calculated for each batch separately, before averaging across all batches to achieve a single score for each gene.
Finally log normalised gene expression for all genes was scaled and centred by subtracting the mean expression of each gene and dividing by its standard deviation.

2.14.2.3 Batch Correction

Two methods of batch correction were compared: the *Seurat* integration method (*Seurat* version 3.1.3; Stuart *et al.* [2019]) and *Harmony* (version 1.0; Korsunsky *et al.* [2019])*. Seurat* integration first finds pairs of mutual nearest neighbours in canonical correlation subspace using the *FindIntegratedAnchors* function from *Seurat.* These pairs are referred to as anchors and are used to correct the expression matrices in a pairwise manner between batches using the *IntegrateData* function.

The *RunHarmony* function from the *Harmony* R package was used to correct PCs for batch effects. *Harmony* uses fuzzy clustering to assign cells to multiple clusters favouring mixed dataset representations. Centroids are calculated for each cluster and for each batch within a cluster. A correction factor is calculated based on the centroids. Cells from each batch are corrected by a cell specific factor; linear combination of the correction factor

and weight of fuzzy clustering. These steps are repeated until batches have converged.

2.14.2.4 Dimensionality Reduction, Clustering, Differential Gene Analysis and Pathway Analysis

Unsupervised (SNN graph-based) clustering, UMAP visualisations and differential gene expression analyses were performed using the *Seurat* R package. The SNN graph and UMAPs were constructed using between 1 and 15 PCs or Harmony dimensions as determined by dataset variability shown in the PCA and the Harmony analysis. All PCA analysis was performed using the 2000 variable genes calculated previously unless otherwise stated. The R-squared values used to assess post normalisation cell wise variability were calculated in the same manner as above using the same PCs or Harmony dimensions as for the UMAP and SNN graph creation. The *clustree* R package was used to generate a cluster tree to assess the resolution parameter. The *silhouette* function from the *cluster* R package was also used to assess the resolution parameter for cell clustering.

All heatmaps, UMAP visualisations and violin plots, were produced using *Seurat* functions in conjunction with the *ggplot2* (version 3.2.1; Wickham [2016]), *grid* (version 3.6.3; R Core Team [2020]) and *viridis* (version 0.5.1; Garnier [2018]) R packages. Differential gene expression analysis was conducted in Seurat using a logistic regression framework. Genes with a log-fold change of at least 0.25 and expression in at least 25% of cells in the cluster under comparison were retained and the rest were excluded. Batch and replicate were included as covariates within the logistic regression analysis. Predicted genes, mitochondrial genes and ribosomal genes were removed for the heatmap displaying the top ten differential expressed genes per cluster (Figure 5.3). This was to prevent uninformative genes being used to assess each cluster. The Cellphonedb database was used for cross referencing of differential genes with known ligand and receptors (Efremova et al., 2020). Pathway enrichment analysis was performed using *ReactomePA* (version

1.30.0; Yu and He [2016]) on significantly differentially expressed (<0.05 adjusted p value) genes with greater than 0.5 average log fold change in expression.

2.14.2.5 Statistical Analysis

All non scRNA-seq data was analysed using GraphPad Prism 8 for macOS (version 10.15.5). All graphs are shown as the mean +/- standard error of the mean (SEM). Statistical analyses are described in the corresponding figures. Significant values **P≤0.05, **P≤0.01, ***P≤0.001, ****P≤0.0001*.

3 Optimising Hepatocyte Isolation for Single-Cell RNA Sequencing

3.1 Introduction

3.1.1 Hepatocyte Replication Dynamics

The replication dynamics of hepatocytes in mice has been described in many studies. Peak hepatocyte proliferation is considered to be between 36 to 48 hours depending on the study (Mitchell and Willenbring, 2008; Zou et al., 2012). This variation is likely a result of differences in mouse strain and the method of proliferation quantification used. Other studies take a broader approach to quantifying hepatocyte proliferation: it is widely accepted that 48h post PHx shows a higher level of hepatocyte proliferation compared to 24h and 72h (Y. Chen et al., 2019; Miyaoka et al., 2012).

A small number of reports have also assessed the spacial dynamics of hepatocyte proliferation. Although evidence is limited, it is now generally documented that hepatocyte proliferation proceeds in a wave from the periportal to pericentral region following PHx (Rabes et al., 1976; Sun et al., 2020).

3.1.2 Fucci2a Transgenic Mouse

Florescent Ubiquitination-base Cell Cycle Indicator (Fucci) technology allows for the monitoring of the cell cycle in live cells (Zielke and Edgar, 2015). Immunohistochemistry is one of the most common techniques used to analyse the cell cycle of hepatocytes. Usually, EdU, or proliferation proteins such as Ki67, are visualised using antibody-based systems to identify which hepatocytes and how many are replicating at any one time. However, these techniques require samples to be fixed and take several days of processing to yield results. As a result, these methods are not suitable for live cell imaging,

or for the sorting of live cells for sequencing experiments. The Fucci construct mitigates these problems and allows for the real time visualisation of cellular replication, enabling hepatocytes at various stages of the cell cycle to be sorted, interrogated and compared.

The Fucci system utilities two components of the cell cycle machinery: Chromatin Licensing and DNA replication factor 1 (Cdt1) and Geminin. The concentration of these proteins oscillates as cells divide and therefore provide an accurate readout of the cell cycle (Figure 3.1). Cdt1 and Geminin have differing effects on DNA replication. Cdt1 is a component of the pre-replication complex and peaks in G1, just before the onset of DNA replication (Arias and Walter, 2007). The levels of Cdt1 quickly decline after the initiation of S phase. On the other hand, Geminin concentration is high throughout the S and G2 phases but declines during late mitosis (M phase), remaining low during this phase and G1 (McGarry and Kirschner, 1998). Geminin is a replication inhibitor and acts by inhibiting Cdt1 and preventing the assembly of pre-replication complexes. The quantity of these proteins are controlled by the E3 ubiquitin ligases APC/C^{Cdh1} and SCF^{Skp2} (Arias and Walter, 2007). APC/C^{Cdh1} targets Geminin for degradation and is active during mid-mitosis and remains so until the end of G1. The reciprocal expression of SCF^{Skp2} during S phase and G2 targets Cdt1 for degradation. Furthermore, SCF^{Skp2} is a substrate for APC/C^{Cdh1}, which ensures the accumulation of these proteins at the correct time.

Several different versions of the Fucci system have been used to create transgenic mice. The Fucci2a construct combines truncated versions of Cdt1 (30/120 amino acids) and Geminin (1/110 amino acids) to mCherry and mVenus fluorophores, respectively (Mort et al., 2014a). The advantage of the 2a Fucci system is the combination of both tagged proteins into a single construct, separated by a *Thosea asigna* virus 2a self-cleaving peptide. The mCherry-Cdt1 protein was found to be optimally placed in the N-terminus portion. Creating the construct in this way results in the production of equimolar amounts of each protein and a reliable signal.

This mouse has previously been used to examine hepatocyte regeneration during PHx in a study by Chen *et al.* (2019), which demonstrated that the Fucci2a system is a reliable and rapid method of visualising hepatocellular regeneration in this particular model of liver damage.

3.1.3 NEBNext® Single-Cell RNA sequencing

I hypothesised that key signals are produced by sub-populations of replicating hepatocytes during liver injury. I therefore sought to use scRNA-seq to analyse the replicating and non-replicating hepatocyte populations of mice that underwent PHx. The use of the Fucci2a mouse enabled the efficient sorting and enrichment of replicating hepatocytes for subsequent analysis. As will be discussed below, the method of scRNA-seq was restricted to a plate-based approach.

The NEBNext® scRNA-seq method is a high-quality, full length, plate-based sequencing approach. This form of scRNA-seq provides full length transcript data from cells sorted into 384 well plates. Technical notes from New England Biolabs® Ltd. (NEB) suggest that this form of library preparation gives high concentrations of cDNA yield per cell, as well as a high number of consistently identified transcripts and good transcript coverage (New England Biolabs, 2018). These are all qualities that are expected and required in order to obtain good quality scRNA-seq data.

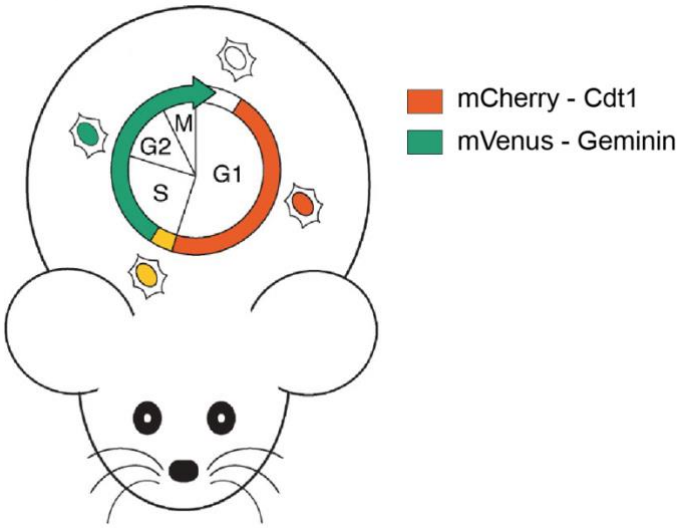


Figure 3.1 – Fucci2a Reporting Schematic

Schematic showing the expression and accumulation of the mCherry – Cdt1 signal and mVenus – Geminin signal during the associated stages of the cell cycle. The yellow area defines the small overlap of dual positive cells. Adapted from Bertero and Vallier (2015).

3.2 Results

3.2.1 Characterisation of 2/3rd Partial Hepatectomy in Mice

To assess the dynamics of hepatocyte replication, I initially characterised the proliferation of hepatocytes over time in mice following PHx. PHx was performed on C57/BL6 wild type (WT) mice. EdU was injected intraperitoneally 3h prior to harvesting the liver. Mice were terminated at 24h, 48h, 72h, 96h, and 168h and the caudate lobe dissected. Liver from uninjured mice was included to represent the quiescent, homeostatic state. Hepatocytes were identified through HNF4α staining. E-cadherin is a well-known marker that defines the periportal region of the liver and was used to define the periportal zones (Hempel et al., 2015). Negative E-cadherin areas were defined as pericentral zones.

To quantify the staining, QuPath software was utilised. Cells were detected based on DAPI signal using an in-built cell detection algorithm. A cell-based classifier was created to identify HNF4α$^+$ cells, and a threshold for EdU was built into the classifier to identify those that were replicating. To assess the possible zonal distribution of hepatocyte proliferation, a pixel classifier was trained to divide the whole slide section into E-Cadherin$^+$ (periportal) and E-Cadherin$^-$ (pericentral) regions. Combining the output of these two classifiers allowed for an automated, semi-bias method of quantifying hepatocyte proliferation dynamics. Representative images from each timepoint are shown in Figure 3.2.

A significant increase in hepatocyte proliferation was observed at 48h post PHx compared to all other timepoints, which is consistent with previous reports (Figure 3.3A; Zou *et al.*, 2012). A downward trend of proliferation occurring between 72h and 168h was observed, which again fits with previous studies (Figure 1.2A; Zou *et al.*, 2012). Taken together, these results demonstrated that I was able to reliably reproduce the key features of PHx that have previously been shown. More specifically, it was clear that between 24

to 72h post PHx was a key time for hepatocyte regeneration. I therefore focused on these timepoints to identify heterogeneity and any novel signalling patterns in the regenerative response.

To assess any potential zonation in hepatocyte proliferation, I used E-cadherin staining to define a boundary between the periportal and pericentral regions. This eliminated the bias of defining such a margin by eye and instead relied on a biologically relevant definition. Furthermore, this removed any variability that arose as a result of sectioning by comparing biologically defined regions between each liver section.

Firstly, E-cadherin^{+} regions were quantified as a percentage of the total area classified. There was no significant difference in the percentage of E-cadherin^{+} area (Figure 3.3B). Proliferation was calculated as a percentage of the total number of hepatocytes per zone (i.e. E-Cadherin^{+} zone or E-Cadherin^{-} zone). No significant differences were seen in the percentages of proliferating hepatocytes between the E-Cadherin^{+} and E-Cadherin^{-} areas at any timepoint (Figure 3.3C). This suggests that at all timepoints measured, hepatocyte proliferation is proportionally equal across the pericentral and periportal zones.

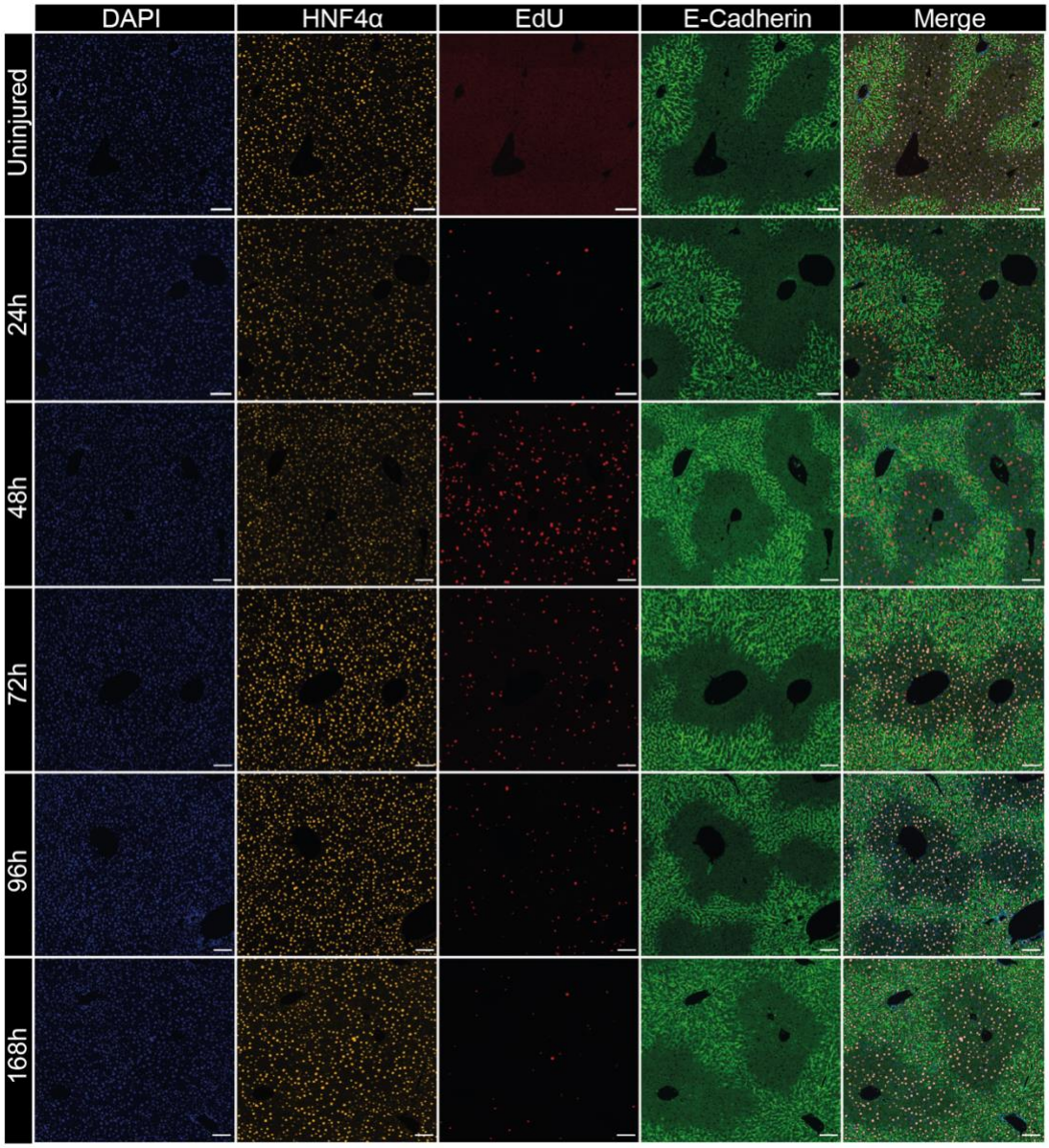


Figure 3.2 – EdU, E-Cadherin and HNF4α Staining in Mouse Livers following 2/3 Partial Hepatectomy

Representative immunofluorescent images of liver sections from uninjured and 24h, 48h, 72h, 96h and 168h post partial hepatectomy. Staining shows HNF4α, EdU, E-cadherin and DAPI. Scale bar = 100μm.

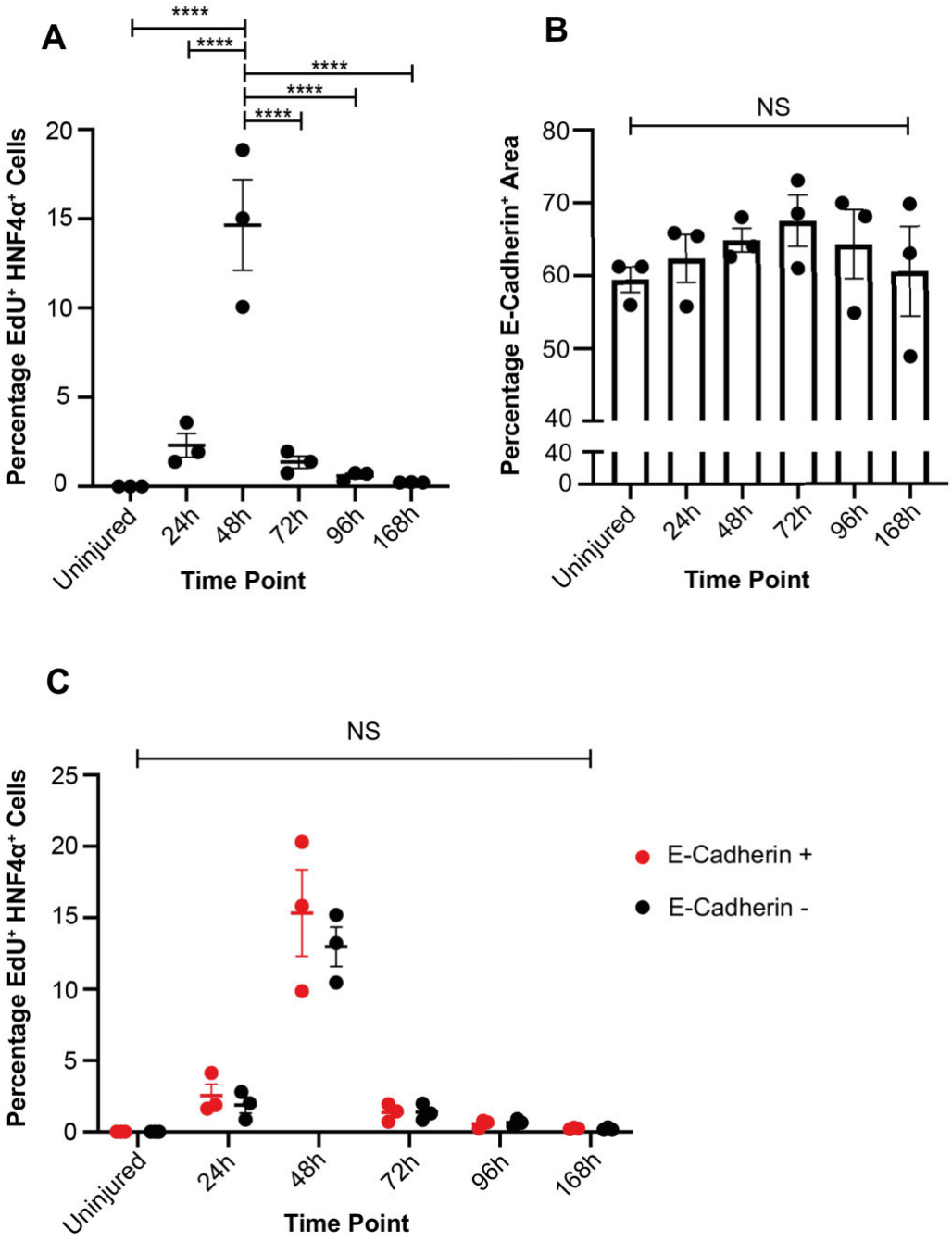


Figure 3.3 – Analysis of Hepatocyte replication dynamics following partial Hepatectomy

A) Quantification of HNF4α⁺ EdU⁺ cells in uninjured mouse liver sections and 24h, 48h, 72h, 96h and 168h post partial hepatectomy. Percentage is measured as the total number of HNF4α⁺ cells per whole liver section.

B) Quantification of E-cadherin$^+$ area in uninjured mouse liver sections and 24h, 48h, 72h, 96h and 168h post partial hepatectomy. Data shown as a percentage of the total area classified.

C) Quantification of HNF4α$^+$ EdU$^+$ cells in uninjured mouse liver sections and 24h, 48h, 72h, 96h and 168h post partial hepatectomy. Data shown as a percentage of total HNF4α$^+$ cells per E-cadherin$^+$ or E-cadherin$^-$ area.

Data was collected from one experiment (n=3). Data shown is mean +/- standard error for each group. A one-way ANOVA was performed to test for differences between timepoints (A & B). A two-way ANOVA was performed to assess the difference between E-cadherin$^+$ and E-cadherin$^-$ groups (C). A Bonferroni multiple comparisons post-test was used to assess differences between groups (C). **** $p < 0.0001$

3.2.2 Optimisation of Hepatocyte Isolation Protocol for Single-Cell RNA Sequencing

One of the key elements of any experiment in which the aim is to measure a parameter of a cell *ex vivo* is the speed at which it is isolated. This ensures that the sample is stable and minimises the chances of changes occurring during processing. In this case, the aim was to lyse the cells as quickly as possible, preventing RNA degradation and transcription of any new genes that result from the isolation procedure. Therefore, my aim was to reduce the time taken to isolate a population of hepatocytes from the mouse liver, ready for single cell isolation.

The original protocol employed by the laboratory that was used to isolate hepatocytes was already a very rapid procedure. However, there was one step I felt that could be omitted (Figure 3.4): the percoll centrifugation, and subsequent wash steps, which aimed to remove dead cells and debris from cell suspensions, as well as enrich for hepatocytes. This step is one of the most time-consuming (with the exception of *in-situ* digestion) and is

predominantly used when isolating hepatocytes for cell culture experiments where no flow cytometry-assisted cell sorting is performed. However, I optimised an isolation protocol in which I sorted live hepatocytes using a FACS sorter to obtain a pure population of live hepatocytes. As a result, the use of percoll became somewhat redundant. Furthermore, this step was performed at room temperature (RT), whilst the remainder of the protocol was carried out at 4^{o}C to slow any RNA degradation. Incorporating a percoll gradient would require that either a) the cells be kept at RT up until this point, potentially allowing for a higher degree of RNA degradation and changes to the hepatocyte transcriptome, or, b) increasing the temperature of the cells back up to RT from 4^{o}C and back down again, which could cause unnecessary stress to the cells and again potential affect the transcriptome. It has also previously been demonstrated that percoll purifications reduce the yield of viable cells (Horner et al., 2019). Removal of any cells at this point based on a density gradient could bias the hepatocyte population and prevent the identification of any rare subpopulations. For these reasons, I decided that removal of the percoll purification gradient would be more beneficial to ensure minimal disruption to the *in-situ* state of the hepatocytes (Figure 3.4). Several other studies have taken a similar approach (Halpern et al., 2017; MacParland et al., 2018).

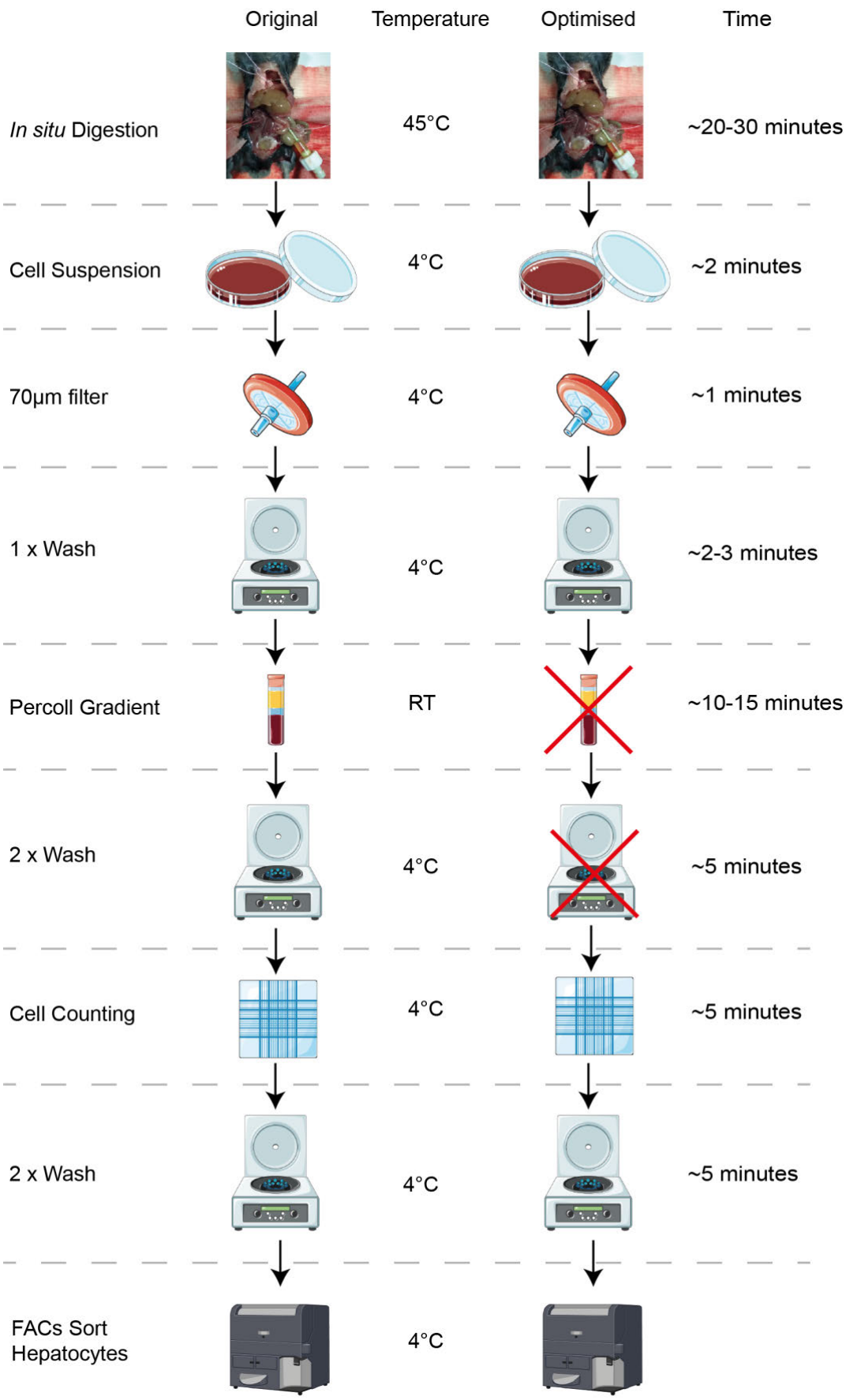

Original
Temperature
Optimised
Time
In situ Digestion
45°C
~20-30 minutes
Cell Suspension
4°C
~2 minutes
70µm filter
4°C
~1 minutes
1 x Wash
4°C
~2-3 minutes
Percoll Gradient
RT
~10-15 minutes
2 x Wash
4°C
~5 minutes
Cell Counting
4°C
~5 minutes
2 x Wash
4°C
~5 minutes
FACs Sort
Hepatocytes
4°C

Figure 3.4 – Diagrammatic Representation of Hepatocyte Isolation.

The key steps for isolating primary mouse hepatocytes are depicted with temperature and time taken shown for each step. Original preparation refers to the protocol that was used prior to optimisation. Optimised protocol depicts the steps that were removed (red crosses) to allow faster isolation.

3.2.3 Droplet based Single-Cell RNA Sequencing of Mouse Hepatocytes During Partial Hepatectomy

Before proceeding with a time course sequencing experiment of mouse hepatocytes following PHx using the 10x Genomics Chromium droplet-based scRNA-seq platform, I decided to generate data from a single sample of uninjured hepatocytes to assess the quality of the data I could obtain using this technique. Live mouse hepatocytes were isolated as previously described in section 2.6 and sorted as shown in section 2.11. Cells were then filtered using a 40µm Flowmi™ cell strainer, as recommended in the 10x Genomics sample preparation guide, prior to re-counting and loading of 7000 cells into the 10x Genomics Chromium controller. This filtering step was deemed necessary to ensure that the hepatocytes were able to flow through the microfluidics of the 10x Genomics chip and prevent the machine from clogging. The reported max diameter of the chip channel width is ~50µm. Mouse hepatocyte are a large cell type and I have recorded mouse hepatocyte diameters at a mean of 28.02±0.5 (S.E.M, n=116), with some hepatocytes as large as 45µm, in suspension. Therefore, doublets and larger hepatocytes could potentially block the chip and result in an unsuccessful droplet capture. Samples were sent to Edinburgh Genomics for sequencing, and the resulting data was aligned and analysed using the Cell ranger pipeline from 10x Genomics.

Unfortunately, the resulting data was unusable. When loading 7000 cells, the expected return in cell number is around 5,000 when taking into consideration the capture efficiency of the 10x Chromium system. However, the QC report from the Cell ranger pipeline showed that ~65,000 cells were

identified (Figure 3.5). Only 177 median genes and 6,101 mean reads were found per cell.

Over 80% of the reads had a Q-score of 30 or above, which is indicative of good quality sequencing. Furthermore, the majority of reads were confidently mapped to the transcriptome. This suggests that the issues arising with the data are not related to the sequence quality. The very large number of cells detected most likely originates from a large concentration of ambient, or extracellular RNA that was captured in droplets when the cells were being isolated. It is possible that large hepatocytes were lysed prior to capture, either by the filtering process or the microfluidic pressure within the 10x Chromium Controller, thereby releasing their RNA. A proportion of the cells captured are likely to be intact hepatocytes. However, most of the sequencing reads and power were likely taken up by the large number of empty droplets containing ambient RNA. As a result, with very few reads sequenced for these cells, the cell detection algorithm of Cell Ranger was unable to confidently distinguish the empty droplets and the intact hepatocytes. Overall the sample had a saturation level of 93.4%, although, this is for the sample as a whole. It is therefore likely that the captured intact cells are under sequenced. Attempting to filter these cells out manually would be challenging and even if successful, the results would be underpowered.

Due to the inability of hepatocytes to survive the 10x Genomics droplet capture procedure, I decided that a plate-based approach would be a more appropriate method. This involves sorting cells into individual wells which removes the possibility of capturing ambient RNA in empty droplets.

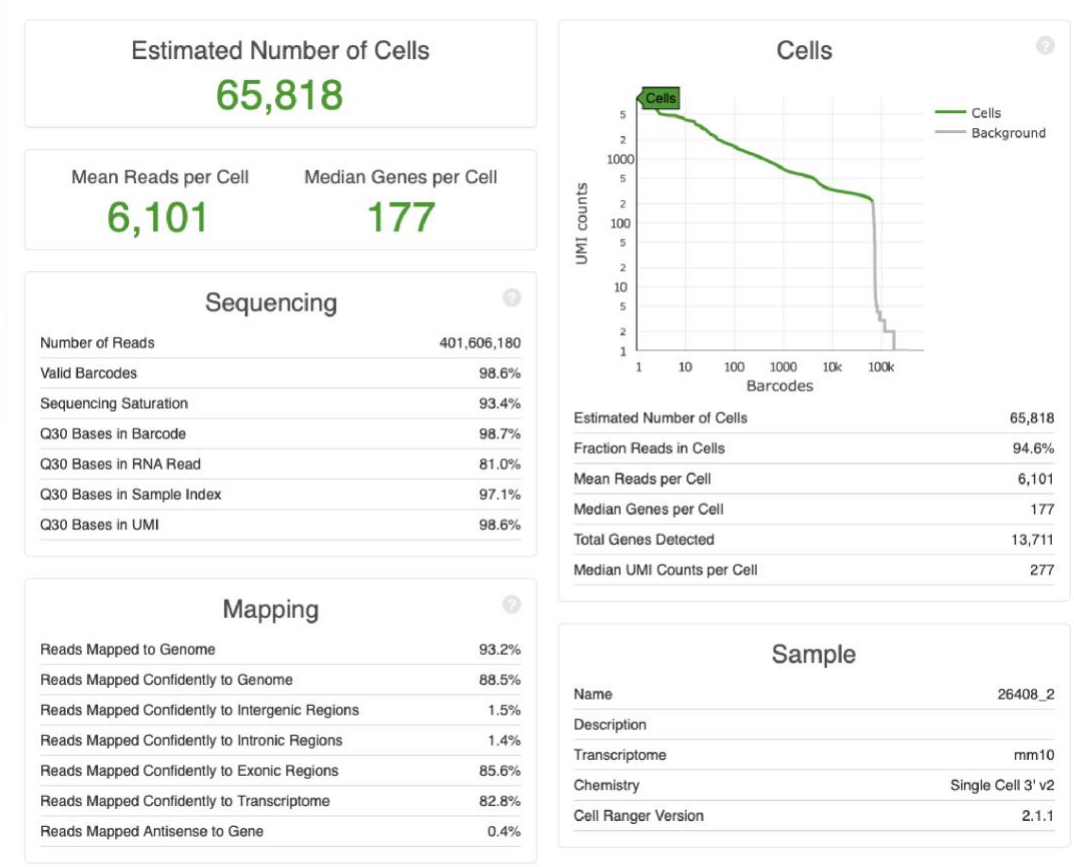


Figure 3.5 – Quality Control Output from Cell Ranger Pipeline of 10x Genomics Single Cell RNA Sequencing of Uninjured Mouse Hepatocytes

Quality control output of 10x Genomic sequencing of uninjured mouse hepatocytes. Estimated number of cells detected alongside the median number of genes and the mean number of reads per cell are shown. Cells – A knee plot shows the distribution of detected UMIs over “cell” barcode. The green line shows the area of the density plot in which cells were detected, while the grey line represents the background non cell containing droplets. Sequencing – Total number of reads, as well as percentages of reads, barcodes, indexes and UMIs that meet the Q30 quality score, which is used to defined good quality reads, are shown. Sequence saturation represents the percentage level at which the sequencing has captured the available reads in the sample. Mapping – Confidently (uniquely) mapped read percentage to the Genome, Intergenic regions, intronic regions, exonic regions and transcriptome are shown.

3.2.4 Isolation of Mouse Hepatocytes Based on Cell Cycle during Partial Hepatectomy

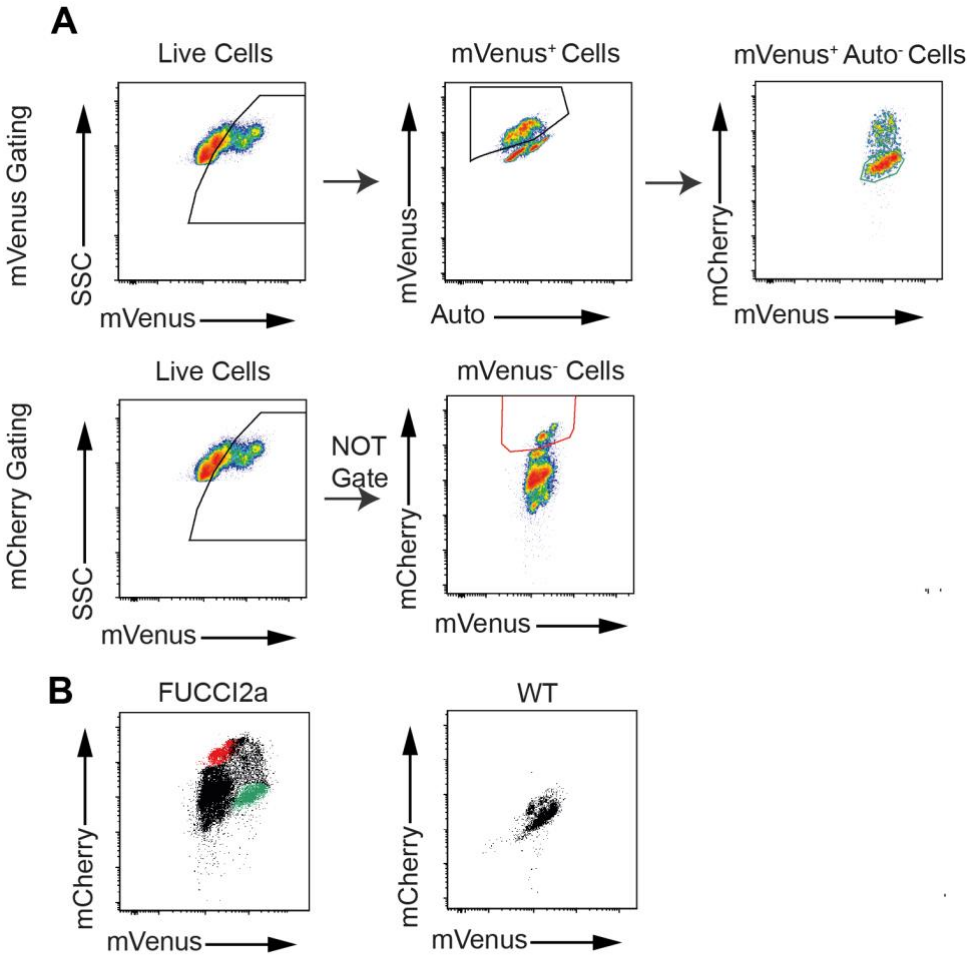
Due to the inability to sequence hepatocytes using a droplet-based methodology, I decided to use a plate-based approach, which allows for hepatocytes to be sorted straight into the well of an Eppendorf TwinTec 384 well PCR plate. As shown in section 3.2.2, ~20% hepatocyte proliferation was observed at 48h post PHx. If hepatocytes were sorted with no enrichment for those replicating, the number of such cells captured would be insufficient to examine the potential heterogeneity in this population. To mitigate this problem, I used the Fucci^{+} mice, which constitutively express the Fucci2a construct in all cells, including hepatocytes. As described previously, this construct allows for the identification of replicating, mVenus expressing cells from the non-replicating, mCherry expressing cells. To assess if the Fucci fluorescent signals could be identified by flow cytometry and sorted by FACS, I performed PHx on Fucci^{+} mice and isolated hepatocytes 48h later. WT mice that also underwent PHx were used as a negative control for the mCherry and mVenus signal.

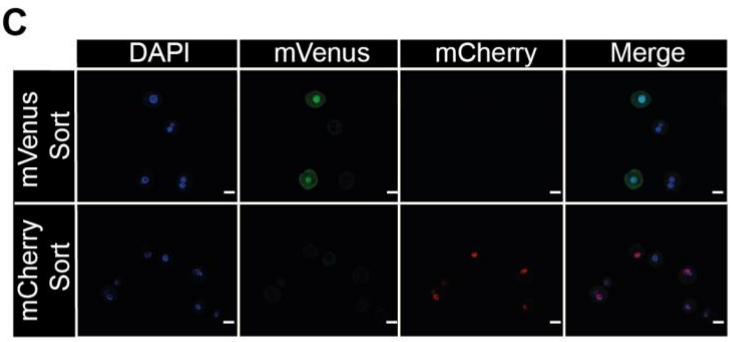
The gating strategy for live (Draq7^{-}) single hepatocytes is outlined in Figure 2.1. Following this, the non-replicating population was defined as mCherry^{+}/mVenus^{-} cells, whereas those replicating were identified as mVenus^{+}/mCherry^{-}/Auto^{-} cells (Figure 3.6A).

The replicating hepatocytes were difficult to accurately identify as they were highly auto-fluorescent in a similar part of the spectrum compared to the mVenus emission. To overcome this and ensure the specific capture of replicating hepatocytes, I used the blue laser to excite the mVenus fluorophore and a 525/50 detector to measure the mVenus signal (and potential auto-fluorescence), together with a 695/40 detector to discern the auto-fluorescence signal. By comparing the intensity captured by these two detectors, I was able to distinguish the auto-florescent cells from mVenus^{+} hepatocytes. More specifically, auto-fluorescent cells had a linear relationship,

while the mVenus^{+} cells were identifiable by a stronger mVenus signal (Figure 3.6A). This is because the signal from the mVenus fluorophore has a more specific peak when using the B525/50 detector while the auto-fluorescent will produce a more widespread emission across multiple detectors.

Using this gating strategy, I was able to see clear replicating cells (mVenus^{+}/mCherry^{-}/Auto^{-}, hence referred to as mVenus^{+}) and non-replicating cells (mCherry^{+}/mVenus^{-}, hence referred to as mCherry^{+}; Figure 3.6B). In order to validate these observations, I sorted the two populations and imaged them using a confocal microscope (Figure 3.6C). Sorted cells were imaged suspended within a droplet of cell suspension on a coverslip and I subsequently counted, per population, how many mVenus^{+} and mCherry^{+} nuclei I could identify. While no mVenus signal was detected in the mCherry population and vice versa, I was unable to identify 100% of cells showing mVenus and mCherry positivity from their respective populations (Figure 3.6D). Only 48% of the mVenus population were mVenus^{+}, and only 25% of the mCherry population were mCherry^{+}. There are several potential explanations for this loss of signal, including; bleaching of the fluorophores, out of focus light from imaging inside a droplet, loss of signal from potentially fragile hepatocytes, or decreased sensitivity to detect the fluorophore from one technique to the next. Nonetheless, I was confident that the results from the FACS sorter were accurate and proceeded to use these mice and gating strategy to isolate the replicating and non-replicating hepatocyte populations.


A
Live Cells
mVenus+ Cells
mVenus+ Auto- Cells
mVenus Gating
SSC
mVenus
Auto
mCherry
Live Cells
mVenus- Cells
mCherry Gating
NOT
Gate
B
FUCCI2a
WT


C
DAPI
mVenus
mCherry
Merge
mVenus Sort
mCherry Sort

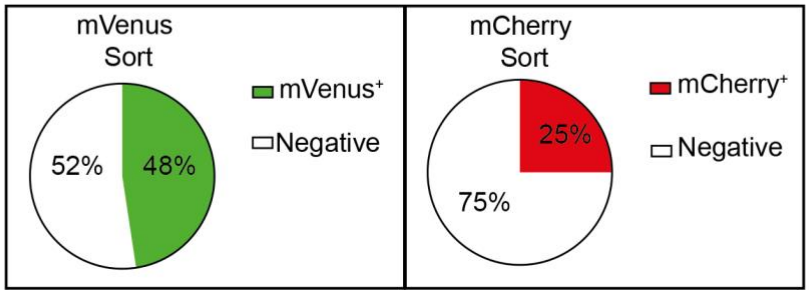

D
mVenus Sort
mVenus+
Negative
52%
48%
mCherry Sort
mCherry+
Negative
25%
75%

Figure 3.6 – Sorting Strategy of Replicating and Non-Replicating Cells from Fucci$^+$ mice at 48h Post-Partial Hepatectomy

A) Live, single hepatocytes were gated prior to the strategy shown. Gating strategy for sorting replicating (mVenus$^+$/mCherry$^-$/Auto$^-$, referred to as mVenus$^+$) and non-replicating (mCherry$^+$/mVenus$^-$, referred to as mCherry$^+$) hepatocytes, using mVenus, auto-fluorescent (Auto) and mCherry signals. "Not gate" refers to cells that are not contained within the gate shown. Title above each plot refers to parent gate.

B) Comparison of mVenus and mCherry signals produced by Fucci$^+$ mice and wild type (WT) mice, 48h post partial hepatectomy. All cells are live, single hepatocytes. Red dots represent the gated, non-replicating cells (mCherry$^+$), while the green dots represent those replicating (mVenus$^+$).

C) Representative confocal images of the mVenus and mCherry signals from sorted mVenus$^+$ and mCherry$^+$ populations. Scale bar = 20μm.

D) Quantitation of mVenus$^+$ and mCherry$^+$ nuclei, identified via confocal imaging, in the mVenus and mCherry populations shown in (C).

3.2.5 Sorting Efficiency of Live Hepatocytes into 384-Well Plates

Having made the decision to move to a plate-based sequencing approach (NEBNext® scRNA-seq), it was important to test the efficiency of sorting single hepatocytes into single wells. Uninjured hepatocytes were isolated and stained with Hoechst 33342, to allow for cellular identification post sorting. Single live hepatocytes were sorted into wells of a 384 well plate (Figure 2.1). Two independent repeats were conducted. Each well was then examined to identify how many hepatocytes were in each well, and recorded as singlets, doublets, empty wells or unknown (Figure 3.7A). Unknown wells were defined as such where it was difficult to establish if there were two hepatocytes or a binucleate cell. Fifteen cells were sorted into the first well as an easily identifiable reference. Very few doublets were observed (~1%), with no wells containing more than two cells (Figure 3.7B). Furthermore, less that 1% of wells were labelled as unknown in both replicates. The number of wells

containing a single hepatocyte were 44% and 56% respectively for each repeat. The remaining wells contained no visible hepatocytes.

A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	15	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	0	1
B	1	1	1	1	0	1	2	0	1	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0
C	0	0	1	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	1	1	0	1	0
D	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	1	1	1	1	0	1	0	0
E	0	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0
F	0	0	0	1	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
G	0	1	1	0	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1
H	1	1	0	1	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	1	1	1	1	0
I	1	1	1	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1	0	1	0	0	1	1
J	1	0	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	0	0	0	1	1	0	0
K	1	0	0	0	0	1	0	1	1	1	1	0	0	0	0	1	1	1	0	0	0	1	0	0
L	1	2	1	0	0	1	1	0	0	?	0	1	1	0	0	1	1	0	1	1	1	0	0	0
M	0	1	1	0	1	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	0	1	0	0
N	1	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	1	1
O	1	0	1	0	1	0	0	1	1	1	2	1	0	1	1	0	1	0	0	0	1	0	0	0
P	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	0	1	1

B

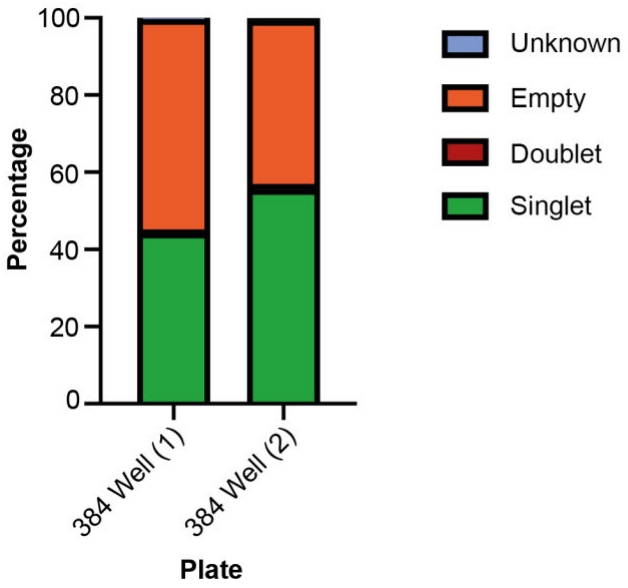


Figure 3.7 – Sorting Efficiency of Hepatocytes using the Becton Dickinson Aria II Fluorescent Activated Cell Sorter

A) A representation of the single cell sorting of live hepatocytes into single wells of a 384 well plate. Cell A1 had 15 hepatocytes sorted to act as a reference point. Empty wells are orange and labelled with a “0”. Wells with a single hepatocyte are labelled with a “1” and are green. Wells containing two wells are labelled with a “2” and are red. Wells coloured blue and labelled with a “?” are wells with unknown number of hepatocytes.

B) Bar plot showing the percentage of wells containing either a single hepatocyte (Green), two hepatocytes (Red), no hepatocytes (Orange) or an unknown number (Blue), from two independent sorts of single, live hepatocytes into single wells of a 384 well plate using a BD Aria II sorter.

3.2.6 Sorting of Replicating and Non-Replicating Hepatocytes for NEBNext® Single Cell RNA Sequencing

With the protocol fully optimised, I performed PHx on four Fucci$^+$ mice. WT mice were included to confirm the gating was accurate for each sort. Hepatocytes from mice at 24h, 48h and 72h post PHx were sequenced, as these represent the timepoints that cover the peak of the hepatocyte regenerative response (Figure 3.8A; Zou *et al.*, 2012). Uninjured hepatocytes were also sequenced as a comparison to quiescent hepatocytes. One mouse was used per timepoint. I decided to limit the number of hepatocyte isolations to two mice per day, one Fucci$^+$ and one WT mouse, to ensure high quality cells were sequenced (i.e. no cells were sitting on ice for long periods of time while waiting to be processed). To further minimise tissue to cell capture time, I processed WT controls first. As a result, only one timepoint was sorted per day, which meant it was not possible to sort multiple timepoints into any one 384 well plate. One plate per timepoint was sorted for sequencing, with the exception of 48h, for which two plates were sequenced, from the same mouse, to increase the number of cells at this critical timepoint (Figure 3.8A). Index sorting was performed for all plates so that the mCherry and mVenus signals could be integrated with the scRNA-seq data. Only non-replicating mCherry$^+$ cells were sorted for the uninjured timepoint. This was due to the very limited number of hepatocytes that replicate in a quiescent liver. As such, time taken to sort a reasonable number of replicating cells would have been impractical. Furthermore, due to the small percentage of mVenus$^+$ only hepatocytes at the 24h and 72h timepoints, I decided to widen my gating strategy to sort the mVenus$^+$/mCherry$^+$ (Dual$^+$) as well as the mVenus$^+$ only cells (at a ratio of 30% Dual$^+$ & mVenus$^+$ to 70% of mCherry$^+$ only hepatocytes; Figure 3.8A &B). This

kept the sort time to a minimum whilst ensuring enough cells were sorted. These dual positive cells represent hepatocytes that have begun to, or will subsequently, enter S phase and should ultimately replicate their genome. One might therefore consider them an important population of hepatocytes that is worth analysing in the context of regeneration. With the higher percentage of replicating hepatocytes at 48h, mVenus^{+} only hepatocytes were sorted at a ratio of 60% to 40% of mCherry^{+} only cells. Hepatocytes were sorted directly into the lysis buffer contained in each well, centrifuged and frozen immediately on dry ice before being stored at -80^{o}C. Each plate took approximately 30 minutes to sort. Plates were shipped to the Sanger Institute for NEBNext® scRNA-seq library preparation and sequencing using the NovaSeq 6000.

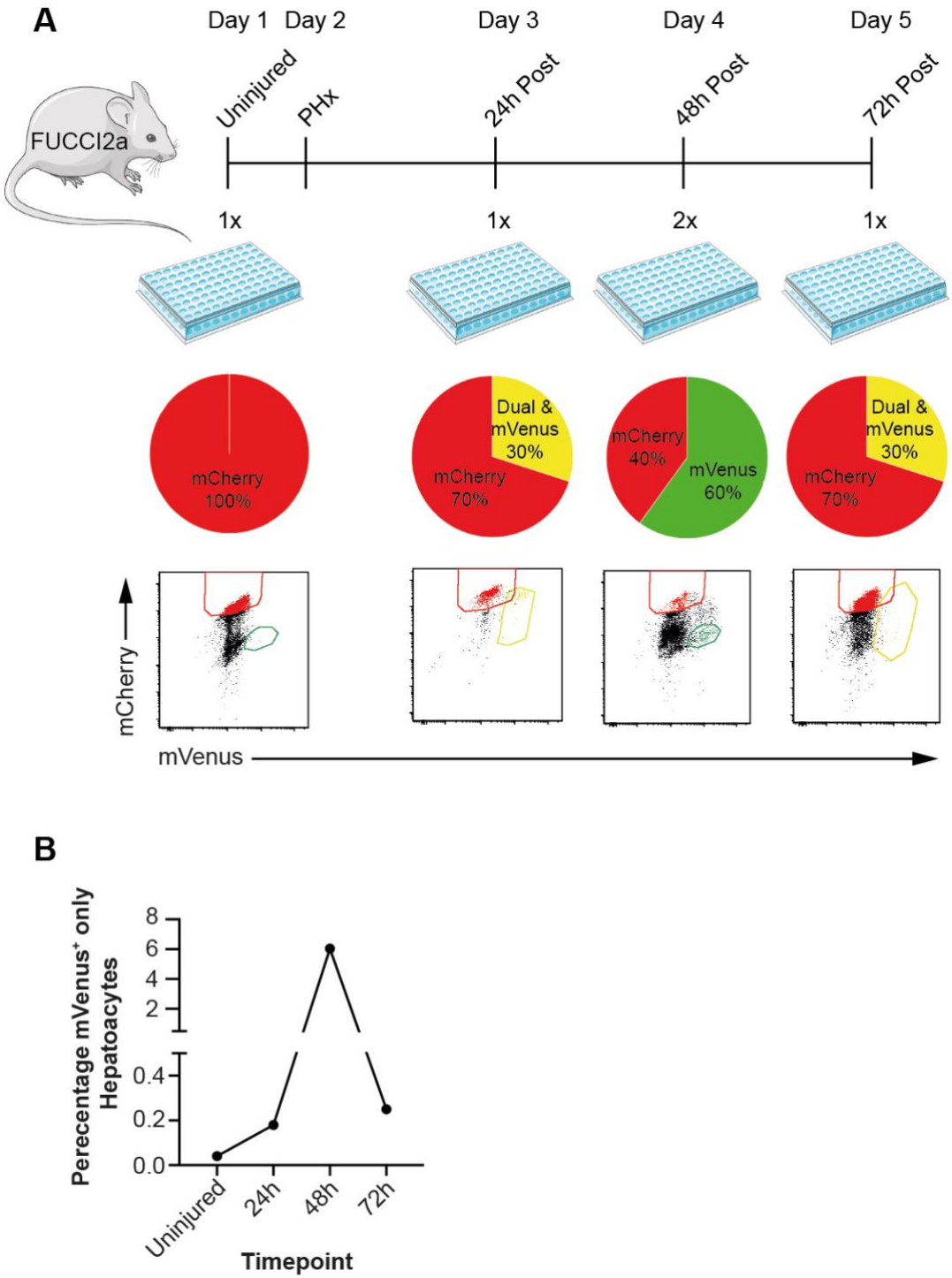


Figure 3.8 – Experimental Design of Sorting Replicating and Non-Replicating Hepatocytes Following Partial Hepatectomy Time Course for NEBNext® Single Cell RNA Sequencing Experiment

A) Pictographic representation of experimental design. Hepatocytes from uninjured mice were isolated and sorted into a 384 well plate on day one. Partial hepatectomies were carried out on day two. Hepatocytes

were isolated and sorted into 384 well plates on days three, four and five. The number of plates and the percentage of mVenus$^+$ only, mCherry$^+$ only and Dual$^+$ plus mVenus$^+$ only hepatocytes contained in each plate are shown. Representative flow plots of the populations sorted are shown, with non-replicating mCherry$^+$ only hepatocytes (red gate), replicating dual$^+$ plus mVenus$^+$ only hepatocytes (yellow gate) and mVenus$^+$ only hepatocytes (green gate).

B) Quantification of the percentage of mVenus$^+$ only hepatocytes using flow cytometry for uninjured hepatocytes and those isolated 24h, 48h and 72h post partial hepatectomy.

3.3 Discussion

Initial characterisation confirmed the replication dynamics of mouse hepatocytes post PHx. Furthermore, I successfully developed a protocol for the rapid isolation of mouse hepatocytes and the sorting of these cells based on cell-cycle.

I confirm 48h post PHx to be peak hepatocyte proliferation. Although this is in line with previous reports (Y. Chen et al., 2019; Miyaoka et al., 2012), performing additional timepoints between 24h and 48h may have identified the peak of hepatocyte proliferation to lie between these two times. However, examining timepoints between 24h and 48h would have been logistically challenging. Furthermore, using these timepoints reduced the number of mice being used which is in accordance with the guidelines stated in the three Rs. Therefore, the decision was made to proceed with the 48h timepoint.

In this study, I found no difference in the proliferation between liver zones. As discussed previously, several studies have found a portal to central wave of hepatocyte proliferation over time following PHx. The first study, conducted in rats, used [^{3}H]thymidine to analyse replicating hepatocytes and split the liver lobule into 9-11 sections depending on size (Rabes et al., 1976). However, dividing the liver in this way results in high variation in region size between samples, making it difficult to compare one section to another. In

addition, research into the use of [^{3}H]thymidine has revealed it can inhibit DNA synthesis and therefore is not a reliable measure of DNA replication (Hu et al., 2002). As such, results from this study should be interpreted with caution. Furthermore, there are known differences in the temporal replication dynamics of hepatocytes in mice compared to rats. For instance, peak hepatocyte proliferation is observed at 24h in rats, as opposed to 36-48h in mice (Michalopoulos, 2007; Zou et al., 2012). It is therefore difficult to make meaningful comparisons between this study and my work. However, Sun *et al.* (2020) recently published data to suggest that this proliferation wave could also be observed in mice following PHx. In this study, the liver lobule was also divided manually into three 100μm regions around the portal triad and central vein, with the remaining area classified as parenchyma. Although the authors did not specify whether the areas that were measured were of equal size. As discussed above, dividing the lobule into regions manually can introduce bias and result in uneven comparisons. In addition, neither study used any marker to identify hepatocytes and therefore one can only assume that they relied on morphological aspects for classification.

In my study, to assess the spatiotemporal dynamics and overcome the issues described above, I used E-Cadherin, a known marker of the periportal region, and HNF4α as a marker of hepatocytes (Hempel et al., 2015). Classifiers were trained in the QuPath software to enable a less bias quantification method than manual counting, and also permit the assessment of whole slide sections as opposed to regions of interest. By comparing HNF4α$^+$ EdU$^+$ cells in E-Cadherin$^+$ areas, I was able to compare similarly defined regions regardless of liver lobule orientation. Using this method, I found that there was no significant change in the percentage of E-Cadherin$^+$ area across the time course. Furthermore, there was no significant difference in proportional proliferation between the E-Cadherin$^+$ area (periportal) or E-Cadherin$^-$ region (pericentral). Taken together, this indicates that at all timepoints, there is no bias of hepatocellular proliferation in one region over the other.

Following hepatocyte isolation optimisation for scRNA-seq, I was unable to generate useable data from a 10x Genomics droplet-based approach. It is likely that the principal cause was bursting of hepatocytes and capture of ambient RNA in empty droplets. There are several possible explanations for this, including the process of sorting, the filtration prior to 10x Genomics Chromium chip loading and/or the microfluidics of the 10x Genomics Chromium Chip. However, omitting the cell sorting step would have required the introduction of an additional process in order to remove dead hepatocytes, such as the re-introduction of the percoll gradient centrifugation. This was not considered a viable option due to the reasons previously discussed (i.e. heat shock and change to hepatocyte transcriptome). Removal of the filtration prior to chip loading was also considered, although doing so would likely have caused blocking of the microfluidics in the chip as outlined in the 10X sample preparation manual. For these reasons, I turned to a plate-based approach.

Plate-based approaches allow a higher level of sequencing depth per cell than that of the 10x Genomics method (Ziegenhain et al., 2017). However, given the higher sequencing costs per cell, the number of cells sequenced in a plate-based workflow tends to be far fewer than in high-throughput droplet-based scRNA-seq protocols. To ensure the capture of hepatocytes of interest (replicating), I decided to use the Fucci^{+} mouse line, which contains the Fucci2a construct. Using this reporter mouse allowed for enrichment of replicating cells, at all-time points. This was key to the experiment; if sorting was left to random sampling, there would not have been enough cells to make any valid conclusions regarding the replicating population.

Before sorting hepatocytes for sequencing, I assessed the sort efficiency. I was able to sort a single live hepatocyte into ~50% of the wells. Although a greater sort efficiency was expected, empty wells pose less of a problem in terms of informatic analysis compared to a doublet. It is harder to distinguish a doublet than an empty well within the resulting scRNA-seq data. Therefore, this low number of doublets was optimal. It is likely the high number of empty wells was a result of hepatocytes' tendency to settle and clump over

time. Indeed clumping of cells is one potential reason for poor sorting efficiency (Cossarizza et al., 2017).

I have found throughout my work with hepatocytes that they are a very fragile cell type. For this reason, it was crucial I minimise processing time. Reducing this time reduces the possibility of any RNA degradation and changes to the hepatocyte transcriptome. This is in line with previous single cell studies in which isolation protocols were as short and gentle on the cells as possible (Halpern et al., 2017; MacParland et al., 2018). The fragility of hepatocytes also had a bearing on the choice of proportion of replicating and non-replicating cells to be FACS sorted for sequencing. A 60% to 40% split of replicating over non-replicating hepatocytes was chosen for 48h, as this was the timepoint with the highest proportion of replicating hepatocytes. Therefore, the time taken to sort a higher percentage was not as long as at 24 and 72h. The slightly higher percentage of replicating cells over non-replicating was chosen to provide a greater chance at identifying heterogeneous pro-regenerative hepatocytes. However, the lower percentage of hepatocellular proliferation at 24h and 72h limited the number of replicating hepatocytes that could be sorted at these times points in a reasonable timeframe. Therefore, I deemed 70% to 30% split as a reasonable trade-off between isolating sufficient replicating cells for analysis and the time taken to sort. Dual positive cells were also included in the sorting of the replicating hepatocytes for these two time points to decrease sort time. These cells represent hepatocytes that have committed to replication and the formation of a regenerative niche (Zielke and Edgar, 2015). Index sorting of each plate enabled me to classify the hepatocyte of each well as being either mVenus^{+} only or dual positive. This enabled the correlation of these hepatocytes with any potential heterogeneity found within the data.

In summary, I was able to utilise the PHx model to isolate and sort replicating and non-replicating hepatocytes across a PHx time-course. The data generated from this experiment allowed me to assess the potential heterogeneous responses of these populations and subsequently identify key signals that are produced to regulate regeneration and the replicative niche.

4 Single-Cell RNA Sequencing: Quality Control, Normalisation and Batch Correction

4.1 Introduction

scRNA-seq is a highly technical process and for this reason, there are several quality control (QC) steps that are required to ensure that good quality data is being produced and analysed. During the library preparation stage of the NEBNext® single-cell RNA seq protocol, RT-PCR and PCR amplification are performed for each cell. The small concentration of RNA captured from each individual cell makes this process challenging: cells may have very little or degraded RNA, despite attempts to mitigate this (Section 3.2.2). Due to the large number of individual reactions performed when using a 384 per plate, it is impractical to measure the quality of the cDNA generated from each individual well. Therefore, a smaller number are randomly sampled from across each plate to get a rough quantification of the quality of cDNA generated. However, it is likely that some wells will contain insufficient quantities of cDNA as a result of either empty wells, or insufficient mRNA capture, RT and amplification due to mRNA degradation. The inevitable capture and sequence of poor quality cells, or empty wells, has been observed by us and others (Ilicic et al., 2016; Luecken and Theis, 2019).

The effective removal of these poor-quality cells is vital for generating accurate results downstream. To achieve this, various QC metrics are commonly inspected at a cellular level. These include number of total genes/features detected per cell (Total genes), number of total read counts per cell (Total counts), and the percentage of read counts that correspond to mitochondrial genes (Luecken and Theis, 2019). Cells that contain a low number of total counts or total genes as a result of either empty wells or degraded mRNA are considered to be poor quality, as are cells with high levels of mitochondrial read count percentages (mitochondrial percentage). While useful, these metrics should be treated with caution and on a case by case

basis. Each single cell technology and dataset may produce QC values that differ from other studies and different cell types will express different numbers of genes (Ziegenhain et al., 2017). Therefore, biological relevance and a pragmatic approach should be taken when defining thresholds for each of these QC metrics.

As well as cell QC, gene QC must be performed. Any given scRNA-seq dataset may contain tens of thousands of genes with a large number of these displaying zero counts in almost all cells (Silverman et al., 2020). ScRNA seq by nature contains a large number of zero counts due to both biological and technical variation (Silverman et al., 2020). Lowly expressed genes that are only expressed in a very small number of cells are generally considered to be uninformative to cellular heterogeneity and only add noise to the data. These genes are therefore removed. Two approaches have been considered. First, a single average approach can be taken in which the expression of each gene is averaged across all cells and then log transformed (Lun et al., 2016). Genes that do not meet a certain average log expression are considered to be uninformative. A second approach one could employ is to filter genes by removing those that don't meet a certain count value per number of cells criterion (Luecken and Theis, 2019; Stuart et al., 2019). For example, all genes must have a count value of five in at least three cells. This is easier to conceptualise and a threshold can be chosen based on the smallest cluster of cells that is expected to be found. For example, when looking for rare populations, removing genes that are expressed in fewer than 10 cells may remove those that are markers for rare cell population.

Once cells and genes have been filtered based on quality, the data is then normalised. The main aim of any normalisation is to effectively correct for library size (total counts), whilst retaining as much biological variance as possible. There are many different normalisation methods that have been suggested over the past few years to tackle the unique challenges of scRNA-seq data (Vieth et al., 2019). The large number of zeros in scRNA-seq data cause issues for standard bulk RNA seq methods of normalisation. One of the first scRNA-seq specific methods was developed for the *Scater* analysis

pipeline (L. Lun et al., 2016). This method involves pooling similar cells and averaging their gene counts. The pseudocell is normalised to an average reference of all cells within a sample, before deconvolving individual normalisation factors for each cell. The act of pooling the cells reduced the incidence of zeros. This method, like bulk methods, still produces a global scaling factor for each cell, whereas other methods such as SCnorm do not (Bacher et al., 2017). The creators of *Scnorm* believe global scaling factors do not account for the difference in gene dependence on sequence depth. This method therefore groups genes based on their count-depth relationship and then finds scaling factors within each group for each cell. These are just two scRNA-seq-specific approaches and it is not uncommon to find bulk methods still being used, such as counts per million (CPM), which take a summed approach to defining a global scaling factor (Stuart et al., 2019). Due to the large differences in technical variation between scRNA-seq datasets, the choice of which normalisation approach to take can be difficult. Recent frameworks such as *scone* aim to alleviate this problem by comparing multiple normalisation approaches and ranking these based on a number of different metrics (Cole et al., 2019).

One final hurdle is the effective integration of multiple scRNA-seq datasets. Technical differences can be introduced unequally when preparing and sequencing datasets. These differences lead to batch effects within the combined data, and cells from each dataset will cluster together and appear more similar to one another solely based on the dataset they originate from. Luckily, there are many tools to rectify this issue such as *Harmony* and the integration approach implemented in *Seurat* (Korsunsky et al., 2019; Stuart et al., 2019). Again, the choice of which of these methods to employ can be challenging. However, the aim remains the same as that of normalisation: removal of technical variation whilst preserving biological variation.

To help the processing of scRNA-seq datasets, multiple pipelines and frameworks have been developed and implemented in the *R* statistical programming language. Examples include *Seurat*, *scater* and *monocle* (Cao et al., 2019; McCarthy et al., 2017; Stuart et al., 2019). Each differ in their

approach and whilst they work reasonably well alone for any single dataset, a more bespoke pipeline is usually required to fully utilise a dataset. The following chapter describes how I performed the QC, normalisation and batch correction of my scRNA-seq datasets.

4.2 Results

4.2.1 Quality Control of cDNA during NEBNext® Single cell RNA sequencing library preparation

Following the sorting of cells into 384 well plates (as described in chapter 3), the five plates were sent to The Welcome Sanger Institute for library preparation and sequencing. Cells were sequenced on a NovaSeq 6000 at a depth of 600,000 reads per cell. To check the quality of cDNA prepared from each plate, 11 wells were randomly sampled from across the plate and analysed using a 2100 Agilent Bioanalyser. Furthermore, the concentration of cDNA per well was measured using an Omega FLUOstar plate reader. Figure 4.1 shows a representative example of the traces observed from a well with <1ng/µl cDNA (empty well; Figure 4.1A), and a well that contains >1ng/µl cDNA (well with cell; Figure 4.1B). The empty well shows a flat trace with a small peak at around 150 base pairs (bp), which is a result of primer dimers. The well that contains a cell shows a clear large peak closer to the larger bp marker (purple), which is indicative of good quality cDNA and is comparable to the example supplied in the NEBNext® protocol manual.

For the remaining wells, The Welcome Sanger Institute provided a matrix of concentrations of cDNA per well. Any well containing over 1ng/µl was deemed to be containing a good quality cell. Previous results suggested (Figure 3.7) that around half of the wells contained a cell. Indeed, ~40-50% of wells contained >1ng/µl cDNA (Figure 4.1C). Once the plates had been sequenced, fastq files were generated. I used the nf-core RNAseq pipeline (Ewels et al., 2020) to align the sequencing reads to a reference genome and generate a counts matrix to be analysed in the *R* programming language.

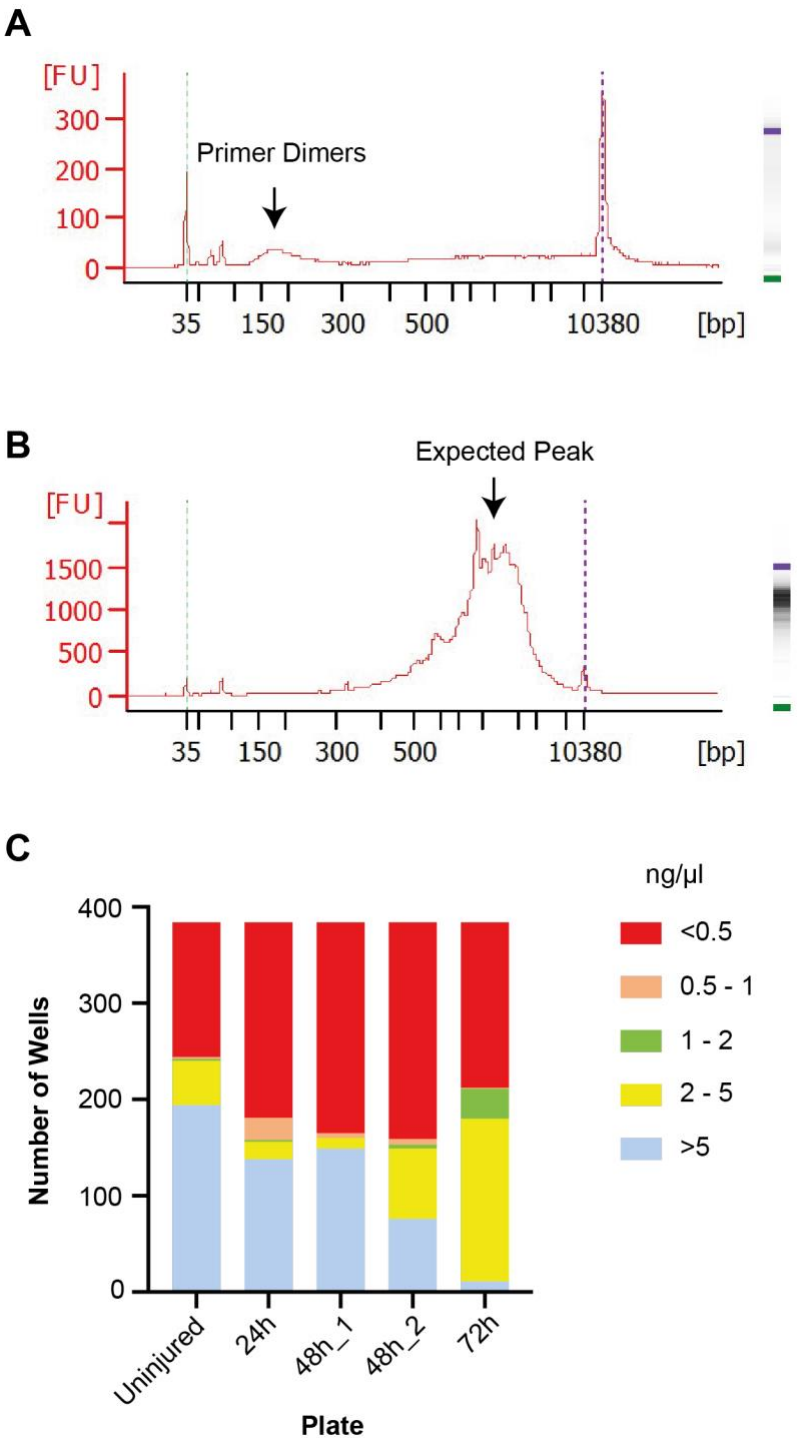


Figure 4.1 – Library Preparation Quality Control

A) Representative 2100 Agilent bioanalyser trace of a well with less than 1ng/µl of cDNA. Primer dimers are denoted on the graph.

B) Representative 2100 Agilent bioanalyser trace of a well with more than 1ng/µl. Expected peak denoted on the graph.

C) Cumulative bar chart showing the number of wells grouped by cDNA concentration at each timepoint post Phx or uninjured.

A & B) Line chart shows DNA Fragment size range (x-axis) and florescent units (FU – representative of signal intensity; y-axis). Gel image of cDNA sample shown to the side of the line graph. Purple line represents high size marker (10380bp) and green line represents low size marker (35bp).

4.2.2 Cell Quality Control

As discussed previously, low quality cells contain low numbers of total counts or genes. By plotting these metrics on a histogram, the profile of each dataset was visualised, which allowed for a threshold to be determined and the removal of poor-quality cells. Firstly, by looking at the number of genes (Total genes) detected in each cell, it was clear that there was a population of cells far below a 5000 gene threshold (Figure 4.2). While individual thresholds can be placed on each dataset separately, a 5000 gene threshold worked well for all 384 well plates in this study. Most of the cells/wells below the threshold likely represent those that contained no cell. Next, the total counts per cell was visualised (Figure 4.3). Rather than setting a threshold here, the histogram was coloured to show those cells that were filtered by the previous threshold and those that remained. It was clear that the cells that had low total genes also had low total counts. Based on this visualisation, it appeared no further count-based threshold was needed.

The final cell QC step was to compare the percentage of mitochondrial read counts of each cell. It has been suggested that poor quality/dying cells have disrupted cellular membranes and as a result, the cellular mRNA leaks out (Luecken and Theis, 2019). However, the mRNA in the mitochondria is protected, for a time, by the mitochondrial organelle membrane. Therefore, such cells would contain a higher number of mitochondrial mRNA read counts. I found that cells from all 384 well plates had a low mitochondrial percentage (Figure 4.4). A threshold of 10% was chosen, as this encapsulated the majority of cells, whilst removing the outliers that had up to 30% mitochondrial count

percentage. The number of cells that past QC can be found in Table 4.1. These individual datasets were merged before continuing with further QC.

Table 4.1 – Cell QC Statistics

Timepoint	Cells/Wells Passed QC (#)	Cells/Wells Removed (#)	% Cells/Wells Retained
Uninjured	235	149	61.2
24h Post PHx	152	232	39.6
48h Post PHx (Plate 1)	160	224	41.7
48h Post PHx (Plate 2)	149	235	38.8
72h Post PHx	204	180	53.1

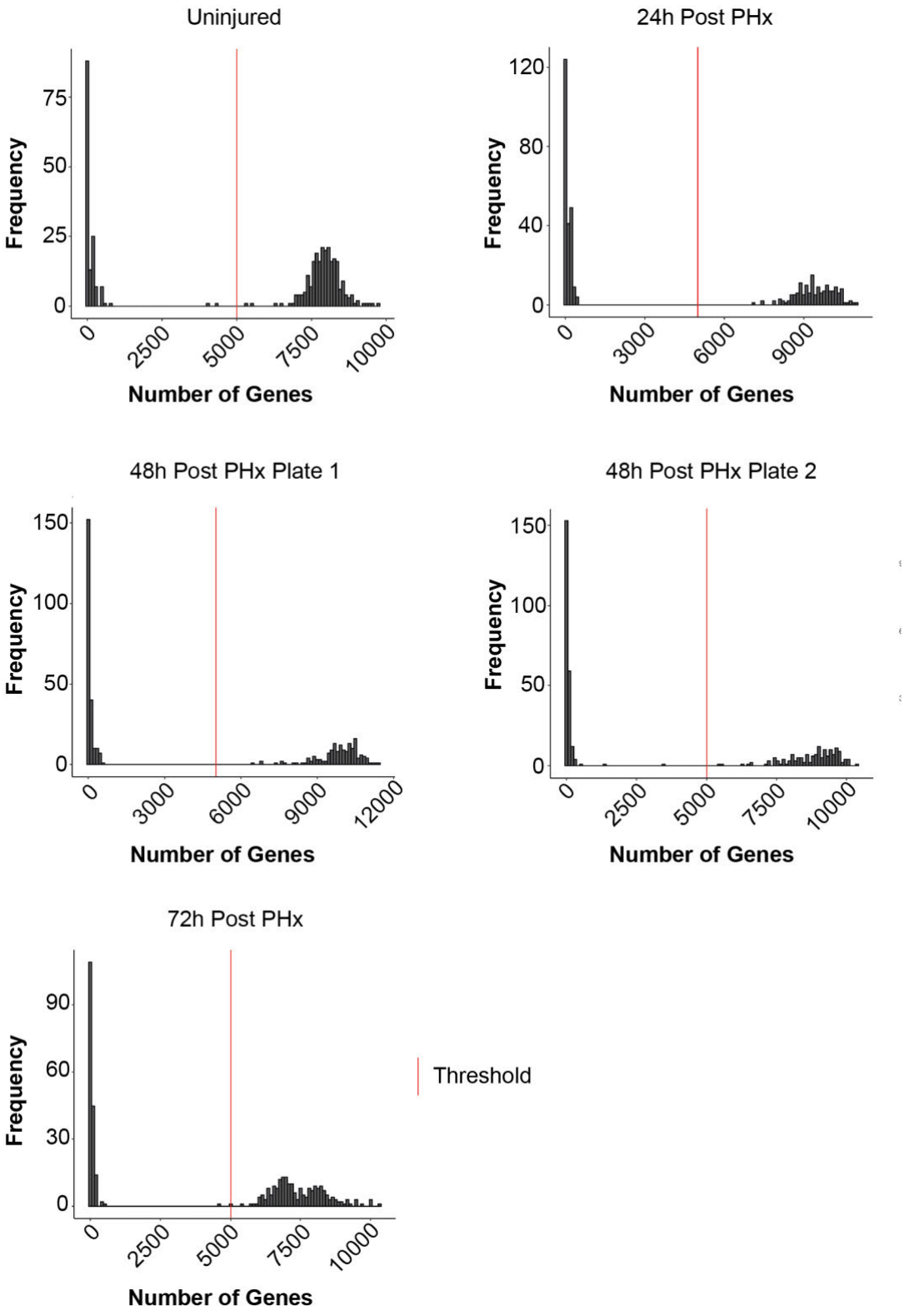


Figure 4.2 – Number of Genes Profiled per Dataset

Histogram representing the total number of genes detected across 384 wells per dataset (timepoint) following scRNA-seq. Each dataset represents a

timepoint post partial hepatectomy, with uninjured representing a timepoint prior to surgery. Red line denotes the threshold at which cells below this were removed.

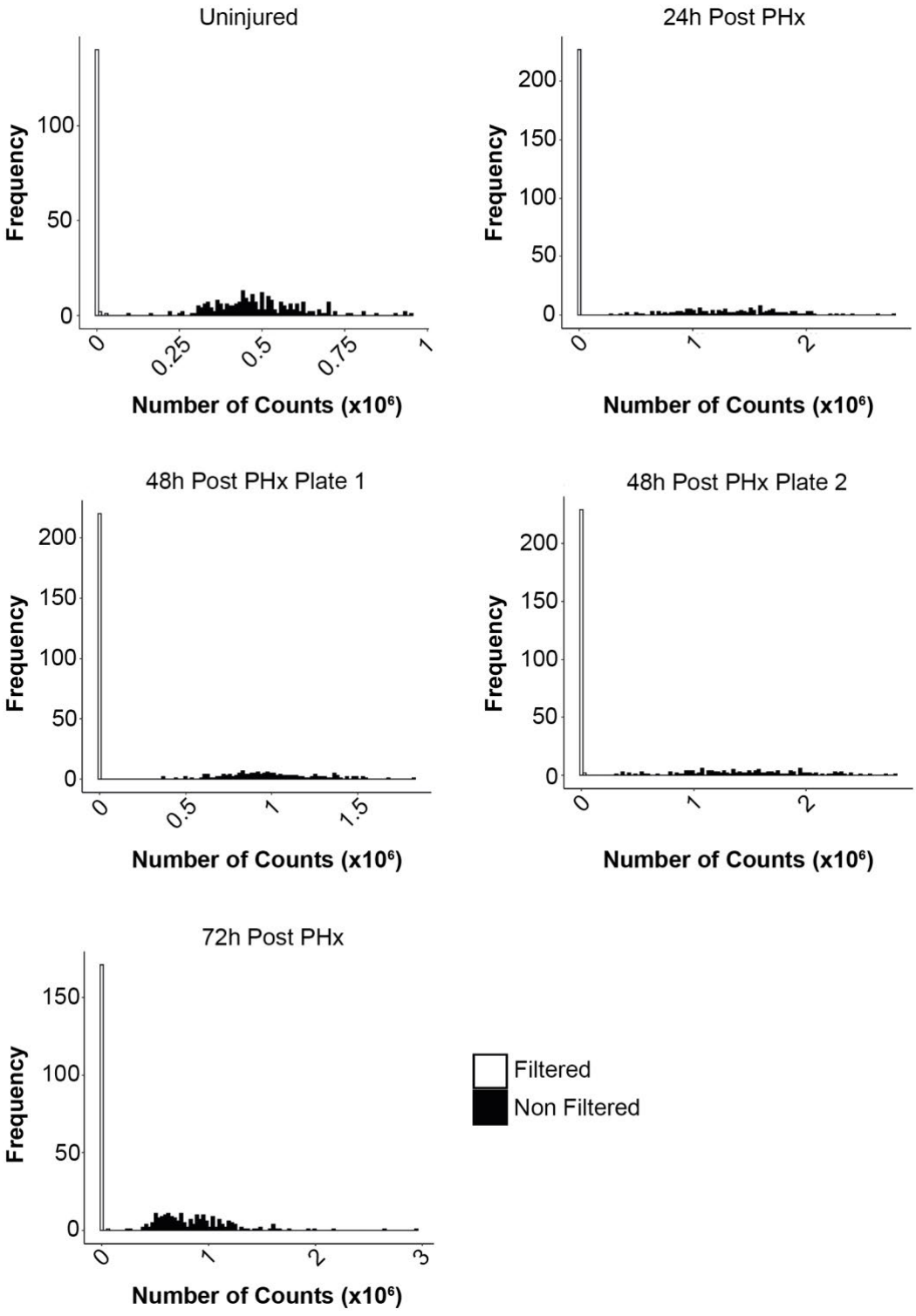

Uninjured
Frequency
Number of Counts (x10⁶)
24h Post PHx
48h Post PHx Plate 1
48h Post PHx Plate 2
72h Post PHx
Filtered
Non Filtered

Figure 4.3 – Number of Counts Profiled per Dataset

Histogram representing the total number of read counts across 384 wells per dataset (timepoint) following scRNA-seq. Each dataset represents a timepoint post partial hepatectomy, with uninjured representing a timepoint prior to surgery. Bars are coloured white for those cells removed by a 5000 threshold on number of genes expressed. Black represents those cells that were kept under the same threshold

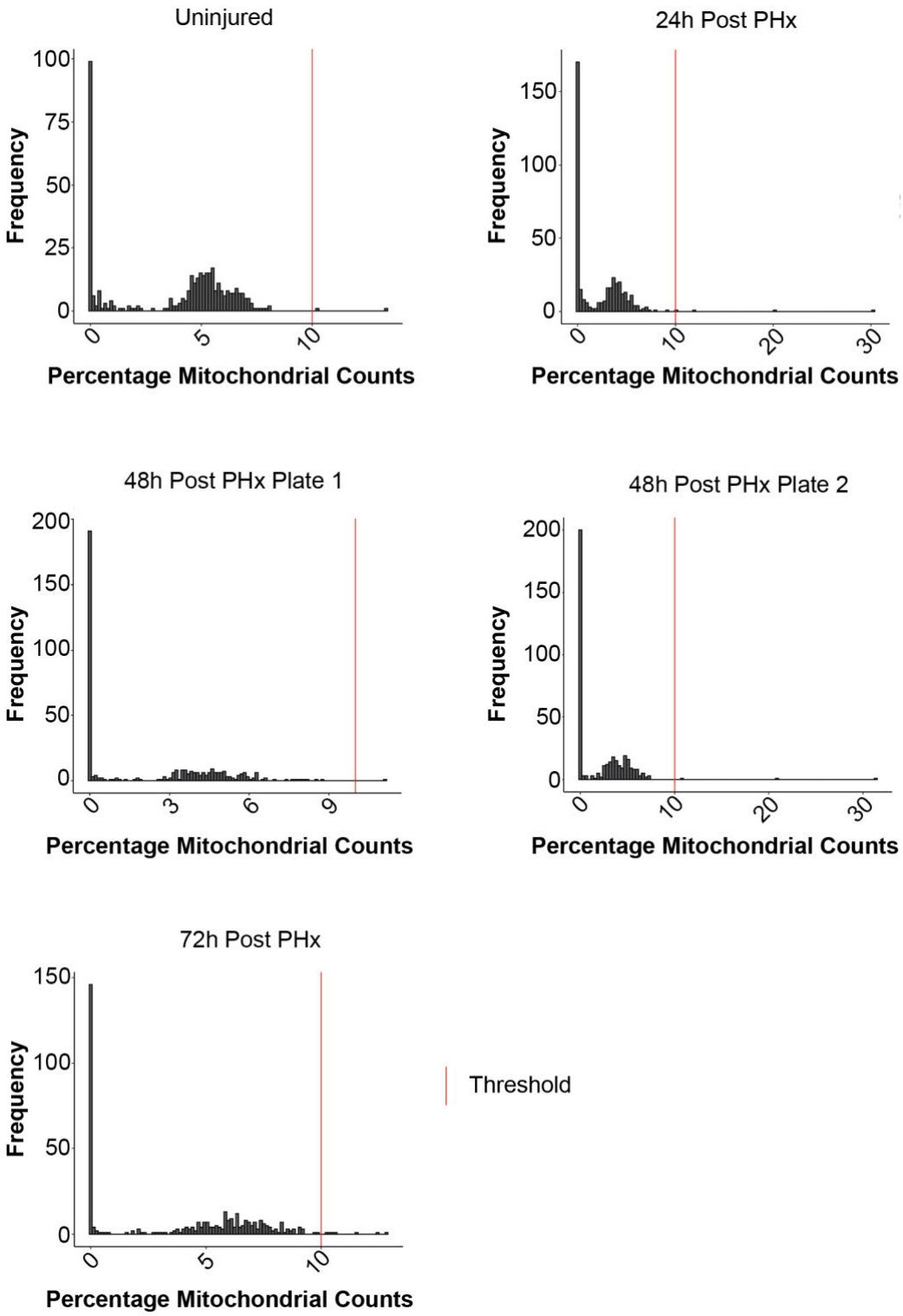


Figure 4.4 – Percentage Mitochondrial Count Profiled per Dataset

Histogram representing the percentage of counts assigned to mitochondrial genes across 384 wells per dataset (timepoint) following scRNA-seq. Each dataset represents a timepoint post partial hepatectomy, with uninjured

representing a timepoint prior to surgery. Red line denotes the threshold at which cells above this were removed.

4.2.3 Gene Quality Control

Removing genes that have a near zero count across all datasets improves statistical inferences downstream (Lun et al., 2016). Furthermore, removal of these uninformative genes saves on computing power and increases the efficiency of the pipeline. However, deciding on a threshold and determining which genes are uninformative can be challenging. As discussed previously, there are two common methods, which I decided to compare. The first is an average expression-based threshold (Lun et al., 2016). An average count of one, or $\log_{10}$ count of zero, is commonly chosen as it usually bisects a histogram of average counts in the flat level portion of the graph (Figure 4.5). This splits the graph into the informative higher expressed genes on the right and the lower potentially uninformative genes on the left. The second detection-based thresholding does not use a standard threshold. The *Seurat* pipeline suggests a default of at least one count per gene in at least three cells (Stuart et al., 2019). The profile of genes removed by such a threshold can be seen in Figure 4.5A. It is clear from this that only a small number of these lowly expressed genes would be removed. This threshold may work well for UMI based data, where the overall counts are lower compared to non-UMI based protocols. Simply increasing the detection per cell to five while keeping the number of cells to express the gene at three, removes a larger number of genes (Figure 4.5B) and accounts for a larger expected count per gene. It has been suggested that downstream analysis should be considered when determining/setting an appropriate threshold using this method (Luecken and Theis, 2019). Keeping the detection at five but increasing the number of cells needing to express the gene to ten results in all genes that are not expressed in at least ten cells to be removed (Figure 4.5C). As mentioned previously, this consequently hinders the identification of clusters of ten cells or less, as the genes used to identify these cells have been removed. As a result, rare

populations could potentially be lost by using this higher threshold. It is important to point out that the average based cut-off of one removes an even greater number of genes than this higher detection-based threshold (Figure 4.5C). Based on this data, I decided that a threshold of five counts in at least three cells would be appropriate. To ensure that this did not remove a large number of read counts from a particular cell or population, a histogram of read counts of genes to be removed was plotted (Figure 4.5D).

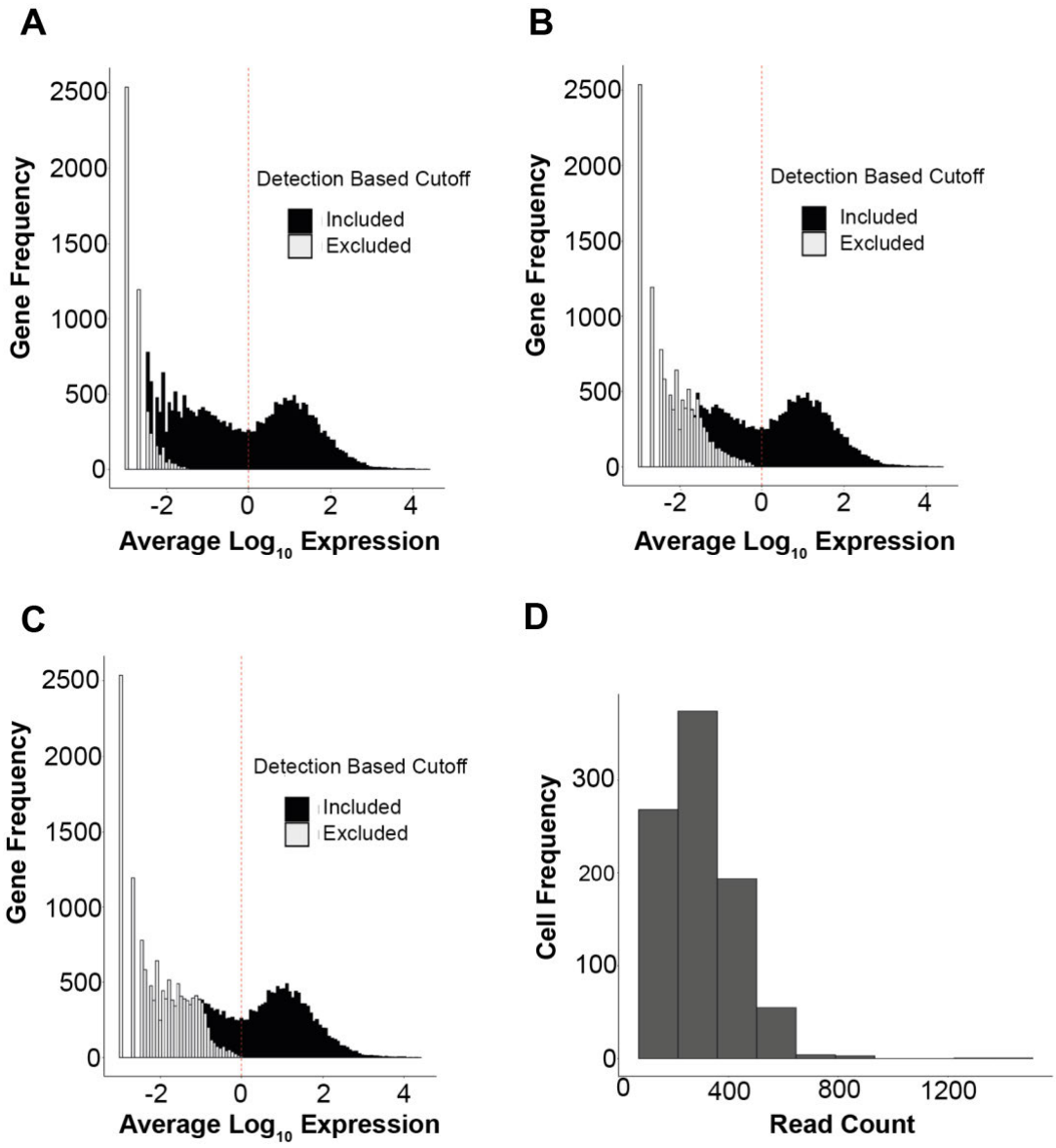


Figure 4.5 – Gene Quality Control

(A,B,C) Histogram of average log_{10} gene expression for single hepatocytes sequenced from uninjured mice, and 24h, 48h and 72h post partial hepatectomy. Average log_{10} expression was calculated from a total of 900 cells that was split into the following datasets; Uninjured – 235 Cells, 24h Post PHx – 152 Cells, 48h Post PHx Plate 1 – 160 Cells, 48h Post PHx Plate 2 – 149 Cells, 72h Post PHx – 204 Cells. Genes that have zero counts in all cells were not used to generate the histogram. Red dashed line represents a threshold of average log_{10} expression 0. Black colour bars and white colour bars

represent genes included or excluded by the following thresholds, respectively. A) Denotes threshold of at least one gene count detected in at least three cells. B) Denotes threshold of at least five gene counts detected in at least three cells. C) Denotes threshold of at least five gene counts detected in at least ten cells.
D) Histogram of read counts removed per cell after a threshold of at least five gene counts in at least three cells was used.

4.2.4 Removal of Contaminating Cells

After gene and cell QC was performed, I proceeded to identify any potential contaminating cells prior to normalisation. In order to do so, I first processed the data through a very rough *Seurat* pipeline of global-scaling normalisation, followed by principle component analysis (PCA) on scaled and centred normalisation data. I then performed uniform manifold approximation and projection (UMAP) dimensionality reduction. The expression of several key marker genes was visualised on the UMAP projection. The data was interrogated for marker genes of several cell types (Table 4.2). No expression of any cholangiocytes, mesenchymal, or leucocyte markers was found (Figure 4.6). Low expression levels of Icam2 and Pecam1 suggested that there was some endothelial contamination (Figure 4.6). These cells also expressed hepatocyte markers, suggesting that they were doublets, or that fragments of endothelia were captured in the wells of these cells. A strict threshold was implemented in which any cell that expressed at least one count of either Pecam1 or Icam2 was removed. This would prevent any aberrant endothelial specific genes appearing in downstream differential analysis. Fifty-three cells were removed during this process. The rest of the cells were verified as hepatocytes through the expression of the hepatocyte markers shown (Figure 4.6). The expression of Mki67 confirmed that the datasets had also captured replicating cells, as intended.

Table 4.2 – Cell Type Marker Genes

Cell Type	**Marker Genes**
Hepatocytes	Hnf4α, Transferrin (Tf), Transthyretin (Ttr), Albumin (Alb)
Cholangiocytes	Epithelial cellular adhesion molecule (Epcam), Keratin 19 (Krt19)
Endothelia	Intercellular adhesion Molecule 2 (Icam2), Platelet endothelial cell adhesion molecule 1 (Pecam1)
Mesenchymal	Platelet derived growth factor receptor beta (Pdgfrβ)
Leucocytes	Protein tyrosine phosphatase receptor type c (Ptprc)
Cell Cycle	Marker of proliferation Ki-67 (Mki67)

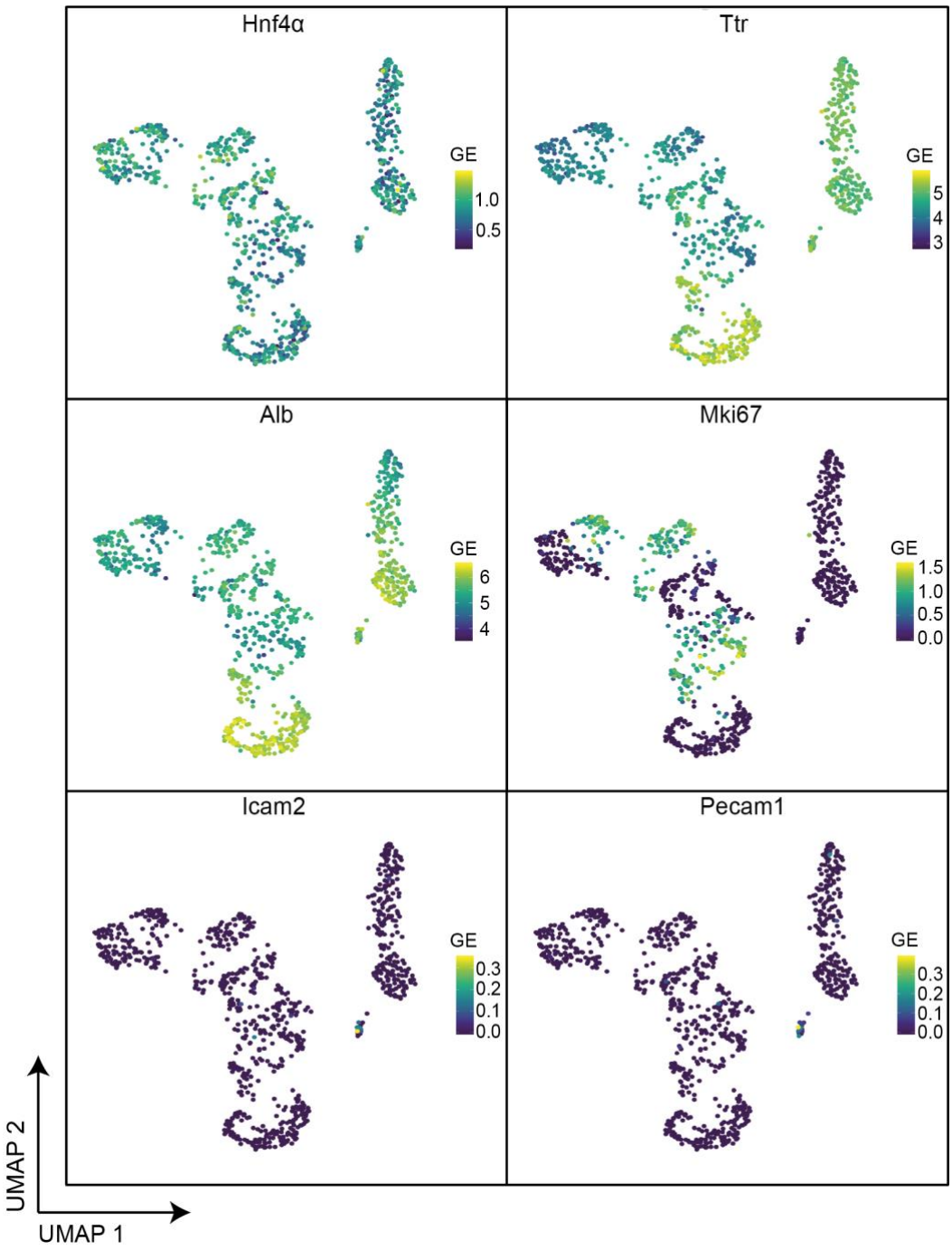


Figure 4.6 – Expression of Cell Type Specific Markers

Uniform manifold approximation and projection (UMAP) of hepatocytes scRNA-seq data following a partial hepatocyte time course (Uninjured, 24h, 48h, 72h). Each dot represents one cell (n=900). UMAPs are coloured by the log normalised expression of the titled gene. GE = Gene expression.

4.2.5 Classification of Fucci Expression State

The normalisation testing framework *scone*, discussed in section 4.1, uses *a priori* biological knowledge to help determine the best normalisation strategy. All hepatocytes were index sorted prior to sequencing and therefore had associated fluorescence data for the mVenus and mCherry fluorophores. During the cell cycle, large changes in the transcriptome occur and as a result, one would generally expect to see differences in gene expression between mVenus and mCherry-expressing cells (Giotti et al., 2017; Liu et al., 2017). Using the fluorescence data, hepatocytes were categorised as mVenus$^+$ only (mVenus$^+$), mCherry$^+$ only (mCherry$^+$) or mCherry$^+$/mVenus$^+$ (Dual$^+$) using the gating shown in Figure 4.7. Based on these classifications , 576 hepatocytes were classified as mCherry, 69 as dual, and 202 as mVenus across all timepoints. Assignment to one of these categories was used to inform the normalisation process downstream.

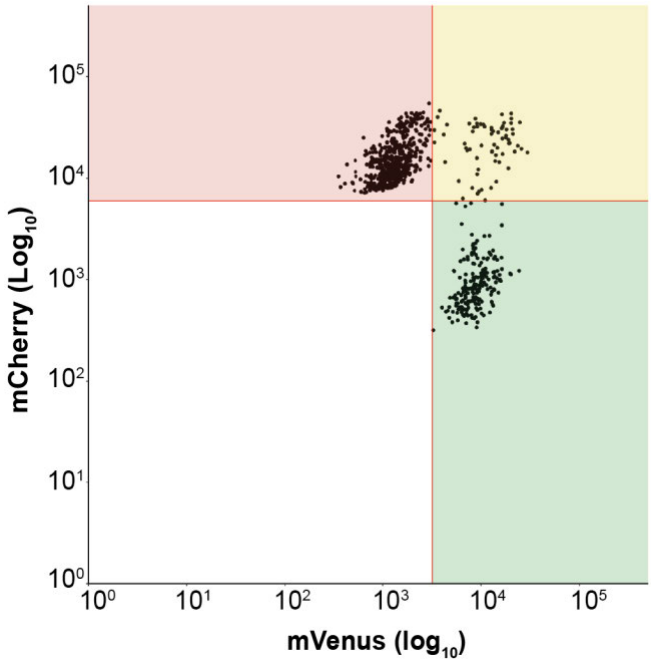


Figure 4.7 – Classification of Hepatocyte FUCCI expression status

Scatter plot showing the mCherry and mVenus expression in hepatocytes that were sorted for scRNA-seq and passed quality control (n=847). Green area shows hepatocytes classified as mVenus$^+$ only, red area shows hepatocytes classified as mCherry$^+$ only and the yellow area represents mCherry$^+$ mVenus$^+$ (Dual$^+$) classified hepatocytes.

4.2.6 Feature Bias and Confounding

4.2.6.1 Gene-wise variability

All steps during sample preparation for scRNA-seq are conducted in a way that minimises technical variability. However, it is not possible to completely abrogate this and as a result, it is critical to interrogate what explanatory variables may confound or introduce technical variance within the data. For example, differences in the totals counts per cell originate from the stochastic random sampling of the cDNA, leading to variation in sequence depth. This does not provide any useful information with regards to differences between cell populations. Other factors were also analysed, such as mitochondrial percentage, total genes, batch (i.e., Uninjured, 24h, 48h plate 1, 48h plate 2, 72h) and biological state (i.e., mCherry, mVenus, or Dual). Ideally, the total counts, mitochondrial percentage and batch would explain small amounts or no gene-wise variance, whereas the biological state would account for the most. The total genes detected is a slightly more problematic feature to consider. On the one hand, the literal number of genes detected may be of little interest, as this may arise from the sampling process. However, the number of genes expressed could also be linked to the underlying biological state. For example, replicating cells will activate a myriad of cell cycle related genes when DNA replication begins (Liu et al., 2017). Therefore, complete removal of this feature may not be desirable as this may also eliminate interesting biological information.

Linear regression was used to model each gene against Total counts, total genes, percentage mitochondrial reads, batch or biological state. The R-squared value for each model was assessed to evaluate the variance explained by each feature. This was then visualised on a density plot (Figure 4.8A). It is clear from this plot that the batch, total genes and total counts were the strongest drivers of variance in the data, with the distribution of these factors showing peaks above 10%. The mitochondrial percentage had less control over the data. For the data to be driven by the biology, we would expect

the distribution of R-squared values for all these QC/explanatory factors to be below that of the biological state.

4.2.6.2 Cell-wise variability

The same factors can also be assessed from a cell-wise perspective. PCA analysis was first performed on logged unnormalized scaled data. The first ten PCs were used in a linear regression model, along with the explanatory variables used above. Looking at the R-squared values, again the total counts explained the most variance within the first PC, showing a high R-squared value (Figure 4.8B). The batch also accounted for a considerable amount of variance across many of the first ten PCs. When this was visualised on a PCA plot, it became clear that the total counts was driving the shape of the data (Figure 4.8C). This factor was therefore the prime candidate to be controlled for/removed by normalisation. Batch is also a potentially important factor to remove, although this is usually achieved in a separate batch correction step and will be discussed below (Section 4.2.10).

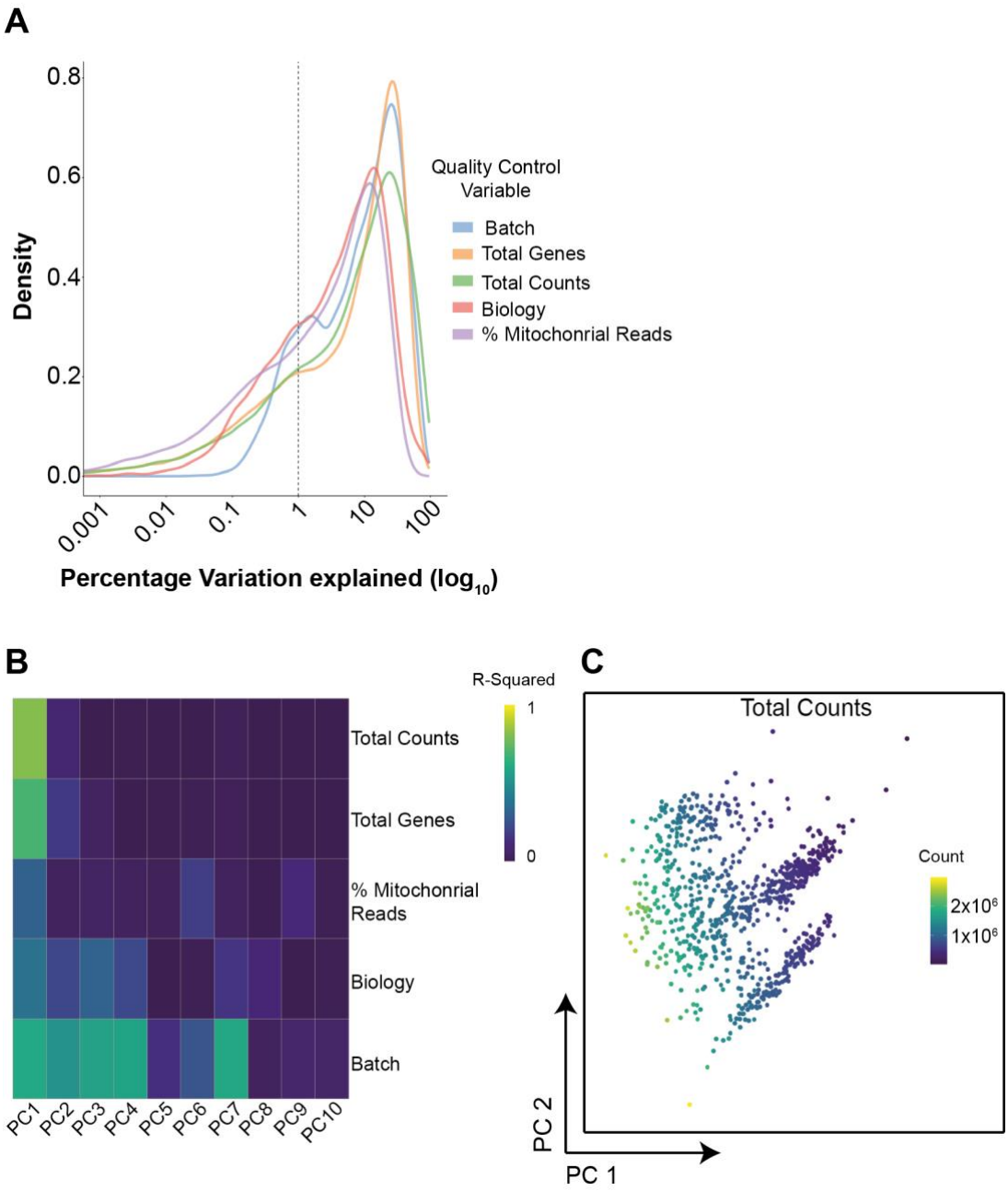


Figure 4.8 – Confounding Quality Control Factors

A) Density plot showing gene wise R-squared value for key quality control variables in scRNA-seq data from mouse hepatocytes following partial hepatectomy. Five key factors are shown, Batch (i.e., Uninjured, 24h, 48h plate 1, 48h plate 2, 72h), total number of expressed genes (Total Genes), Total number of read counts (Total counts), biological state (Biology; i.e., $mCherry^+$ only, $mVenus^+$ only, or $Dual^+$) and percentage mitochondrial counts (% Mitochondrial reads). The further the distribution lies to the right, the greater the contribution of variables

towards the gene expression variability. A dashed line represents a percentage variability of 1%. Distributions below this threshold are considered acceptable.

B) Heatmap showing the cell wise (linear regression on principle components) R-squared value for key quality control variables in scRNA-seq data from mouse hepatocytes following partial hepatectomy. The same five factors are represented as in (A).

C) Uniform manifold approximation and projection of hepatocytes scRNA-seq data following a partial hepatocyte time course. Cells are coloured by the number of read counts they possess. PC 1 shows a clear correlation with the number of read counts.

4.2.7 Normalisation

Considering total counts accounted for the most variability in the data, library size normalisation was required to remove this effect. There are many different ways to normalise scRNA-seq data. Fortunately, a framework called *scone* has been developed to test various normalisation strategies and compare their effectiveness using performance metrics (defined in Table 4.3). Desirable scores for these metrics can be summarised as follows;

Higher values - BIO_SIL, PAM_SIL

Lower values - BATCH_SIL, EXP_QC_COR, RLE_IQR, RLE_MED

Table 4.3 – Scone Metrics

Metric Name	Definition
BIO_SIL	Average silhouette width of cells grouped by biological component. High scores show that a normalisation has retained similarity of the following biological groups; mCherry, mVenus and Dual.
BATCH_SIL	Average silhouette width of cells grouped by batch component (i.e., Uninjured, 24h, 48h plate 1, 48h plate 2, 72h). Low scores are desired and show batch variation is a smaller driver of cell similarities.
PAM_SIL	Maximum average silhouette score of 2-10 clusters using partitioning around medoids (PAM). High scores represent normalisation strategies that preserve the heterogeneity within the data.
EXP_QC_COR	R-squared measure for the regression of first three log-count PCs on QC PCs (number of read counts and percentage mitochondrial counts). Low score shows the normalisation strategy has reduced the explanatory power of the QC variables.
RLE_IQR	Variance of inter-quantile range of the relative log expression. Low scores represent a reduction in the global differential variability.
RLE_MED	Mean of squared-median relative log expression. Low scores represent a reduction in the global differential expression.

I tested several normalisation strategies within the *scone* framework to assess which best fits this data. These included *scran*, *SCnorm*, *TMM* (a bulk based normalisation strategy), *Sum* (counts corrected by the summed library size of each cell, similar to CPM), and *Batchelor*. No normalisation was also used as a comparison. *Batchelor* normalisation is an extension of the *scran* method, in which multiple batches are present. Rather than using the entire combined data to generate the size factors, size factors for each individual batch are first generated and then downscaled to match the coverage of the shallowest batch. The metrics described above were generated for each normalisation strategy and visualised by PCA (Figure 4.9A). I found *Batchelor* to be the superior normalisation strategy. This was able to balance effective normalisation, showing good RLE_IQR and RLE_MED scores, whilst retaining

biological variation and heterogeneity with good scores for BIO_SIL and PAM_SIL.

The effect normalisation has on the data can be easily visualised using a relative log expression (RLE) plot. Good library size normalisation will produce a median RLE close to zero for most cells. Indeed, looking at the RLE plot for the various normalisation strategies tested, *Batchelor* showed all cells had a median RLE close to/equal to zero and the overall variability of RLE was lower than the other normalisations (Figure 4.9B). The other normalisation strategies showed larger RLE variation or cells that did not have an RLE centred on zero. This was most prominent when looking at the values for no normalisation and demonstrates the need to normalise the data. I therefore used *Batchelor* to normalise these datasets.

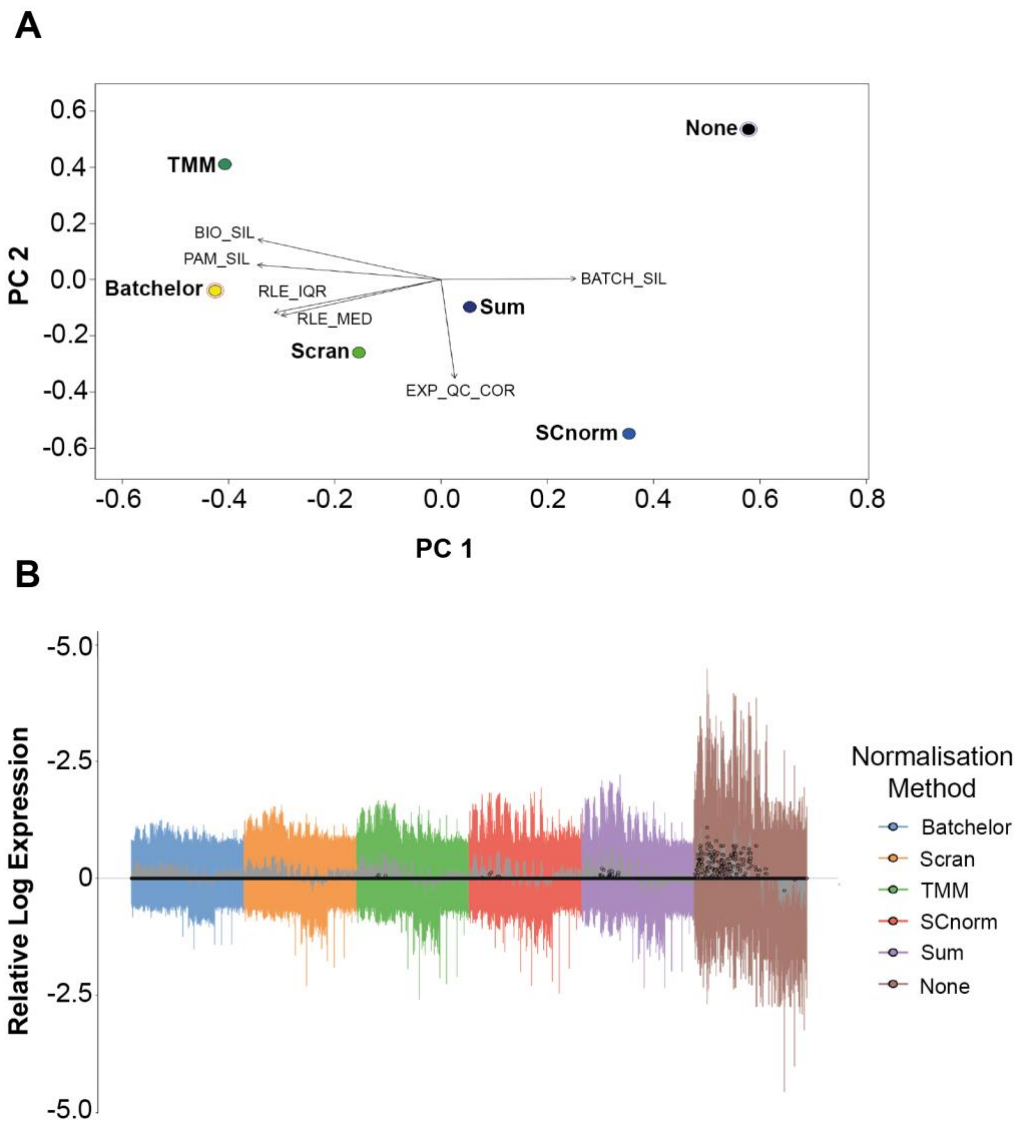


Figure 4.9 – Normalisation Comparisons

A) Biplot showing the first two principle components of the six ranked scone performance metrics; BIO_SIL (Measure of preservation of biological difference), BATCH_SIL (Measure of removal of batch structure), PAM_SIL (Measure of heterogeneity preservation), EXP_QC_COR (Measure of QC artefact removal), RLE_IQR (Measure of the reduction in global differential variability) RLE_MED (Measure of the reduction in global differential expression). Each dot represents a normalisation approach that are coloured blue (low) to yellow (high) based on rank score. The least effective normalisation is further circled in blue and the best circled in red. Length and direction of black arrows

are a measure of how each metric contributes to the principle components. Normalisation approaches shown include; *Batchelor*, *Scran*, *SCnorm*, *TMM*, *Sum*, *None* (No normalisation).

B) Tufte-style boxplot of relative log expression for each of the normalisation approaches. Each coloured region represents all the cells from the combined scRNA-seq datasets of hepatocytes across a partial hepatectomy time course. The median is represented by a circle, grey lines represent inter quantile range, and coloured lines represent the whiskers (1.5 times the inter quantile range).

4.2.8 Highly Variable Genes

Once normalised, highly variable genes were selected for use in downstream analyses. The purpose of picking highly variable genes is to remove genes that add random noise within the data and are “uninteresting” with respect to understanding differences between the cells. However, the genes that are chosen need to preserve the interesting biological structure. As with gene filtering, selection of a smaller subset of genes also improves computation efficiency.

To choose a list of genes, I used the *modelGeneVar* function from *scater*. This functions first computed the variance of the log-normalised counts for each gene across all the cells. The log-normal counts, as opposed to raw counts or unlogged values, were used to remove the mean-variance relationship seen within the scRNA-seq data. This is where the variance of the gene is most affected by the abundance of the gene than the underlying biology. However, the log transformation does not fully stabilise the variance and some relationships still persist. The *modelGeneVar* function therefore fitted a trend to the variances with respect to mean log-expression, which was then used to identify technical (uninteresting) variance. This was then subtracted from the total variance to yield the biological component. As my dataset contained several batches, this process was repeated across each batch separately as opposed to calculating this trend on all the data together.

This was to prevent any specific batch dominating the selection. Once the technical and biological component had been calculated for each gene in each dataset, these were averaged across each batch. I decided to use the top 2000 genes with the most biological variation for downstream analysis. Highlighting these genes on a mean variance graph showed that selecting the first 2000 covers many of the genes above the trend for all datasets; these genes still capture the interesting biological information (Figure 4.10). Choosing a higher number could potentially introduce genes that represent uninteresting and noisy variation.

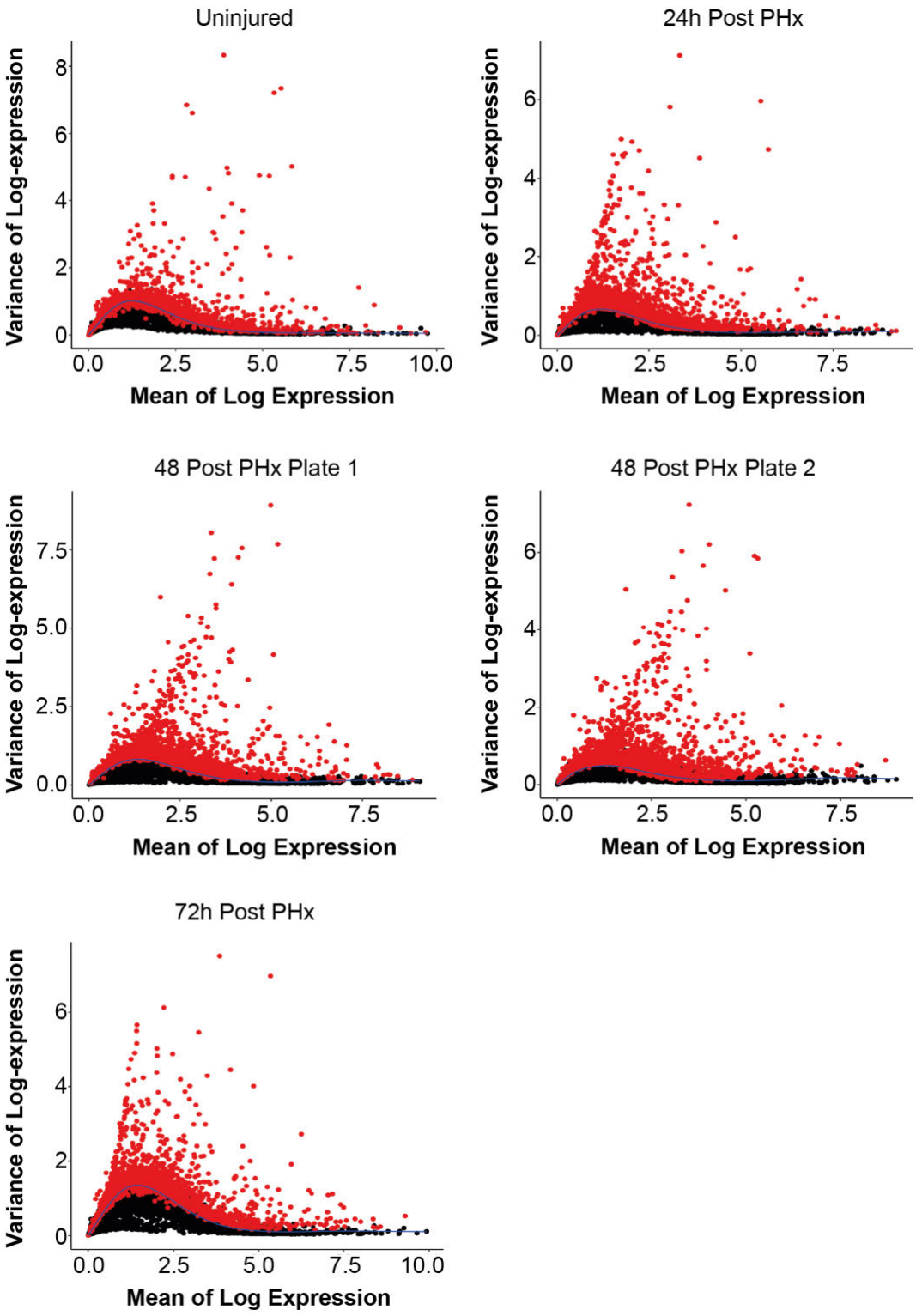


Figure 4.10 – Highly Variable Genes per Dataset

Dot plot represents the variation of log-expression with respect to mean log expression of genes across 384 hepatocyte cells per dataset (timepoint)

following scRNA-seq. Each dataset represents a timepoint post partial hepatectomy, with uninjured representing a timepoint prior to surgery. Blue line represents the trend fit produced by *modelGeneVar*. Red dots represent the chosen 2000 highly variable genes.

4.2.9 Principle Component Analysis

Following highly variable gene selection, the log-normal counts were scaled and centred prior to PCA analysis. PCA analysis was the first dimensionality reduction step in the workflow and was conducted using the highly variable genes chosen above. Again, at this point in the analysis, a subset of the PCs were chosen based on the variation they capture. This was to eliminate those that were of little interest and had the potential to add technical noise to the analysis. One way to decide the number of PCs is to use a heuristic approach, in which the variation contribution of each PC is plotted on an Elbow plot (Figure 4.11A). An appropriate threshold was considered to be the point at which the graph reaches the “joint” of the elbow, at which the addition of more PCs does not contribute a large amount of variation and may start to add uninteresting noise. My data revealed that between 10 to 20 PCs could have been chosen.

Another method is to visualise the expression of the genes with the most positive and negative PC scores of each PC (Appendix 1). This showed the biological components captured, as well as the degree of separation of each PC. Based on all the information, I decided that 15 PCs was the ideal number, as there was a small dip in variation captured when using more. This was also mirrored in a lower degree of separation of gene expression in PCs above 15.

4.2.10 Batch Correction

Using these 15 PCs, a UMAP visualisation was constructed. This is a non-linear dimensionality reduction method that allows the visualisation of

cells in 2D space. Presenting the data in this way highlighted a large batch component, in which replicating cells clustered together, as did the non-replicating cells, but each separated by the individual plate from where they originated (Figure 4.12A & B). It is possible that the difference was biologically driven, however, the separation of the two 48h groupings of replicating hepatocytes would suggest that there was a significant batch effect (Figure 4.12A). Again, linear regression was used to assess the degree of cell-wise variance that is explained by the QC/explanatory factors. The R-squared value from the linear regression for batch was high throughout many of the first 10 PCs. The dependency on total counts and mitochondrial percentage had, reassuringly, decreased, showing effective normalisation (Figure 4.13A).

To circumvent the batch component, a batch correction was performed on the data. Two methods were chosen and compared to find the most optimal approach. The first was a method available in the *Seurat* pipeline and will be referred to from this point onwards as *Seurat integration*. *Seurat integration* changes the log-normal counts to correct for the batch component. Following this correction, these corrected counts are re-scaled and re-analysed using PCA and UMAP visualisation. Using the same approach as described previously, 15 PCs was found to be the most suitable for the corrected data (Figure 4.11B, Appendix 2). The UMAP clearly showed overlap of the different batches, and while the cells were somewhat separated by the biological component, many of the non-replicating uninjured hepatocytes sat closer to the injured replicating hepatocytes, which was not expected (Figure 4.12C & D).

A second approach called *Harmony* displayed a more optimal result. This method corrects the original PCs generated to eliminate the batch component. These new harmony components (HCs) were used to create the UMAP representation. The same heuristic elbow plot can still be used to assess the optimal number of HCs (Figure 4.11C, Appendix 3). Fifteen HCs were chosen, as this, once again, appeared to the most optimal number. Much like the *Seurat integration*, hepatocytes from the various datasets overlapped. However, *Harmony* kept a greater separation of the non-replicating and

replicating hepatocytes (Figure 4.12E & F). This can be seen by a larger R-squared value for the biological factor when using *Harmony* compared to *Seurat Integration* (Figure 4.13B & C). While the total genes showed a slightly higher R-squared value for *Harmony* than with *Seurat integration*, the value was reduced when compared to that which was obtained prior to normalisation. Furthermore, some or all of the remaining explanatory power in total genes may originate from biological variation. Considering the large number of cell cycle associated genes, it is likely that this variation is cell cycle related (Giotti et al., 2017). Based on these visualisations and analyses, I decided that *Harmony* was the best approach to employ for batch correction in order to inform downstream analyses.

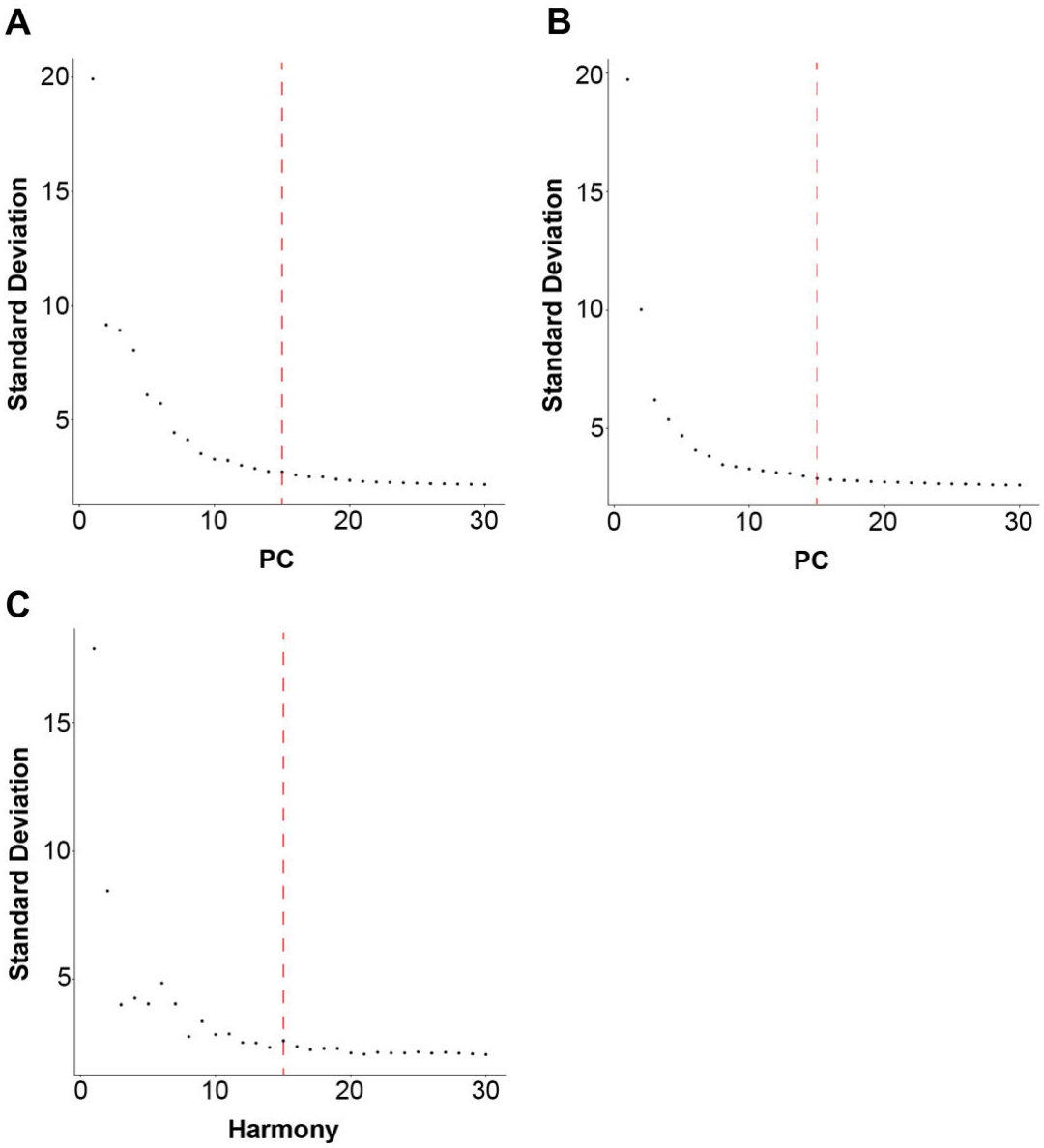


Figure 4.11 – Comparison of Batch Correction – Elbow Plot

Elbow plot showing the standard deviations of each of the first 30 principle components (A & B) or harmony components (C). All data shown above was generated from scRNA-seq of hepatocyte cells following a partial hepatectomy time course (Uninjured, 24h, 48h plate 1, 48h plate 2, 72h). Data was normalised and either processed with no batch correction (A), corrected using the *Seurat integration* method (B) or corrected using the *Harmony* method (C) prior to this analysis. Red dashed line denotes the number of principle or harmony components chosen for each method.

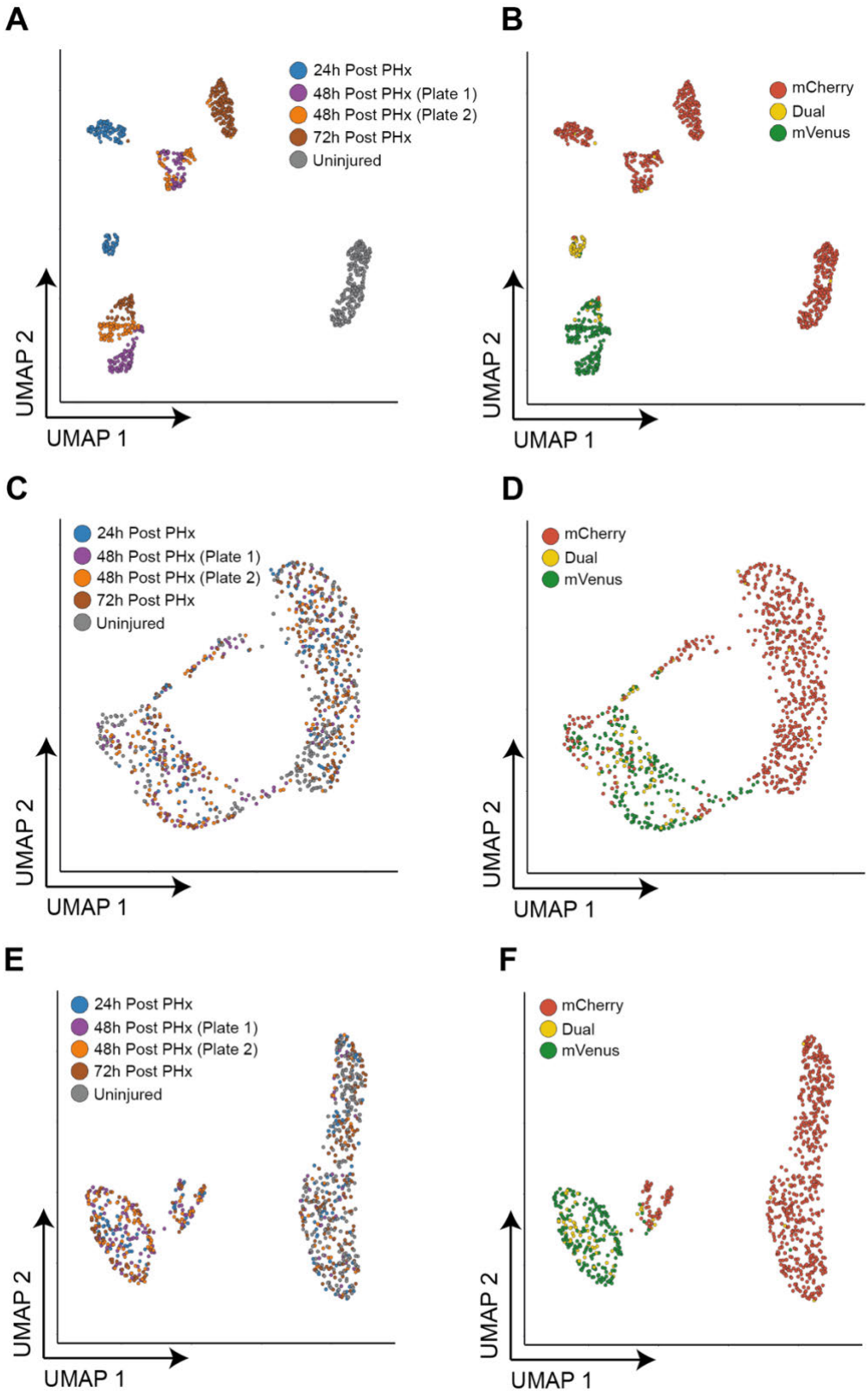

A
24h Post PHx
48h Post PHx (Plate 1)
48h Post PHx (Plate 2)
72h Post PHx
Uninjured
UMAP 2
UMAP 1
B
mCherry
Dual
mVenus
UMAP 2
UMAP 1
C
24h Post PHx
48h Post PHx (Plate 1)
48h Post PHx (Plate 2)
72h Post PHx
Uninjured
UMAP 2
UMAP 1
D
mCherry
Dual
mVenus
UMAP 2
UMAP 1
E
24h Post PHx
48h Post PHx (Plate 1)
48h Post PHx (Plate 2)
72h Post PHx
Uninjured
UMAP 2
UMAP 1
F
mCherry
Dual
mVenus
UMAP 2
UMAP 1

Figure 4.12 – Comparison of Batch Correction – UMAP Visualisations

Uniform manifold approximation and projection, where each dot represents one hepatocyte. Colour denotes the batch (A, C & E) or biological classification (B, D & E) of each cell. All data shown above were generated from scRNA-seq of hepatocytes following a partial hepatectomy time course (Uninjured, 24h, 48h plate 1, 48h plate 2, 72h). Data were normalised and either processed with no batch correction (A & B), corrected using the *Seurat integration* method (C & D) or corrected using the *Harmony* method (E & F) prior to this analysis.

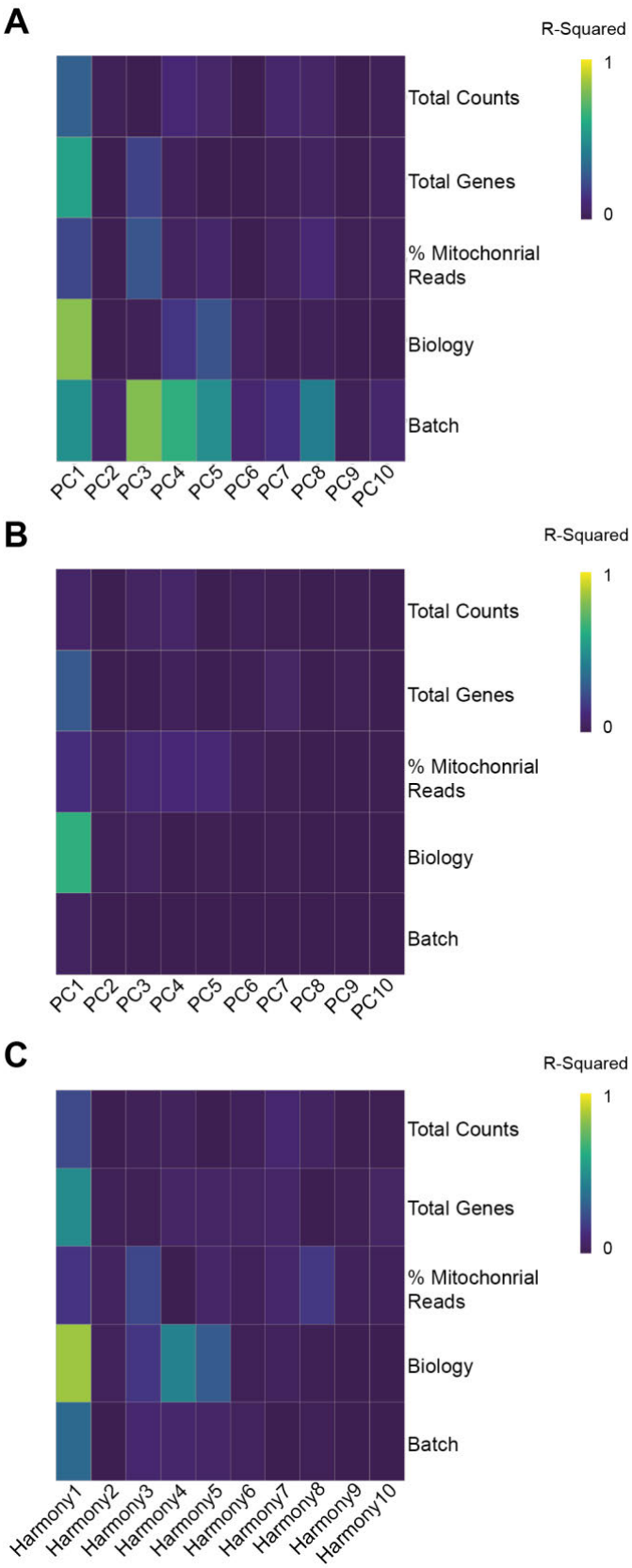

A
R-Squared
1
0
Total Counts
Total Genes
% Mitochonrial Reads
Biology
Batch
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10
B
R-Squared
1
0
Total Counts
Total Genes
% Mitochonrial Reads
Biology
Batch
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10
C
R-Squared
1
0
Total Counts
Total Genes
% Mitochonrial Reads
Biology
Batch
Harmony1
Harmony2
Harmony3
Harmony4
Harmony5
Harmony6
Harmony7
Harmony8
Harmony9
Harmony10

Figure 4.13 – Comparison of Batch Correction – Analysis of Explanatory Variables

Heatmap showing the R-squared value of linear regression models fitted individually to the first 10 principle components (A & B) or first 10 harmony components (C) against key five explanatory variables. The five explanatory variables are; Batch (i.e., Uninjured, 24h, 48h plate 1, 48h plate 2, 72h), total number of expressed genes (Total Genes), Total number of read counts (Total counts), biological state (Biology; i.e., $mCherry^+$ only, $mVenus^+$ only, or $Dual^+$) and percentage mitochondrial counts (% Mitochondrial reads). Data was normalised and either processed with no batch correction (A), corrected using the *Seurat integration* method (B) or corrected using the *Harmony* method (C) prior to this analysis.

4.3 Discussion

The aim of this chapter was to develop a bespoke scRNA-seq pipeline to analyse NEBNext® scRNA-seq data from hepatocytes. I have shown within this chapter that I was able to determine optimal cell based and gene-based thresholds to isolate high-quality cells, followed by effective removal of contaminating cell populations. Subsequent steps aimed to remove technical variations, whilst preserving biological variation. Explanatory QC variables were used to assess how much technical variation remained in the data at each stage. Classification of hepatocytes as either $mVenus^+$, $mCherry^+$ or $Dual^+$ was used to ensure that the biological variation was maintained within the data. Several normalisation and batch correction strategies were compared to ascertain the most appropriate. Determining these optimal approaches was key to generating highly accurate downstream results.

The NEBNext® scRNA-seq workflow produced high quality scRNA-seq data from live mouse hepatocytes following a PHx time course, sorted by cell cycle phase. The QC thresholds described in this chapter are important to remove data generated from empty wells or cells/wells containing degraded RNA, which is often observed in scRNA-seq experiment (Luecken and Theis,

2019; McCarthy et al., 2017; Stuart et al., 2019). However, the point at which these thresholds are placed is very subjective and dataset dependent. For example, previous studies on hepatocyte scRNA-seq data use higher thresholds for mitochondrial genes (Chembazhi et al., 2020; MacParland et al., 2018). The reason for this is the metabolic capacity of hepatocytes, which some have suggested causes an increase in mitochondrial gene percentage. While this may be a reasonable assumption, it is clear from my data that hepatocytes can be sequenced without obtaining high mitochondrial read percentage. One main difference between this study and those published previously is the method of scRNA-seq. For example, Chembazhi *et al.* (2020) and MacParland *et al.* (2018) both used a droplet-based approach as opposed to a plate-based technique. Based on my experience, isolating hepatocytes using a droplet-based technique is challenging as discussed in (section 3.2.3). One might suggest that the hepatocytes sequenced in the aforementioned studies were, to some degree, stressed. In order to reliably identify the source of these differences between different studies, a direct comparison of scRNA-seq methods of hepatocytes isolated from the same source should be performed.

I was able to perform effective gene level QC and remove cell contaminants. A strict threshold was placed to remove potential endothelial doublets and completely mitigate the identification of any endothelial driven gene expression in downstream differential gene analysis. Had a more lenient threshold been used, there would have been an increased chance of false positives occurring.

Several QC factors were then identified as confounding variables, which explained most of the variation in the first few PCs. It has been previously reported that detection rate (i.e. number of genes detected above a certain threshold) is the main driver of PC 1 in most single cell datasets. This was somewhat the case in this study, however total counts (number of read counts) and batch also explained a lot of the variance (Hicks et al., 2017). The total counts and batch were effectively removed with the chosen normalisation and batch correction methods, as shown by a decrease in R-squared values

(Figure 4.13C). The number of genes detected still accounted for some of the variation in the first PC, although this was much lower than that of the biological component. Furthermore, the variability described by the number of genes in this case were likely to originate from a biological perspective. Looking at the *Seurat integration,* where uninjured non-replicating cells were found to group with those that were injured and replicating, the variance explained by the total genes was reduced (Figure 4.13B). This supports the underlying assumption that the variance accounted for by total genes is linked to the biological classification of the hepatocytes. This led me to choose the Harmony method, despite the higher R-squared value for total genes.

Although I have described batch as something that requires removal, in this case, an argument can be made that the difference in batch is actually driven more by timepoint and biology. Unfortunately, due to the constraints of the experimental design, the batch and timepoints are confounded. However, the disparity between the two 48h plates suggested that technical batch differences was causing cells from these technical replicates to separate. The choice to perform batch correction was therefore made for two reasons. If the separation of cells from these two plates was driven by technical variation, removal of the batch component enabled the identification of common differential genes between the timepoints. On the other hand, if biology was the main driver of this separation, grouping the hepatocytes into similar groups allowed for easier classification of heterogeneous populations. Differential expression between clusters and between timepoints can then be made. Nonetheless, the latter should be treated with caution, as this analysis may contain an inflated number of false positives. As a result, validation would be key for any genes identified in this way.

5 Single-Cell RNA Sequencing: Clustering and Differential Gene Expression Analysis

5.1 Introduction

Following QC and normalisation, the subsequent steps in a scRNA-seq analysis pipeline aim to interrogate the data to find novel biology. Two of the main analyses conducted in most workflows involve clustering the cells and performing differential gene expression (DGE) analysis (Luecken and Theis, 2019; McCarthy et al., 2017; Stuart et al., 2019). These analyses are used to identify different populations of cells, as well as the genes they express. These are therefore the two steps I used to answer the core aims and hypothesis of this study. More specifically, potential replicating and non-replicating populations of hepatocytes, key genes expressed by these populations, and in particular the replicating hepatocytes, were identified and assessed to understand their impact on the replicative niche and liver regeneration.

There are many different methodologies of unsupervised clustering of cells in a scRNA-seq pipeline, including K-means, hierarchical and graph-based (Andrews and Hemberg, 2018). Each have individual advantages and disadvantages. However, a graph-based approach has become the most popular, in part due to its primary use in the *Seurat* package (Stuart et al., 2019). Furthermore, an independent study found that the shared nearest neighbour (SNN) graph-based approach implemented in *Seurat* performed better overall than the other algorithms tested (Duò et al., 2018). Graph-based approaches are highly scalable and faster than other methodologies for large scRNA-seq datasets. Moreover, they make fewer assumptions about the shape of the clusters and cell distributions within each cluster (Amezquita et al., 2020).

One difficulty with this form of clustering is the choice of an “optimal” resolution parameter. The resolution parameter controls the number of clusters that are obtained from the SNN clustering. Lower resolutions will yield fewer

clusters and higher resolutions will yield more clusters. The word “optimal” is used lightly here as the choice of resolution can vary depending on the underlying questions. Luckily, there are several parameters that can help guide the selection process. One method is to visualise how the relationships between clusters change across multiple resolutions using a cluster tree (Huang et al., 2017). Stability scores can also be overlaid on such visualisations to assess how stable a cluster is across the various resolutions. The stability score is measured from 0-1, where 1 represents a cluster that does not change at all across all resolutions.

Another commonly used measure is the silhouette score, which ranges from -1 to +1 (Kiselev et al., 2017). This is calculated per cell and is a measure of how similar a cell is to others of its own cluster, compared to the other clusters. Box plots of silhouette scores can be plotted for each cluster at each resolution to assess how the cohesion changes. Again, high scores represent high cohesion between the cells in a cluster. While it might appear counter-intuitive, choosing the resolution with the highest scores is not always the best option. A delicate balance is required that involves selecting the higher scoring methods, deciding on the number of clusters expected and the number of clusters that will give the best chance at discovering novel biology. A pragmatic approach therefore has to be taken in order to make this decision.

Following clustering of cells in scRNA-seq data, DGE analysis can be performed to identify what cell types or subpopulations are present within the data, based on known biology. Pre-defined markers of cell populations and pathway analysis can also help to define the clusters. The DGE analysis also enables identification of novel markers or genes that are expressed by the various cell populations. Despite the fact that specific DGE analyses have been developed for scRNA-seq data, these may not be the best methods for a particular dataset. Recent evidence has suggested that the choice of normalisation effects the performance of the DGE analysis. A simple t-test can perform well with size factor-based normalisations, while MAST, a specific scRNA-seq based approach, performed consistently worse than other methods tested (Vieth et al., 2019). In practice, no consensus has been

reached as to which approach is best when performing DGE analysis. Several methods have been used, including Wilcox rank sum, area under the curve classifier and negative binomial modelling (Aizarani et al., 2019; Halpern et al., 2017; Ramachandran et al., 2019). For this study, a logistic regression (LR) framework was implemented, as this is a fast approach that leverages the large number of cells within single cell data (Ntranos et al., 2018). Furthermore, co-variables can be introduced into the LR model to account for batch when performing the DGE test.

Within this chapter, I use the methods described above to inform my decision on the cluster resolution and use the LR to analyse the differential genes between the different hepatocyte populations.

5.2 Results

5.2.1 Clustering of Hepatocyte Single Cell RNA Sequencing Data

As discussed in chapter 4, *Harmony* was the most optimal batch correction method for this data. As a result, the harmony dimensions used to produce the UMAP visualisations were also used to generate the shared nearest neighbour graph for *Seurat*'s graph-based clustering. A range of resolutions were chosen from 0.1 to 2 to assess which resolutions best fits the data. The team who created *Seurat* recommend using 0.4 – 1.2, which usually provides good results. I therefore included this range span within my choice of resolutions, as well as extending beyond to capture clustering structure that may not fall within the norm. In order to assess which resolutions best fits the data, I initially used the *clustree* R package. This allowed me to visualise the number of clusters found at each resolution and how the proportion and number of cells shift from the clustering of the previous resolution to the next. Overlaying of the cluster's stability score across the tested resolutions was also helpful in deciding the best resolution. I also calculated silhouette scores for each cell at each resolution, which were then plotted via a box plot for each cluster at each resolution.

Clusters at low resolutions (0.1 - 0.5) had high silhouette scores but low stability scores with the exception of cluster ‘3’ (Figure 5.1 & Figure 5.2). While high silhouette scores are desirable, the low number of clusters identified at these low resolutions does not allow for the identification of novel heterogeneity. At the higher resolutions (1.3 – 2), clusters exhibited low stability scores due to the shifting of cells (Figure 5.1). Furthermore, the larger number of clusters results in lower silhouette scores due to the similarity of cells within one cluster to its neighbours (Figure 5.2). Clusters defined by these high resolutions are therefore unlikely to represent realistic biological subpopulations. The middle resolutions provided a more optimal clustering. I subsequently chose resolution 0.9 over 0.7 or 1.1, as most of the clusters had the same, if not better, stability scores (Figure 5.1). In addition, the median silhouette score was higher at 0.9 over 1.1, although lower in comparison to 0.7 (Figure 5.2). Choosing resolution 0.9 balanced the ability to identify more clusters, giving a greater chance of identifying novel heterogeneity, while still retaining a reasonable median silhouette score and good stability scores.

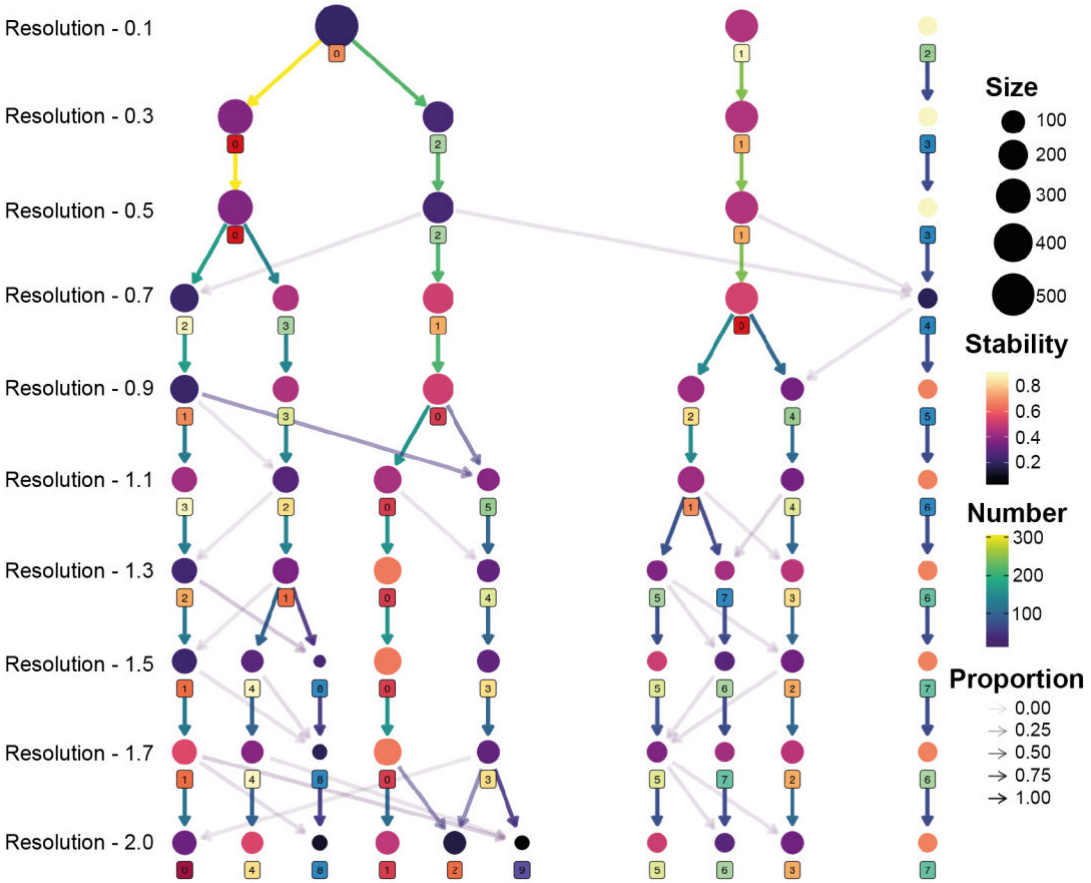


Figure 5.1 – Cluster Tree of Hepatocyte Single Cell RNA Sequencing Data.

Cluster tree showing the number of clusters found from *Seurat* shared nearest neighbour graph-based clustering of the single cell RNA sequencing data of mouse hepatocytes. Data contains cells from uninjured mice and 24h, 48h and 72h post partial hepatectomy. Resolutions 0.1, 0.3, 0.5, 0.7, 1.1, 1.3, 1.5, 1.7 and 2.0 were assessed. The size of each spot represents the number of cells within each cluster, with each cluster spot defined by a number. The numbering of clusters is sequential from large clusters to small. Each spot is shaded by the stability score of that cluster. Arrows show the movement of cells between clusters of different resolutions where the colour representing the number of cells and the transparency represents the proportion of cells.

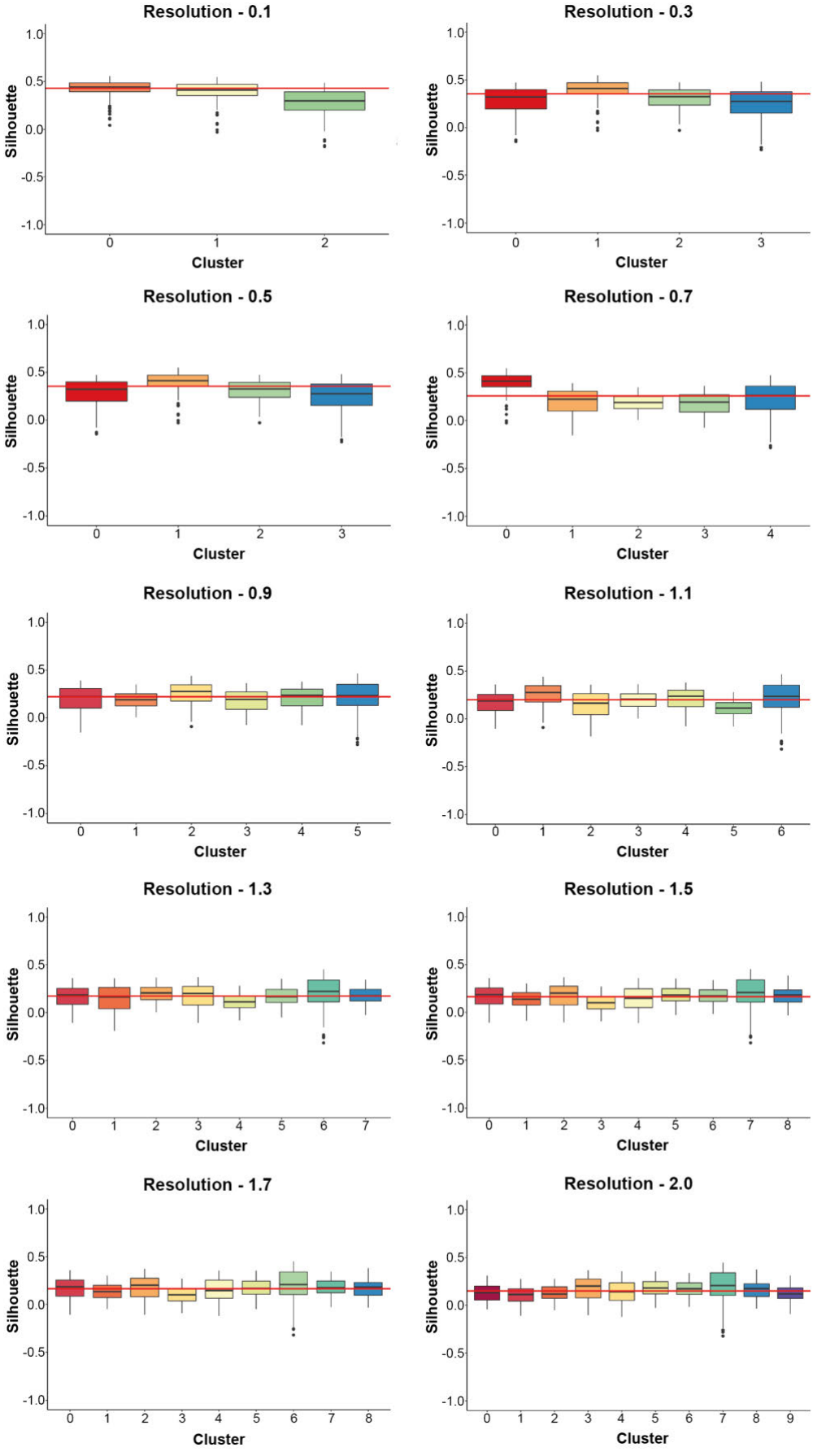

Resolution - 0.1
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
Cluster
Resolution - 0.3
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
Cluster
Resolution - 0.5
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
Cluster
Resolution - 0.7
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
Cluster
Resolution - 0.9
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
Cluster
Resolution - 1.1
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
6
Cluster
Resolution - 1.3
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
6
7
Cluster
Resolution - 1.5
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
6
7
8
Cluster
Resolution - 1.7
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
6
7
8
Cluster
Resolution - 2.0
Silhouette
1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
6
7
8
9
Cluster

Figure 5.2 – Cluster Silhouette Scores of Hepatocyte Single Cell RNA Sequencing Data.

Box plot of the silhouette scores for each cell in each cluster found using *Seurat* shared nearest neighbour graph-based clustering of the single cell RNA sequencing data of mouse hepatocytes. One graph per resolution is shown and labelled. Data contains cells from uninjured mice and 24h, 48h and 72h post partial hepatectomy. Resolutions 0.1, 0.3, 0.5, 0.7, 1.1, 1.3, 1.5, 1.7 and 2.0 were assessed. A red line in each graph represents the median silhouette scores of all cells at each resolution.

5.2.2 Differential Gene Expression Analysis

The next stage in the analysis was to define each of the six clusters identified at resolution 0.9. Figure 5.3A shows a UMAP coloured by cluster assignment. The most common method for achieving this is DGE analysis and known marker interrogation. DGE analysis can give a quick insight into what cell type or subpopulation each cluster may represent by examining the differentially upregulated genes. This can then be confirmed by visualisation of the expression of known marker genes. DGE analysis was performed using a logistic regression (LR) framework for all clusters, where all cells of one cluster were compared to all cells of the remaining clusters. This comparison is made for each cluster individually to generate a list of differentially expressed genes. Plotting the top ten genes by positive average log fold change on a heatmap is a useful method to qualitatively visualise the differences between clusters. This type of DGE analysis is not designed to identify unique markers. However, in practice, many of the top differentially expressed genes will distinguish one or two clusters depending on the overall variation in cell types captured. The top ten genes are expected to show blocks of high scaled expression in one or two clusters, and low expression in the remaining clusters. The genes associated with each cluster also give insight into what cell type or subpopulation each represents.

Performing this analysis showed blocks of expression within the hepatocyte dataset (Figure 5.3B, Appendix 4). Clusters 1 and 3 showed distinct expression between one another but overlapped expression with cluster 1. Clusters 2 and 4 showed very similar expression patterns as well as shared low-level expression of genes associated with clusters 0 and 3 respectively. Cluster 5 expressed distinct genes that were not found to a high degree in any other cluster.

5.2.2.1 Defining Clusters 0, 1, and 3

The reciprocal expression of genes such as Histidine Ammonia-Lyase (*Hal)* and cytochrome P450 genes between clusters 0 and 3 suggested that these clusters defined the periportal and pericentral zones, respectively (Brosch *et al.*, 2018; Kietzmann, 2017; Figure 1.3B). The shared expression of genes in cluster 1 with clusters 0 and 3 therefore suggested that this cluster represents the midzone hepatocyte populations. Building upon this assumption, I plotted known marker genes of the periportal, pericentral and midzone regions (Halpern et al., 2017).

First, cluster 0 had higher expression for Phosphoenolpyruvate Carboxykinase 1 (*Pck1*), E-cadherin (*Cdh1*), and Argininosuccinate Synthase 1 (*Ass1*; Figure 5.4A). These genes have previously been reported as markers of the periportal zone (Halpern et al., 2017; Hempel et al., 2015). Next, I performed pathway enrichment analysis using the significantly differentially expressed genes of this cluster that had an average log fold increase of 0.5 or above, which revealed one term: Metabolism of amino acids and derivatives (Figure 5.4B, Appendix 6). The metabolism of amino acids is a periportal associated process (Kietzmann, 2017). I was therefore able to confidently label this cluster as periportal hepatocytes.

Cluster 3 showed high scaled expression of Cytochrome P450 Family 1 Subfamily A Member 2 (*Cyp1a2*), Cytochrome P450 Family 2 Subfamily E Member 1 (*Cyp2e1*) and Glutamate-Ammonia Ligase (*Glul*; Figure 5.5A). These genes have previously been reported to mark pericentral hepatocytes

(Halpern et al., 2017). Pathway enrichment analysis was performed using the same criteria of DGE genes, which revealed that several pathways were enriched within this cluster (Appendix 9). These terms included cytochrome P450 activity and bile acid synthesis (Figure 5.5B), both of which supported the classification of these hepatocytes as pericentral.

I hypothesised that cluster 1 most likely represented midzonal hepatocytes. However, there are fewer distinct markers for cells within this region as they share expression of genes from both the pericentral and periportal zones. Nevertheless, I used the few non-monotonic markers that have been discovered to aid my confirmation of this cluster as midzonal hepatocytes (Halpern et al., 2017). More specifically, Hepcidin Antimicrobial Peptide (*Hamp*) and Insulin Like Growth Factor Binding Protein 2 (*Igfbp2*) both showed higher expression in cluster 1 hepatocytes (Figure 5.6A). Pathway enrichment analysis of genes from this cluster appeared to show more pericentral associated pathways, such as cytochrome p450 activity and bile acid metabolism and synthesis (Figure 5.6B, Appendix 7). However, the associated p values were higher and number of terms defining each pathway were lower than that of the pericentral cluster (3). I therefore defined these cells as midzonal hepatocytes.

Almost all hepatocytes contained within these three clusters were mCherry^{+} only (Figure 4.12) and all timepoints contributed to each of the three clusters. As such, these clusters represented the non-replicating hepatocyte populations.

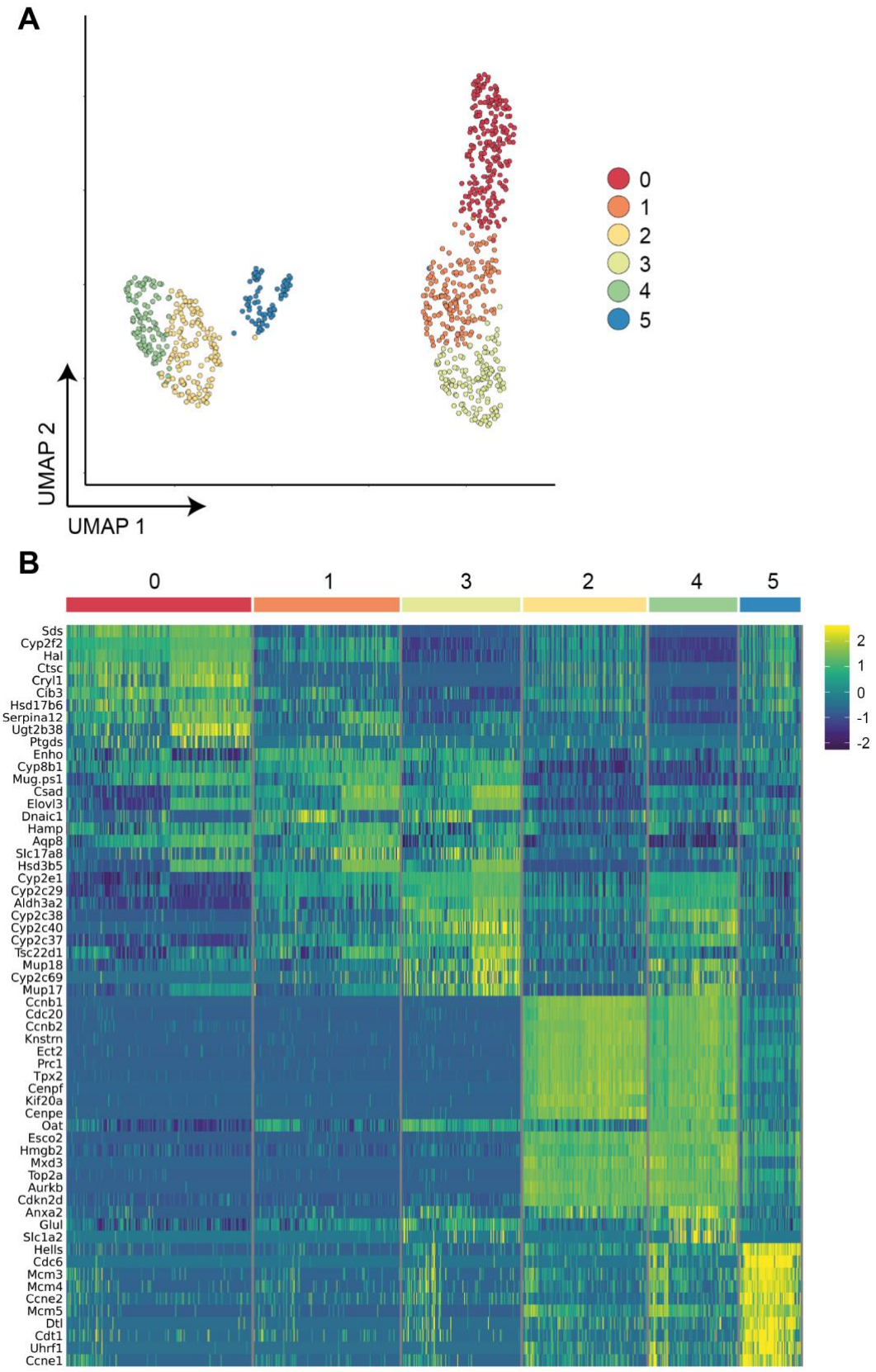

A
0
1
2
3
4
5
UMAP 2
UMAP 1
B
0
1
3
2
4
5
Sds
Cyp2f2
Hal
Ctsc
Cryl1
Cib3
Hsd17b6
Serpina12
Ugt2b38
Ptgds
Enho
Cyp8b1
Mug.ps1
Csad
Elovl3
Dnaic1
Hamp
Aqp8
Slc17a8
Hsd3b5
Cyp2e1
Cyp2c29
Aldh3a2
Cyp2c38
Cyp2c40
Cyp2c37
Tsc22d1
Mup18
Cyp2c69
Mup17
Ccnb1
Cdc20
Ccnb2
Knstrn
Ect2
Prc1
Tpx2
Cenpf
Kif20a
Cenpe
Oat
Esco2
Hmgb2
Mxd3
Top2a
Aurkb
Cdkn2d
Anxa2
Glul
Slc1a2
Hells
Cdc6
Mcm3
Mcm4
Ccne2
Mcm5
Dtl
Cdt1
Uhrf1
Ccne1
2
1
0
-1
-2

Figure 5.3 – Initial Clustering and Differential Gene Analysis of Hepatocyte Single Cell RNA Sequencing Data.

A) Uniform manifold approximation and projection, where each dot represents one hepatocyte. Colour denotes the cluster classification of each cell, based on *Seurat* shared nearest neighbour clustering.

B) Heatmap showing scaled expression of up to 10 significantly differentially expressed genes (<0.05 adjusted p value) weighted by average log fold change for each cluster. Cells are represented by columns and genes are represented by rows. Cells are grouped by cluster annotation.

Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy. Logistic regression was used to identify differential expressed genes.

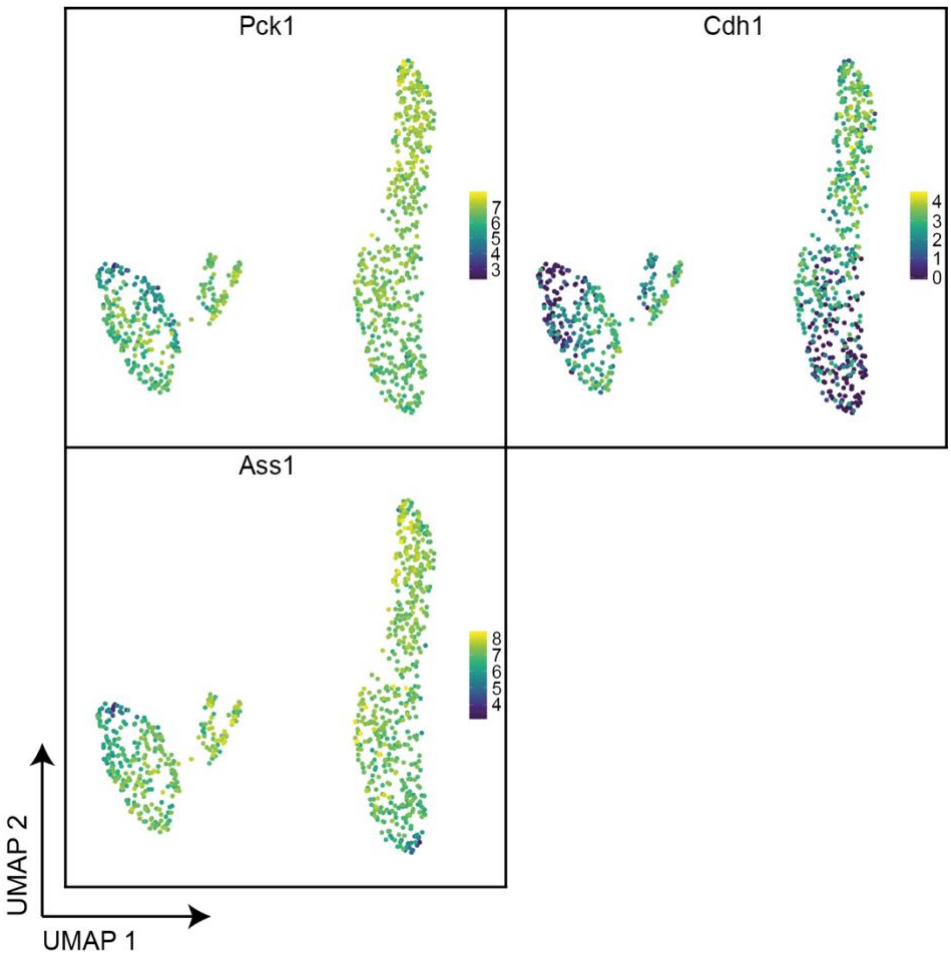


Figure 5.4 – Cluster 0 Marker Gene Expression and Pathway Enrichment

Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the labelled gene. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

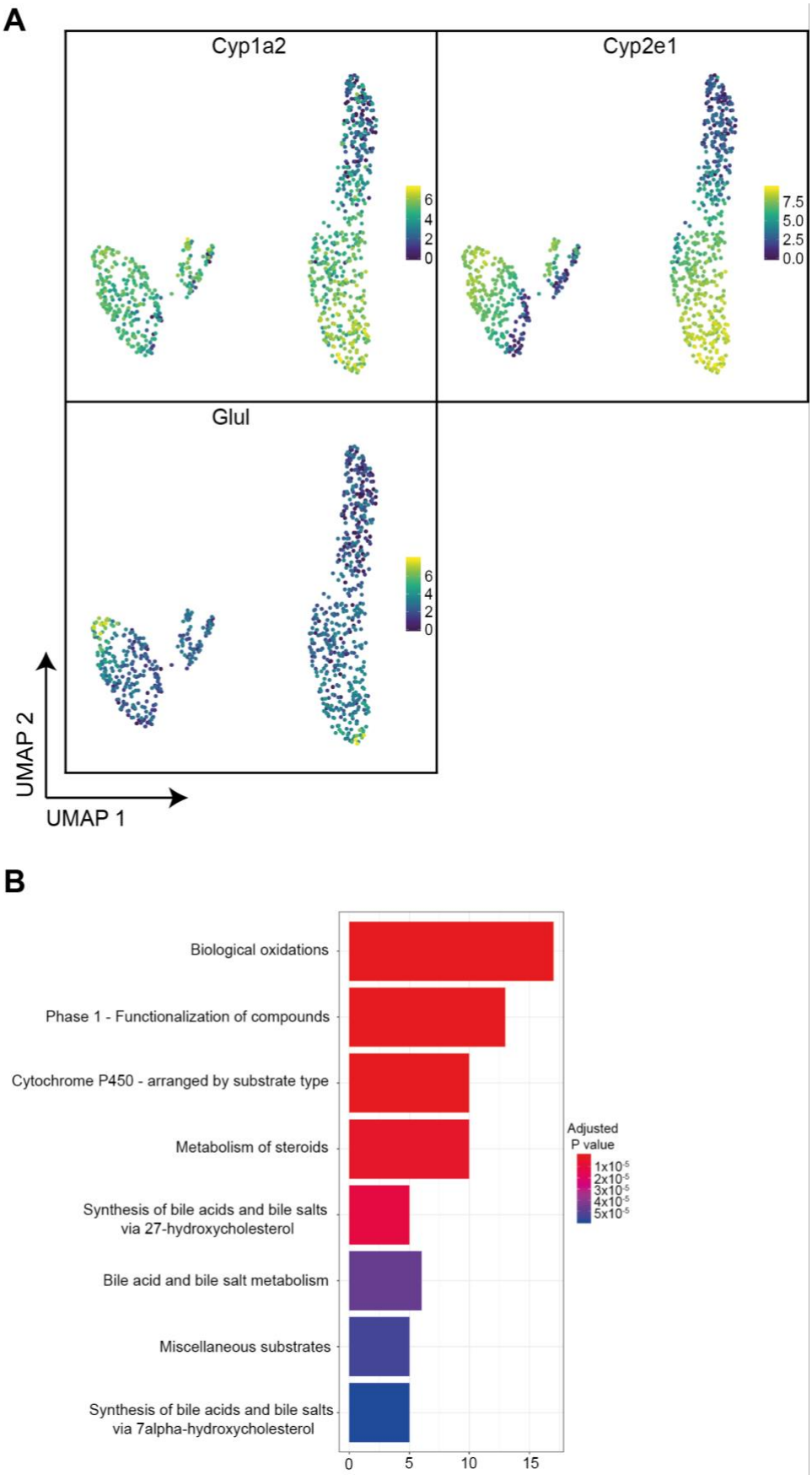

A
Cyp1a2
Cyp2e1
Glul
6
4
2
0
7.5
5.0
2.5
0.0
UMAP 2
UMAP 1
B
Biological oxidations
Phase 1 - Functionalization of compounds
Cytochrome P450 - arranged by substrate type
Metabolism of steroids
Synthesis of bile acids and bile salts via 27-hydroxycholesterol
Bile acid and bile salt metabolism
Miscellaneous substrates
Synthesis of bile acids and bile salts via 7alpha-hydroxycholesterol
0
5
10
15
Adjusted P value
1x10-5
2x10-5
3x10-5
4x10-5
5x10-5

Figure 5.5 – Cluster 3 Marker Gene Expression and Pathway Enrichment

A) Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the labelled gene. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

B) Bar plot showing the top 8 results of pathway enrichment analysis using all significantly expressed differential expressed genes for cluster 3 (<0.05 adjusted p value) with an average log fold change above 0.5. Bars are coloured by adjusted p value and represent the number of genes that are associated with each term. A hypergeometric model in the *ReactomePA* package was used to perform the enrichment test against the Reactome database.

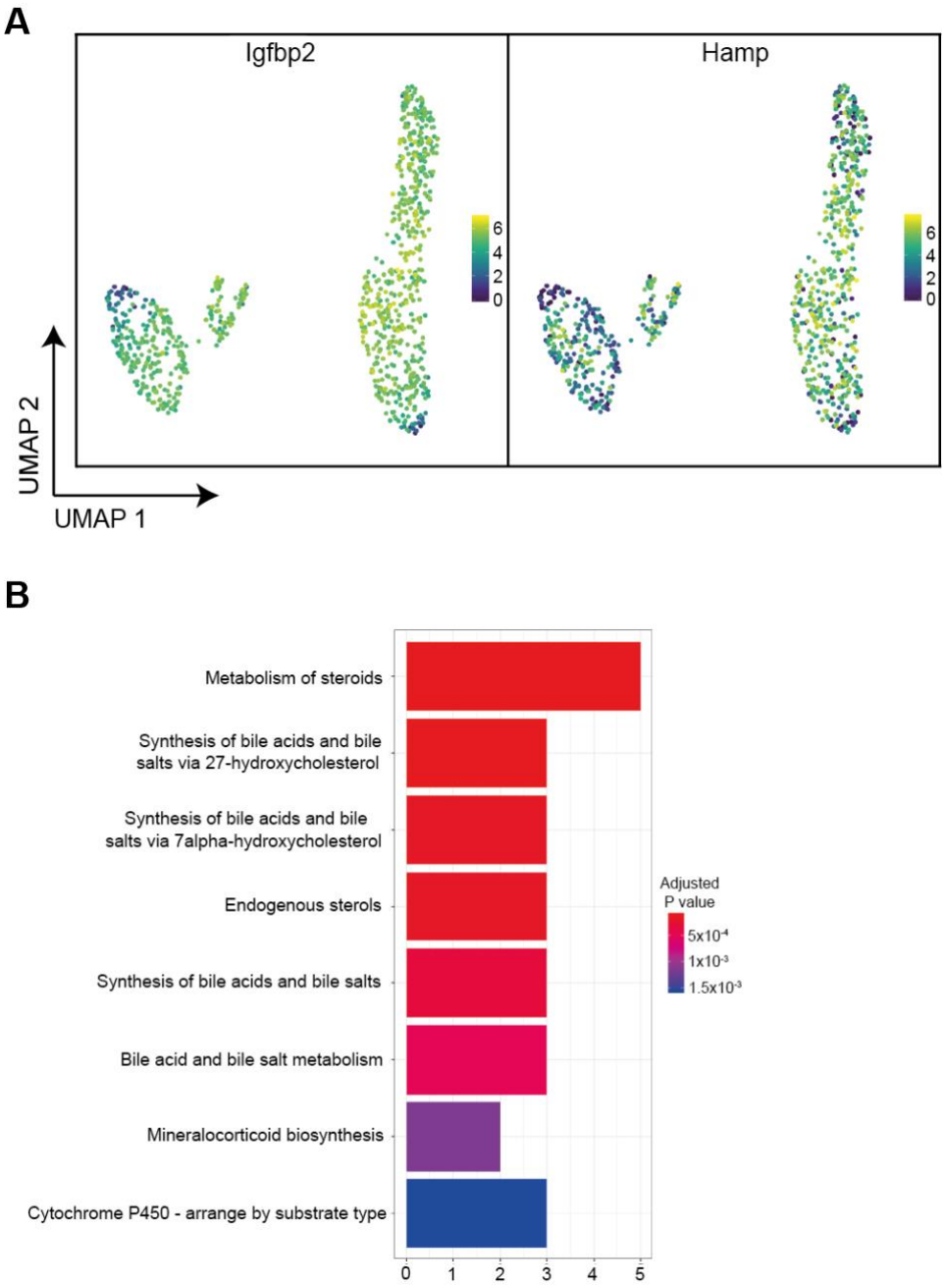


Figure 5.6 – Cluster 1 Marker Gene Expression and Pathway Enrichment

A) Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the labelled gene. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

B) Bar plot showing the top 8 results of pathway enrichment analysis using all significantly differentially expressed genes for cluster 1 (<0.05

adjusted p value) with an average log fold change above 0.5. Bars are coloured by adjusted p value and represent the number of genes that are associated with each term. A hypergeometric model in the *ReactomePA* package was used to perform the enrichment test against the Reactome database.

5.2.2.2 Defining Clusters 2 and 4

Cluster 2 and 4 shared higher scaled expression for many of the same marker genes (Figure 5.3B). These included cell cycle division proteins and cyclins, suggesting that these two clusters represented the replicating hepatocytes. This was consistent with the classification of the hepatocytes in this cluster as either $Dual^{+}$ or $mVenus^{+}$ only (Figure 4.12). Nevertheless, I decided to plot two well-known markers of proliferation on a UMAP: Marker of Proliferation Ki-67 (*Mki67*) and DNA Topoisomerase II Alpha (*Top2a*; Figure 5.7A). This confirmed that all cells in this cluster were expressing cell cycle markers and that there were no outliers.

The pathway analysis for cluster 2 and 4 revealed that these cells were likely to be in M phase or late S phase (preparing to enter M phase; Figure 5.7B &C, Appendix 8, Appendix 10). While both clusters displayed similar expression levels of cell cycle genes, the expression of zonated genes differed, resulting in the separation of cells into 2 clusters. Cluster 2 expressed many of the same zonal markers as cluster 0, defining these as the periportal replicating hepatocytes (Figure 5.4A). On the other hand, cluster 4 expressed many of the same zonal markers as cluster 3, suggesting that these cells represented the pericentral replicating hepatocytes (Figure 5.5A). Capturing this zonal division in the replicating populations enabled the examination of potential differences in gene expression between periportal and pericentral replicating hepatocytes.

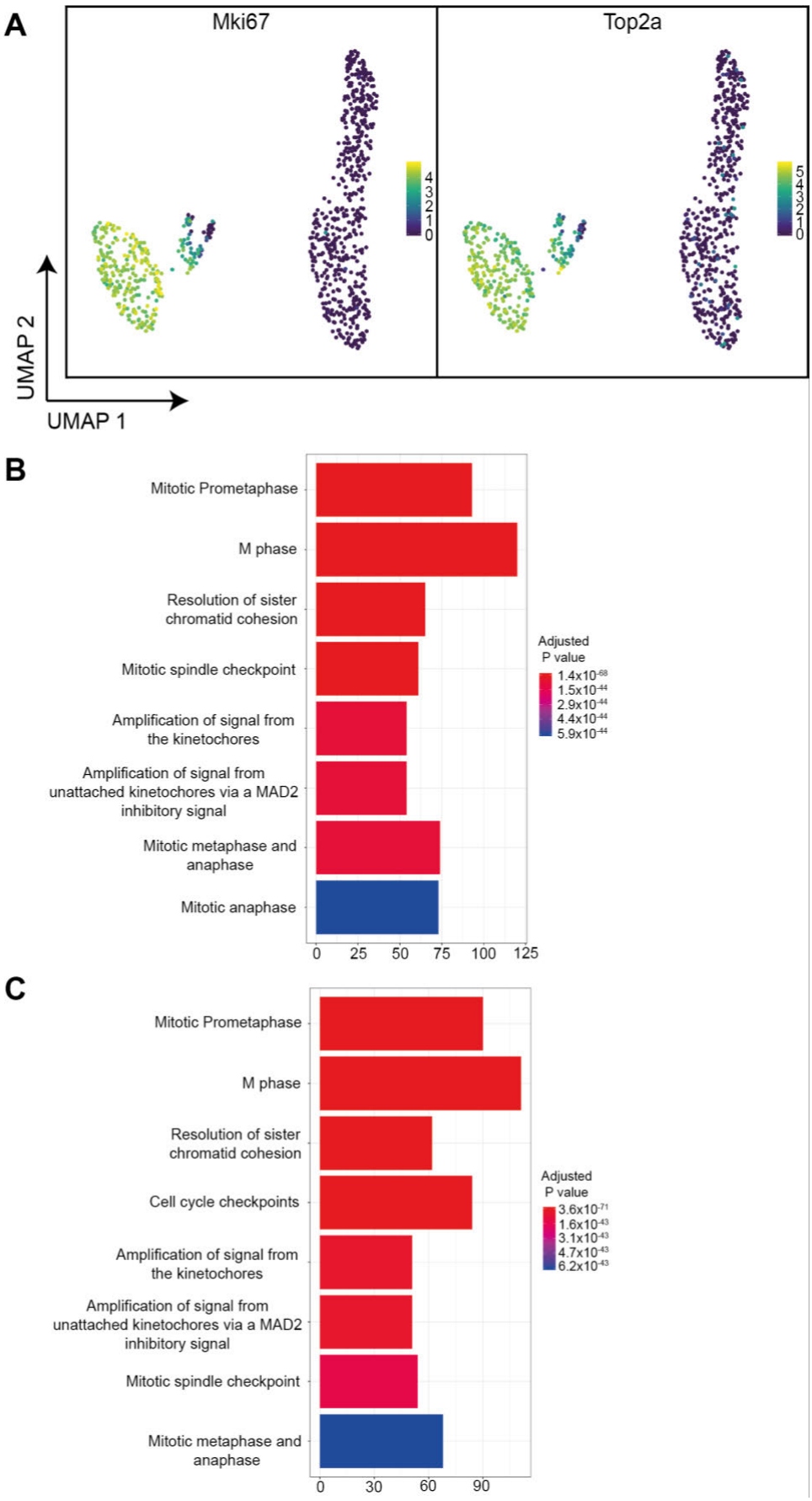

A
Mki67
Top2a
UMAP 2
UMAP 1
B
Mitotic Prometaphase
M phase
Resolution of sister chromatid cohesion
Mitotic spindle checkpoint
Amplification of signal from the kinetochores
Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal
Mitotic metaphase and anaphase
Mitotic anaphase
Adjusted P value
C
Mitotic Prometaphase
M phase
Resolution of sister chromatid cohesion
Cell cycle checkpoints
Amplification of signal from the kinetochores
Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal
Mitotic spindle checkpoint
Mitotic metaphase and anaphase
Adjusted P value

Figure 5.7 – Cluster 2 and 4 Marker Gene Expression and Pathway Enrichment

A) Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the labelled gene. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

B & C) Bar plot showing the top 8 results of pathway enrichment analysis using all significantly differentially expressed genes for cluster 2 (B) or cluster 4 (C) (<0.05 adjusted p value) with an average log fold change above 0.5. Bars are coloured by adjusted p value and represent the number of genes that are associated with each term. A hypergeometric model in the *ReactomePA* package was used to perform the enrichment test against the Reactome database.

5.2.2.3 Defining Cluster 5

Cluster 5 was fairly distinct from the other hepatocyte populations (Figure 5.3A). Similarly to clusters 2 and 4, cluster 5 had higher scaled expression for cell cycle related genes (Helicase, Lymphoid Specific - *Hells,* Cell Division Cycle 6 - *Cdc6*, Cyclin E2 - *Ccne2),* although didn't share the same expression patterns as that of cluster 2 and 4. *Cdt1* was within the top ten differentially expressed genes for cluster 5. Plotting the expression of *Cdt1* on a UMAP showed the specificity of its expression to this one cluster (Figure 5.8A). In the Fucci system, CDT1 protein abundance is used as a marker of the G1 cell population (Mort et al., 2014b). Indeed, the majority of cells in this cluster were defined as $mCherry^{+}$ only (Figure 4.12). However, expression of the Geminin gene (*Gmnn*) was also high in these hepatocytes, as well as the other replicating populations (Figure 5.8A). The mCherry protein signal within these cells suggests that these are non-replicating hepatocytes, although their transcriptional profile appears to be similar to those replicating.

Pathway analysis was conducted in the same manner as above for this cluster and revealed interesting terms (Appendix 11). DNA replication and S phase were the top enriched terms suggesting that these cells were replicating. Whereas, the Mitotic-G1-G1/S phases term suggested that these cells maybe in a transition state (Figure 5.8B). Moreover, separate clustering of these hepatocytes from the two replicating clusters showed that these cells were clearly in a different phase of the cell cycle. For these reasons, I decided to describe these cells as primed hepatocytes. At a protein level, these hepatocytes are still in G1 and therefore non-replicative. However, they expressed many genes associated with the S phase, demonstrating that they were preparing or “primed” for replication, or potentially in transition.

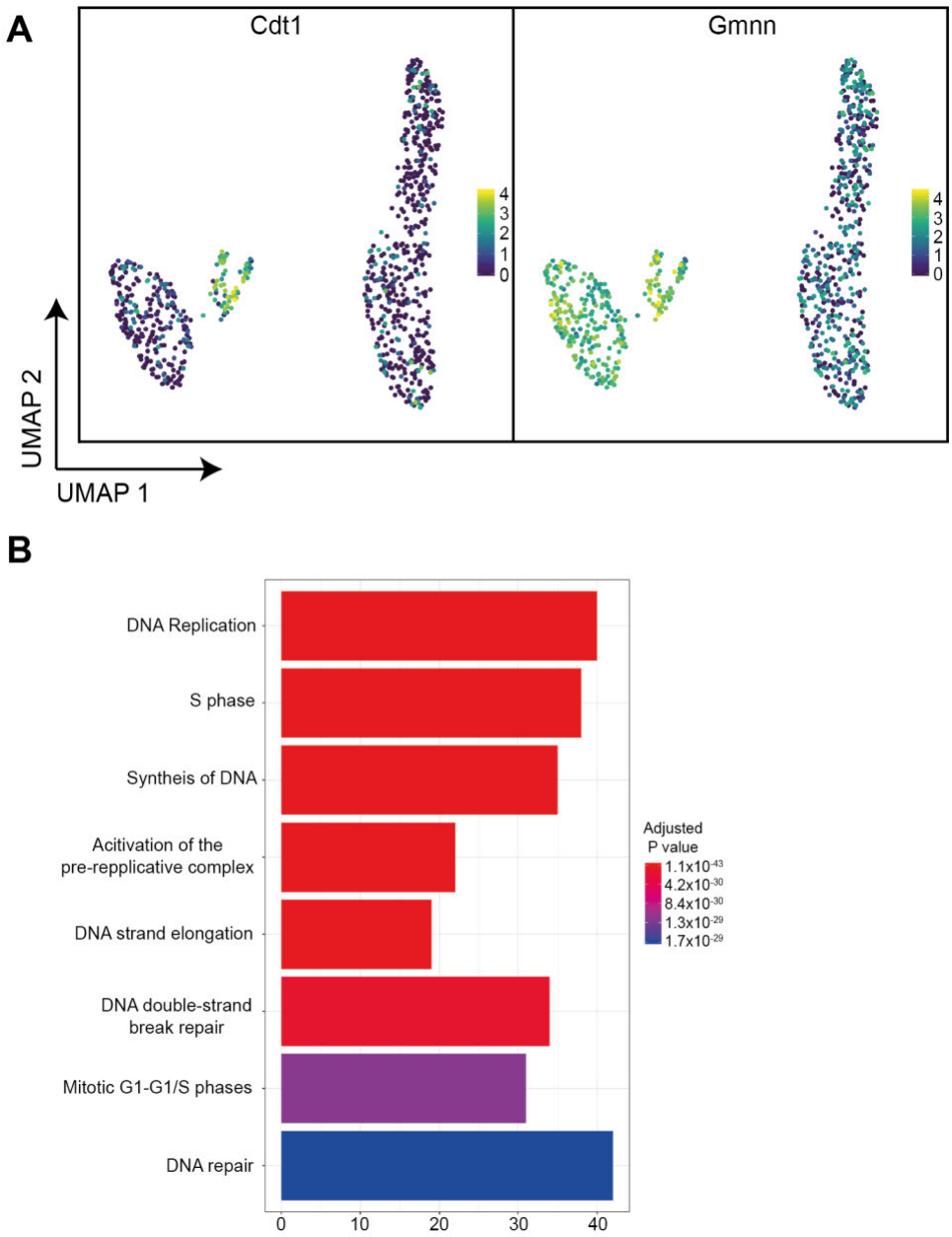


Figure 5.8 – Cluster 5 Marker Gene Expression and Pathway Enrichment

A) Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the labelled gene. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

B) Bar plot showing the top 8 results of pathway enrichment analysis using all significantly differentially expressed genes for cluster 5 (0.05 adjusted p value) with an average log fold change above 0.5. Bars are

coloured by adjusted p value and represent the number of genes that are associated with each term. A hypergeometric model in the *ReactomePA* package was used to perform the enrichment test against the Reactome database.

5.2.3 Ligand/Receptor Analysis

Once each cluster was defined (Figure 5.9), the aim was to assess key signals that are produced by replicating hepatocytes that can control and influence the regenerative niche. The primary mechanism through which cells communicate with the environment is through cell surface receptors and by excreting or presenting ligands. In order to identify genes that are upregulated by the replicating population, I compared the differentially expressed genes from the DGE analysis to a known list of ligands and receptors. Such a list can be found from CellphoneDB, a manually curated database of ligands and receptors (Efremova et al., 2020). The comparison was performed with the current DGE gene list, as the two replicating clusters (2 & 4) only contained proliferating injured hepatocytes (Figure 5.7 & Figure 4.12). The resulting genes were visualised on a heatmap of scaled expression (Figure 5.10, Appendix 5).

While some genes were expressed predominantly in only one of the replicating clusters (e.g. Leucine Rich Repeat Containing G Protein-Coupled Receptor 5 [*Lgr5*], Natriuretic Peptide Receptor 2 [*Npr2*]), these genes were also expressed in the corresponding non-replicating cluster (Figure 5.10). Most genes appeared to be consistently upregulated in both replicating clusters.

This does not mean no new genes of interest were identified. Interestingly, two interleukins (IL) were expressed predominantly in the replicating populations: *Il15* and *Il33* (Figure 5.11 & Figure 5.12). IL15 has previously been described to play an important pro-regenerative role during PHx (Suzuki et al., 2006). However, this is the first evidence to suggest that it is produced mainly by the replicating population. IL33 has been shown to play

a protective role during viral hepatitis, but its role during liver regeneration and PHx remain unknown (Carrière et al., 2017). The expression of *Il33* also appeared to be higher in the periportal replicating hepatocytes compared to the pericentral hepatocytes. This hints at potential zonation of the gene. The replicating hepatocytes also appeared to express another immunomodulatory molecule, Chemokine Like Factor (*Cklf*; Figure 5.11 & Figure 5.12). Again, the functions of the ligand have not yet been studied in the context of liver regeneration.

Among the expressed cell surface proteins, Delta 1 (*Dll1*) was predominantly expressed in these two cluster (Clusters 2 & 4; Figure 5.11 & Figure 5.12). Many studies have looked at the implication of Notch signalling during liver regeneration. Removal of Notch has been shown to be detrimental to the regenerative response, although it is not yet completely understood if the signals are pro-regenerative or inhibitory (Croquelois et al., 2005). Notch and Jagged are the two main components that have been studied in this pathway; the implications of Delta remain unclear.

As well as marking the two replicating clusters, *Il33*, *Cklf* and *Dll1* were also expressed in the primed cluster but to a lesser extent (Figure 5.11 & Figure 5.12). Given potential priming of these hepatocytes to enter into DNA replication, it is not surprising they upregulate the same markers that are expressed in the two replicating populations.

As discussed previously, many of the ligand and receptor genes are expressed in both replicating clusters. However, this is not the case for *Bmp7,* which appeared to show a specific pericentral expression pattern (Figure 5.11 & Figure 5.12). Bone Morphogenetic Protein 7 (*Bmp7*) is an important pro-regenerative molecule during PHx, although its potential zonal expression is unknown (Sugimoto et al., 2007). *Bmp7* expression was seen to a lesser degree in the primed cluster (Figure 5.11 & Figure 5.12). As this did not separate into periportal and pericentral populations within the current clustering strategy, I was unable to confirm whether the same zonal pattern of expression persists in this population. Nevertheless, one might speculate that

the expression of *Bmp7* is zonated within this cluster based on its limited expression within the cluster.

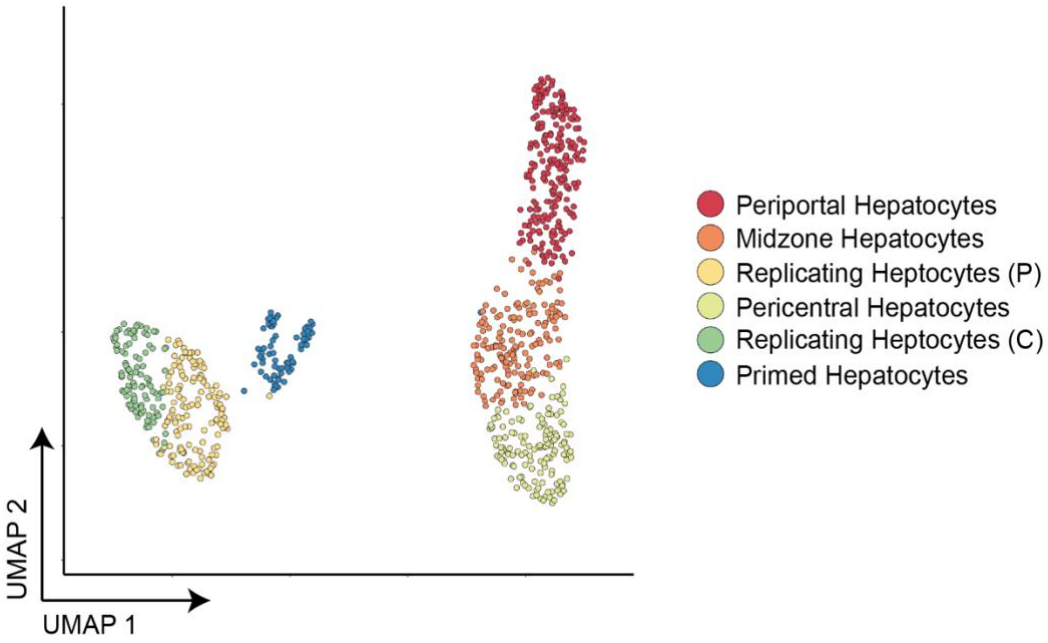


Figure 5.9 – Final Hepatocyte Cluster Annotation

Uniform manifold approximation and projection, where each dot represents one hepatocyte. Colour denotes the cluster classification of each cell, based on *Seurat* shared nearest neighbour clustering. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

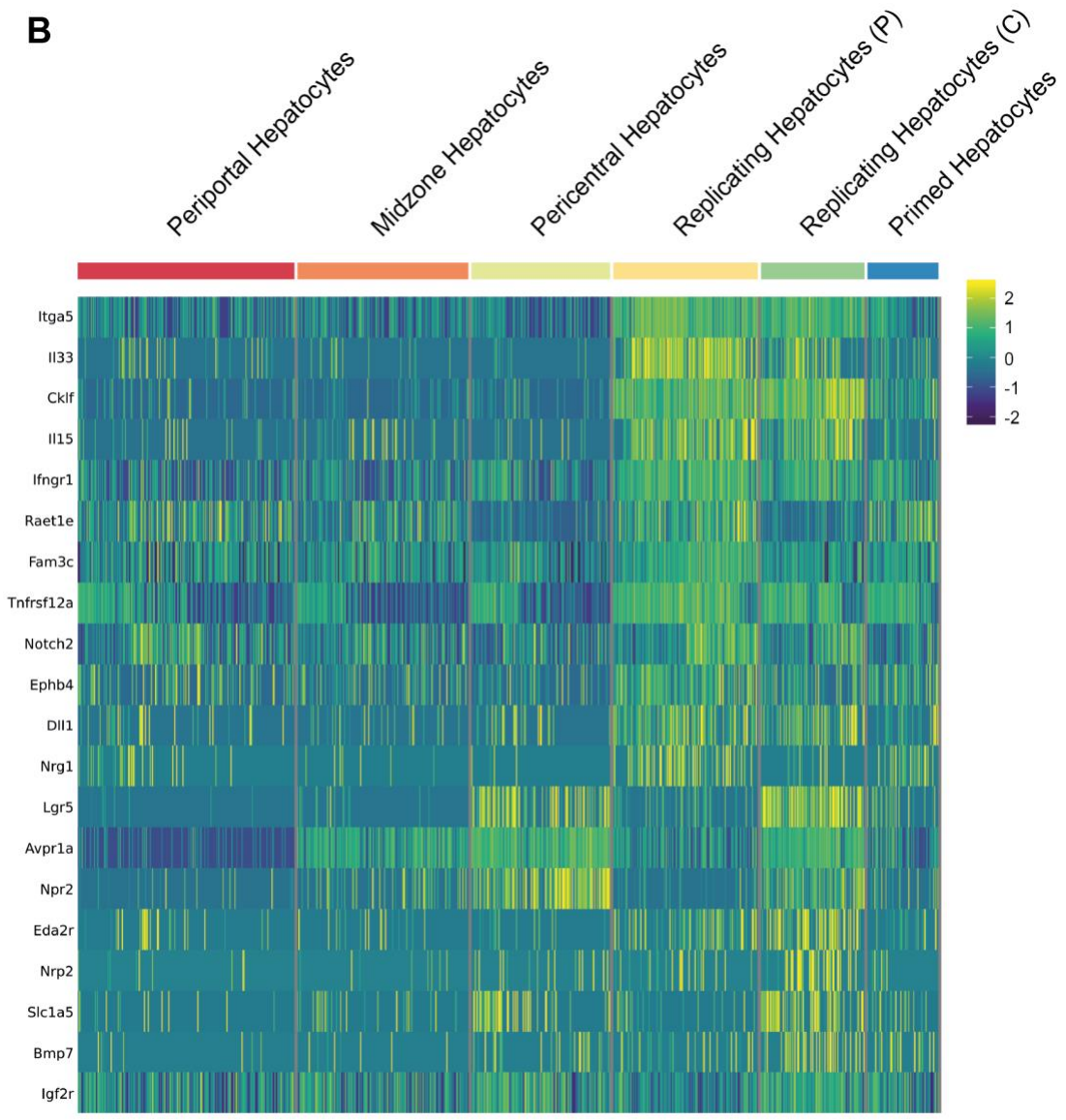


Figure 5.10 – Heatmap of CellphoneDB Expressed Genes

Heatmap showing scaled expression of ligand and receptor genes significantly upregulated in the replicating hepatocyte populations. Logistic regression was used to generate differentially expressed genes. Genes significantly (<0.05 adjusted p value) upregulated in the replicating hepatocyte (P) and (C) populations were cross referenced with genes from CellphoneDB. Cells are represented by columns and genes are represented by rows. Cells are grouped by cluster annotation. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

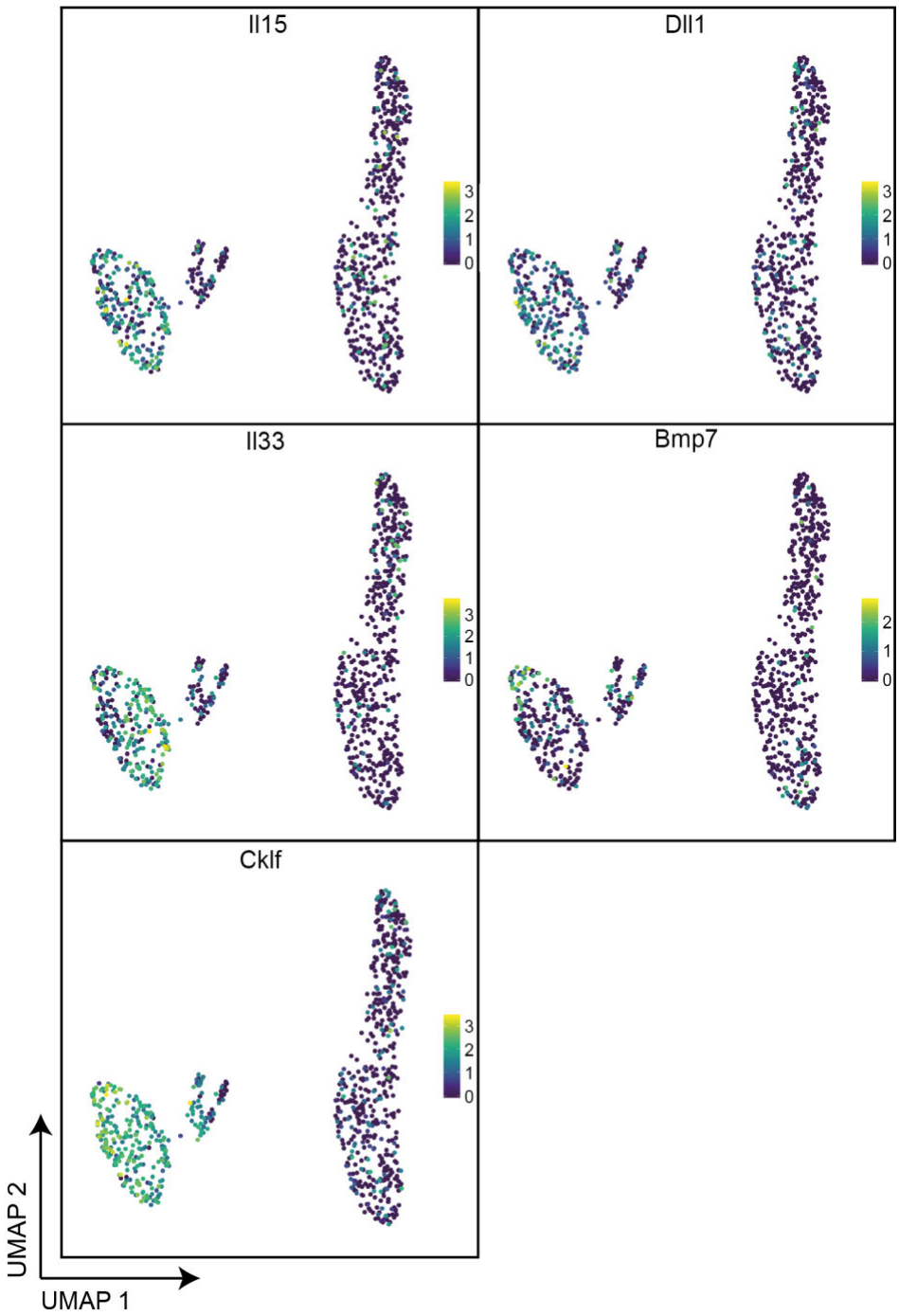


Figure 5.11 – Feature Plots of Key CellphoneDB Genes

Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the labelled gene. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

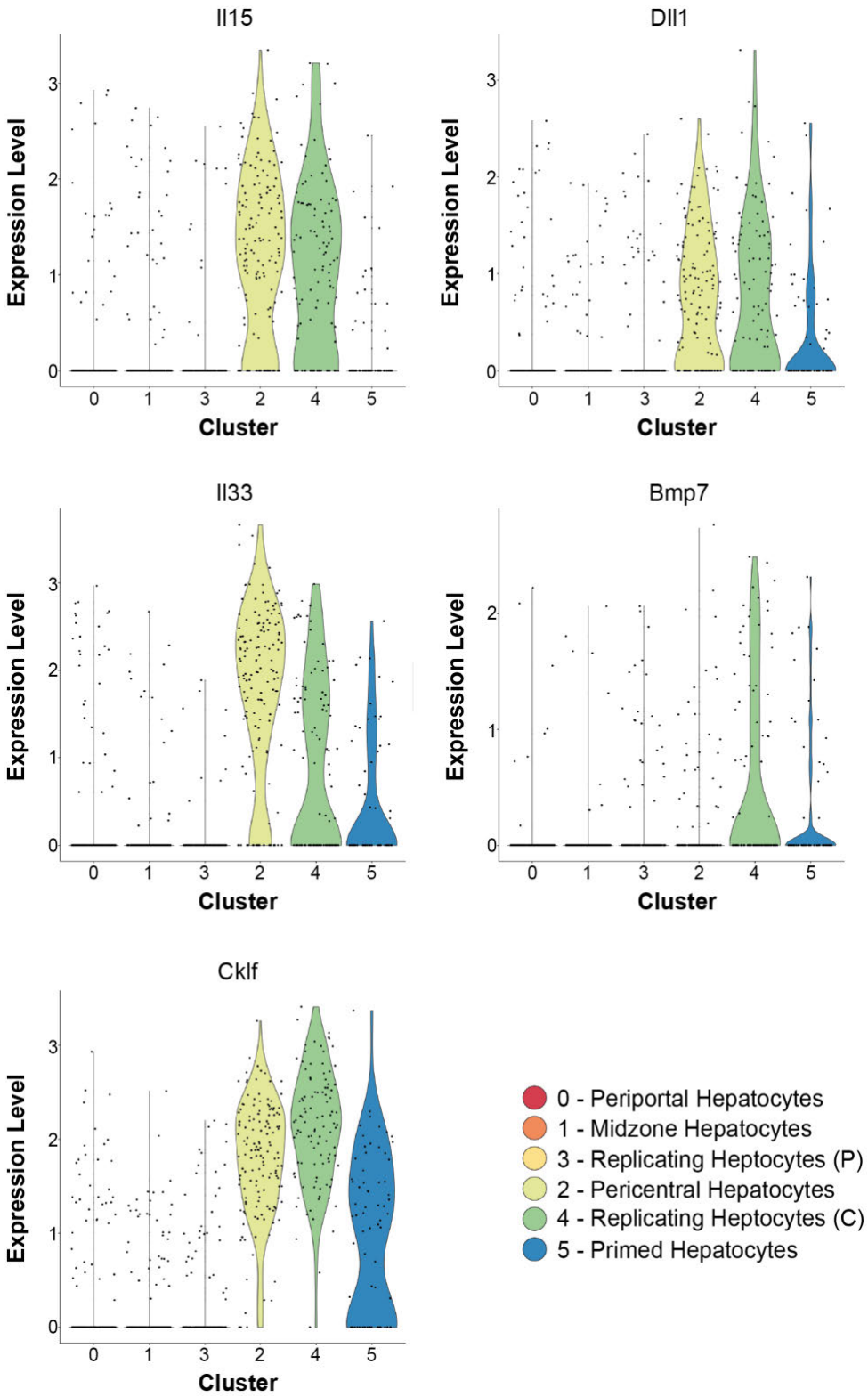


Figure 5.12 – Violin Plots of Key CellphoneDB Genes

Violin plots showing the log normalised expression level of the labelled genes. Each violin represents expression profile from one cluster. Each dot represents

one hepatocyte. Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

5.2.4 Circadian Rhythm, Cell Cycling and Timeless Gene Expression Following Partial Hepatectomy

Following interrogation of the replicating populations, I sought to determine whether any key ligands or receptors were upregulated by the primed hepatocytes compared to other populations. By doing so, I sought to identify signals that trigger a non-replicating hepatocyte to enter the cell cycle in order to promote the regenerative process. Unfortunately, no ligand or receptor genes from the Cellphone database were uniquely expressed by the primed hepatocytes. However, the *Timeless* gene was highly specific to this cluster. The protein produced by this gene has been implicated in both cell cycle and circadian rhythm regulation, thereby connecting the two processes. This therefore implies a potential role for the circadian rhythm in controlling the hepatic regenerative response following PHx.

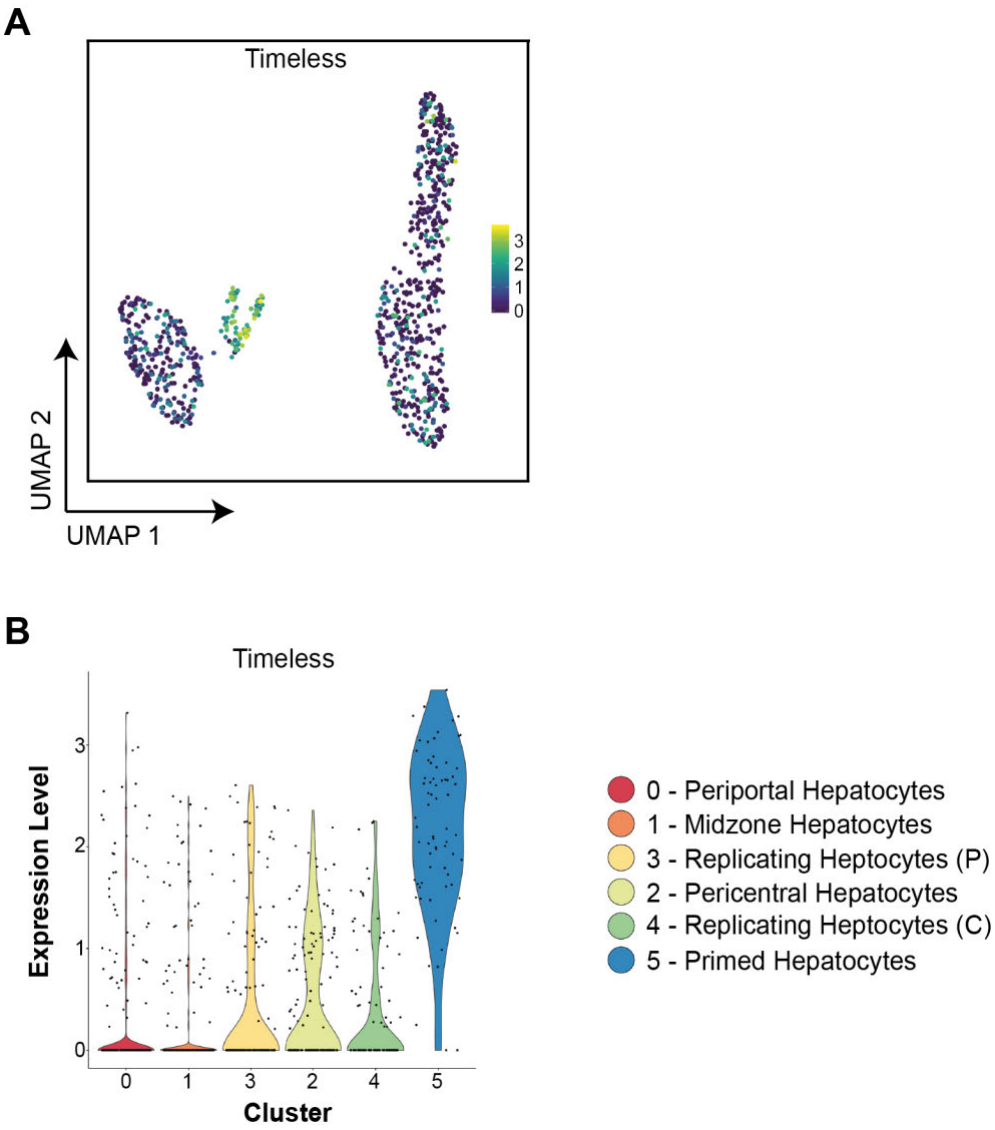


Figure 5.13 – Timeless Gene Expression

A) Uniform manifold approximation and projection, where each dot represents one hepatocyte. Dots are coloured by the log normalised expression of the *Timeless* gene.

B) Violin plots showing the log normalised expression level of the *Timeless* genes. Each violin represents expression profile from one cluster. Each dot represents one hepatocyte.

Data used to generate the plots was from single cell sequencing of uninjured mouse hepatocytes and 24h, 48h and 72h post partial hepatectomy.

5.3 Discussion

The aim of this chapter was to identify clusters of hepatocyte subpopulations within my scRNA-seq dataset. Furthermore, I aimed to assess their differential markers with the goal of identifying key signalling molecules produced by the replicating hepatocytes. Within this chapter I was able to use scoring metrics to choose the optimal resolution for *Seurat* shared nearest neighbour clustering. DGE analysis enabled the classification of the six populations: three non-replicating populations that were split by their zonation profile, two replicating clusters also split by zonation, and one cluster of primed hepatocytes. Cross referencing the differentially expressed genes with CellphoneDB highlighted key markers produced by the replicating hepatocytes that may control and regulate the regenerative niche. Furthermore, I provide evidence that suggests the circadian rhythm plays a role during the hepatic regenerative process.

Identification of hepatocyte clusters associated with zones has previously been identified in other scRNA-seq studies. Furthermore, changes in gene expression across these zones has been extensively studied by Halpern *et al.* (2017). The identification of these three clusters (0, 1 & 3) of non-replicating hepatocytes therefore fits with the current literature. All uninjured cells reside within these three clusters, which is as expected based on the experimental design.

Based on my previous observation that there is no proportional difference in the potential of periportal and pericentral hepatocytes to replicate following partial hepatectomy, it was reassuring to see that cells from all timepoints post PHx contributed to the two replicating clusters. Within the replicating populations, I was surprised to only identify zonal heterogeneity. Previous studies have shown only half of the hepatocytes that replicate their genome continue to M phase and cytokinesis (Miyaoka et al., 2012; Miyaoka and Miyajima, 2013). I therefore expected to see more heterogeneous subpopulations that show differences in their expression due to the lack of M phase transition. One potential explanation for this is insufficient capture of

such cells or that the mechanisms that control these events are not regulated at a transcriptional level. While no heterogeneity was found in hepatocyte cell cycle completion, I was able to identify a novel “primed” hepatocyte population.

I was also able to identify key signals produced by these replicating and “primed” cells, confirming the hypothesis that replicating hepatocytes produce specific signals following PHx. Of the 20 ligand/receptor genes identified following cross referencing, five were of particular interest.

A previous study investigating the impact of Il15 on liver regeneration has already been conducted (Suzuki et al., 2006). Suzuki *et al.* (2006) demonstrated that administration of mice with IL15 increased mature hepatocyte proliferation, over vehicle controls. Furthermore, they also observed an accumulation of oval cells, the liver stem cell that can differentiate into both hepatocytes and biliary cells. The authors speculated the accumulated oval cells differentiated into hepatocytes to enhance liver regeneration. However, the study only used whole liver lysates to uncover the hepatic expression of IL15. My work provides evidence that expression of IL15 is restricted or upregulated in proliferating hepatocytes, potentially forming a positive feedback loop. Il15 has also been associated with the progression of fibrosis (Jiao et al., 2016). Knock out of IL15 receptor α showed an increase in liver fibrosis, suggesting IL15 plays a protective role. One might speculate that hepatocyte senescence caused by liver fibrosis exacerbates the condition as a result of a decrease in hepatocyte replication and therefore Il15 production.

IL33 has not yet been studied in the context of liver regeneration. There have been several published reports of IL33 being both beneficial and harmful in certain liver diseases. For instance, IL33 has positive effects on viral hepatitis but can also promote HSC activation, which can exacerbate liver fibrosis (Carrière et al., 2017; Sun et al., 2017). The immunomodulatory role of IL33 might be of most interest in the context of liver regeneration. IL33 is a known potent activator of eosinophils, which in turn have been linked with improved liver regeneration (Johansson et al., 2017). Eosinophils produce IL4, which acts directly through Il4 receptor α to promote hepatocyte proliferation (Goh et al., 2013).

CKLF is a known potent chemoattractant for a variety of immune cells, including neutrophils, monocytes and lymphocytes (Han et al., 2001). However, this immunoregulatory property may not be the most interesting function in this context. CKLF1 and CKLF2, two of the isoforms of the CKLF gene, possess the ability to increase proliferation of skeletal muscle cells (Han et al., 2001; Xia et al., 2002). It remains to be seen whether CKLF has the same pro-proliferative trait with other cells types and through which mechanism CKLF acts to achieve this. Nevertheless, the strong expression of this gene by replicating hepatocytes following PHx warrants further investigation. One might speculate that CKLF may act to enhance hepatocyte proliferation, or potentially that of other cell types within the liver.

The upregulation of *Bmp7* in the regenerating hepatocytes is interesting with respect to the current literature. For instance, administration of recombinant human BMP7 was shown to facilitate liver regeneration following PHx in mice (Sugimoto et al., 2007). However, BMP7 was not expressed in the liver and was believed to be released from the kidney. My results diverge from this: *Bmp7* appeared to be expressed by pericentral replicating hepatocytes. One potential reason for this discrepancy is the methodology employed to detect BMP7. Sugimoto *et al.* (2007) used whole liver lysates to assess BMP7 expression but given the confined expression of *Bmp7* to the pericentral replicating hepatocytes, detecting the protein might prove difficult from a whole lysate. Protein analysis from a sorted population of replicating pericentral hepatocytes might show an increase in protein expression, as indicated by the increased gene expression.

While one would hypothesize a pro-regenerative role for the four molecules/genes (IL15, IL33, CKLF, BMP7) mentioned so far, DLL1 may function to inhibit hepatocyte proliferation. As discussed previously, there is supporting evidence of NOTCH1 playing an inhibitory role on hepatocyte proliferation (Croquelois et al., 2005). For example, KO of NOTCH1 after birth led to nodular hyperplasia in mouse livers. Regeneration post PHx in these mice was decreased, however, the authors speculated the larger liver to body weight ratio in the KO mouse was the primary reason for this, with regenerative

potential inversely correlated to the amount of remnant liver. The expression of *Dll1* in replicating hepatocytes therefore creates a paradigm, in which those that have committed to proliferation signal to prevent the surrounding hepatocytes from entering the cell cycle. This may force the consistent equal replication of hepatocytes across the liver lobule and prevent the nodular regeneration seen in NOTCH1 KO mice. The spacial control of liver regeneration remains unclear, although Dll1 may represent a prime candidate to examine in order to elucidate this mystery. Furthermore, DLL1 may also play a role in hepatocyte fate specification of hepatic progenitor cells (HPCs). Notch signalling is known to coordinate the differentiation of HPCs into biliary cells following PHx (Lu et al., 2016). However, knockdown of DLL1 in cultured liver progenitor cells results in a decrease in HNF4α expression and an increase in osteopontin^{+} (OPN) cells, a biliary cell marker (Kaylan et al., 2018). One might therefore suggest that DLL1 modulates the surrounding regenerative niche to enhance hepatocyte fate trajectories of HPCs.

There is strong evidence that suggests all five genes/proteins discussed above play important roles in liver regeneration. The identification of these upregulated ligands has allowed me to form several hypotheses about their function within the context of liver repair. Overall, replicating hepatocytes potentially produce a range of secreted ligands that enhance and promote proliferation in a global context. On the other hand, the cell surface ligand DLL1 may inhibit local hepatocytes from replicating in order to maintain the structure of the liver lobes and prevent nodular regeneration.

Finally, as discussed previously, I identified a population of primed hepatocytes. These were referred to as primed due to the discrepancy between their protein and transcriptional expression profiles. The lack of an mVenus^{+} signal in these cells places them in either G1/G0 of the cell cycle. However, at the transcript level, many of their differentially expressed genes are associated with either the S phase or the G1/S phase transition. For this reason, I believe these cells were either preparing to enter S phase and were waiting for the right signals, or had received the correct signal and were just beginning to replicate. Zou *et al.* (2012) have demonstrated that hepatocytes

have a synchronous cell cycle entry with mitosis peaking at Zeitgeber time (ZT) 0 throughout the PHx time course. Matsuo *et al.* (2003) proposed that the Bmal1-Clock/Wee1/Cdc2 pathway controls the circadian rhythm of hepatocyte G2/M transition, and therefore the synchronicity of the regenerative response. Bmal1 should be low at ZT0, however, Zou *et al.* (2012) found evidence that Bmal1 was high in hepatocytes at this point. Furthermore, Wee1 did not follow the same pattern as Bmal1. This suggested that the Bmal1-Clock/Wee1/Cdc2 pathway may not be in control of the hepatocyte cell cycle. As such, it was interesting to identify *Timeless* as a differentially expressed gene for the primed cluster. Studies have shown a circadian regulation of TIMELESS expression in mammals. Furthermore it has been suggested that TIMELESS functions as an intra-S and replicating checkpoint protein (Barnes et al., 2003; Ünsal-Kaçmaz et al., 2005). Knock down of TIMELESS can result in defective replication checkpoints and the early entry of cells into mitosis under DNA damaging conditions. Ünsal-Kaçmaz *et al.* (2005) found TIMELESS expression to be high in S/G2/M phases and low during G1/G0 in normal human fibroblasts (NHF1). The gene expression data from this study suggests that TIMELESS expression is high in late G1 and early S phase, whereas no/low expression of *Timeless* was observed in the replicating clusters that were in G2 and M phase. Considering the defective and early entry of cells into M phase after UV treatment and when TIMELESS is down regulated, it appears this protein prevents cells from progressing through the cell cycle (Ünsal-Kaçmaz et al., 2005). In support of this, TIMELESS interacts with Checkpoint Kinase 1 (CHK1), which when phosphorylated, can lead to slowing or stalling of DNA replication (Patil et al., 2013; Ünsal-Kaçmaz et al., 2005). This mechanistic paradigm would suggest that hepatocytes are stalled around S phase as a result of the expression of TIMELESS. Following downregulation of TIMELESS, hepatocytes can enter S phase, doing so in a synchronous manner. Further work will be required to assess protein expression of TIMELESS and understand its mechanism of action.

6 Discussion

The PHx model has been extensively studied over the past few decades and many of the signalling pathways and mechanisms of regeneration have been uncovered. However, it is clear that we do not have the full picture and new signals are still being uncovered as well as changes in our understanding of the overall mechanisms involved (Fazel Modares et al., 2019; Miyaoka and Miyajima, 2013; Wang et al., 2019). The heterogeneous nature of hepatocytes has also been widely studied, although primarily in the context of metabolism. For instance, since the mid to late 1900s, it has been known that hepatocytes from the periportal zone express different genes and carry out different functions compared to those residing in the pericentral region. Nevertheless, Halpern *et al.* (2017) were able to uncover novel biology with the advent of single cell RNA sequencing, a technology that enables the heterogenous nature of any cell type or tissue to be investigated. Despite this, the central dogma for many years was that hepatocytes react in a homogeneous manner during partial hepatectomy, with all hepatocytes entering the cell cycle and dividing at least once. However, in 2012, Miyaoka *et al.* (2012) demonstrated that this is not the case and only a certain percentage of hepatocytes fully complete mitosis. This suggested a heterogeneous hepatocyte response to regeneration. The aim of this project was therefore to use scRNA-seq to examine the potential heterogeneous nature of hepatocellular regeneration and identify novel signalling molecules that could play a role. I hypothesized that subpopulations of replicating hepatocytes produce key signals that regulate the hepatic regenerative niche.

I first demonstrated that the PHx model could be reliably reproduced and determined that peak hepatocyte proliferation occurred at 48h. However, my results suggested that a periportal to pericentral wave of hepatocyte proliferation does not occur following PHx, which diverges from previously published studies (Rabes et al., 1976; Sun et al., 2020). Following model characterisation, a hepatocyte isolation protocol was optimised for the purpose of scRNA-seq. I was unable to produce usable sequencing data from a droplet-

based sequencing approach. However, I successfully produced scRNA-seq data using a plate-based technique.

I also developed a bespoke analysis pipeline that enabled me to maintain complete control of all the analysis approaches used in this study. For example, I considered several different approaches of thresholding in order to remove poor quality cells (and genes) and examined the use of multiple sophisticated normalisation strategies. In addition, I compared two batch correction methods and assessed their suitability using the intrinsic *a priori* Fucci signal classifications.

Following QC, downstream analysis revealed that hepatocytes formed 6 clusters. Clusters 0, 1 and 3 were identified as non-replicating periportal, midzone and pericentral hepatocytes, respectively. Clusters 2 and 4 consisted of replicating periportal and pericentral cells, respectively. Finally, cluster 5 was suggested to be a novel “primed” non-replicating hepatocyte population. As expected, uninjured hepatocytes were only found in clusters 0, 1 and 3, as $mCherry^{+}$ only hepatocytes were sorted form uninjured mice. Hepatocytes isolated following PHx were present in all 6 clusters.

I further identify several signals specifically expressed by replicating cells. Based on previously published literature, I hypothesize that these molecules potentially play important roles in the regenerative process and even control the regenerative niche. Moreover, the primed hepatocyte cluster represents a potentially novel phase within the cell cycle of hepatocytes, in which they are pre-programmed to undergo DNA replication. Interestingly, TIMELESS was found to be expressed specifically by hepatocytes within this cluster. The circadian rhythm and cell cycle regulatory functions of this protein suggest that the expression of TIMELESS may regulate hepatocyte cell cycle entry (Ünsal-Kaçmaz et al., 2005).

6.1 Future and Ongoing Work

Due to time constraints, only one biological scRNA-seq replicate was performed for each timepoint. While additional timepoints are unlikely to

change the overall conclusions drawn in this study, performing further replicates may enable the identification of smaller differences with more confidence. More specifically, a pseudobulk approach could be employed in order to compare different timepoints. This procedure accounts for type 1 error and reduces the number of incorrectly identified differentially expressed genes in the DGE analysis (Lun and Marioni, 2017).

In addition to generating more replicates, re-evaluation of the existing data can continue as further advances in scRNA-seq analyses are made. Over the past few years, many new analysis methods have been developed for scRNA-seq data and it is likely that this will continue (Luecken and Theis, 2019). Datasets can therefore be analysed using these new approaches, which could potentially lead to the generation of new hypotheses. In addition, as more single cell experiments are conducted, a large pool of data will become publicly available. This will allow for comparisons with my data and enable further discoveries. For example, by integrating datasets from PHx that contain different cell types, inferences about cell to cell interactions can be made. Programs such as *Cellphone* or *NicheNet* aim to answer such questions (Browaeys et al., 2020; Efremova et al., 2020). Pseudotime is another area of active study with a multitude of different algorithms to infer a time trajectory within single cell data (Saelens et al., 2019). What can be considered as the main pseudotime R package, *Monocle*, is currently in *beta* testing of its third iteration, demonstrating how quickly techniques are updated as we gain greater understanding of scRNA-seq data (Qiu et al., 2017).

As discussed in chapter 5, several genes of interest were identified that were specifically expressed by the replicating and primed populations, including *Il33*, *Il15*, *Cklf*, *Dll1*, *Bmp7* and *Timeless*. Further work is required to validate the protein expression of these markers. For example, immunofluorescence staining could be performed to determine whether or not these markers are exclusively expressed/upregulated in specific hepatocyte populations, as observed in the scRNA-seq data. Where commercially available antibodies are unavailable, the scRNA-seq results could be confirmed by Florescence In-Situ Hybridisation (FISH), in which mRNA is

directly stained in tissue samples. Proteins specific to replicating hepatocytes could also be validated by western blot or qPCR, as a bulk sample could be obtained by sorting these cells from a Fucci mouse.

Following validation, one could assess the function of each protein in the context of liver regeneration and PHx. For secreted molecules such as CKLF, which I hypothesised to act directly on hepatocytes, *in vitro* techniques could be employed similar to those used to discover the auxiliary mitogen norepinephrine (Cruise et al., 1985). For example, the proliferation of primary hepatocytes in culture could be quantified with and without the addition of CKLF in order to assess its mitogenic potential. The effect of CKLF can be tested both with and without additional growth factors, such as HGF and EGF, two commonly used mitogens in hepatocyte cell culture (Greenhalgh et al., 2019). On the other hand, in cases where the protein of interest may act indirectly to increase hepatocyte proliferation, co-cultures or conditioned media experiments could be performed to evaluate their potential function. For example, I previously hypothesised that IL33 activates eosinophils to produce IL4, which subsequently causes an increase in hepatocyte proliferation (Goh et al., 2013). Media from eosinophils cultured in the presence of IL33 could be applied to hepatocytes *in vitro* to assess cellular replication. In addition, a similar experiment could be performed in which IL33 is administered to co-cultures of hepatocytes and eosinophils. An increase in the number of proliferating hepatocytes of those cultured with eosinophils compared to those without, would also support this hypothesis. Protein analysis of the culture media could reveal these molecular signals produced by the eosinophils under IL33 treatment.

While *in vitro* techniques are very useful tools, they cannot fully recapitulate the complex nature of the PHx model. Conditional and/or inducible KO mice can be created for each gene and the impact on PHx assessed. Changes in the regenerative response, such as decreased or increased hepatocyte proliferation, changes in lobular structure or in liver to body weight ratio all indicate an important role in liver regeneration. For instance, similar to the inducible KO of Notch1 shown by Croquelois *et al.* (2005), a conditional

and inducible Dll1 KO could be generated and tested in a similar manner. If nodular regenerative hyperplasia was observed in a Dll1 KO, this would provide evidence that Dll1 is the key ligand for Notch1 mediating this phenomenon. Taking this further, deletion of Dll1 prior to PHx may lead to dysregulation of the regenerative response, and an increase in proliferation. Croquelois *et al.* (2005) noted that Notch1 KO mice displayed lower hepatocyte proliferation over WT due to a larger liver to body weight ratio. The aforementioned study had waited several days after the removal of Notch1 and allowed the liver to increase in size before performing the PHx. I would aim to assess the impact on PHx more immediately after removal of Dll1 to assess its direct impact on PHx. Any increase in proliferation or changes in lobular structure would provide evidence that Dll1 controls the spatial organisation of hepatocellular regeneration.

As well as assessing the importance of these genes in the context of PHx, the wider influence of these genes should also be considered. Comparison to scRNA-seq data generated from acetaminophen or carbon tetrachloride-induced liver injury models could aid our overall understanding of the functions of these genes during liver regeneration.

Ultimately, I would like to assess the expression of Dll1, CKLF, IL33, IL55, BMP7 and TIMELESS in human tissue through immunofluorescence and flow cytometry. Interrogation and validation of the data generated within this project in humans is vital with regard to gaining insights into the human liver regenerative response. In conjunction, comparison of these target genes can be cross referenced with human scRNA-seq data. A handful of human liver single cell datasets are publicly available (Aizarani et al., 2019; MacParland et al., 2018). However, many of these are from ischemic livers due to the nature of tissue collection. For instance, those obtained are either diseased or deemed unsuitable for transplantation and as such, would have been without a blood supply for an extended period of time. In order to overcome these issues, our group has developed a method for removing a small segment of the liver, prior to interruption of blood flow. Previous attempts to isolate live hepatocytes from this small cube of human liver tissue proved difficult,

however advances in single nuclei RNA-seq protocols mean that in the future we will be able to accrue hepatocyte data from these biopsy samples.

6.2 Concluding Remarks

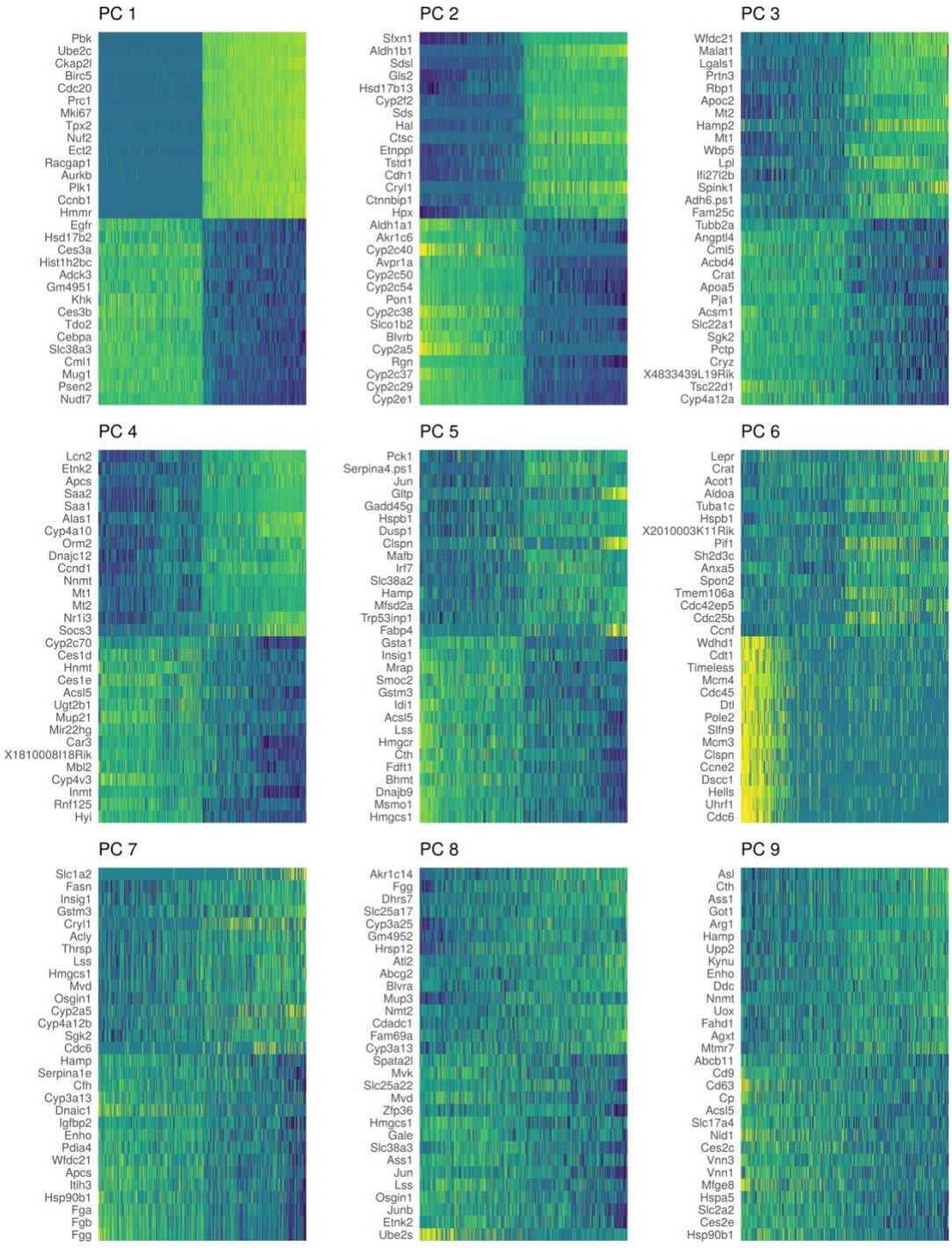
Using scRNA-seq, I have shown that multiple signalling molecules are specifically produced by replicating hepatocytes. Several of these, such as Il15, Il33, and CKLF, may be pro-regenerative and therefore provide a potential therapeutic target for liver disease. Direct administration may not be feasible for some of these molecules as their roles go beyond liver regeneration (Han et al., 2001; Miller, 2011; Perera et al., 2012). IL15, IL133 and CKLF all have immunomodulatory roles and direct administration of these may lead to off target effects that would not be desirable. However, understanding the downstream effector molecules and signalling will generate a deeper understanding of the regenerative process. This in turn will lead to new avenues to explore, producing more specific targets. On the other hand, potential proliferation inhibitors such as Dll1 may lead to new treatments for hepatocellular carcinoma.

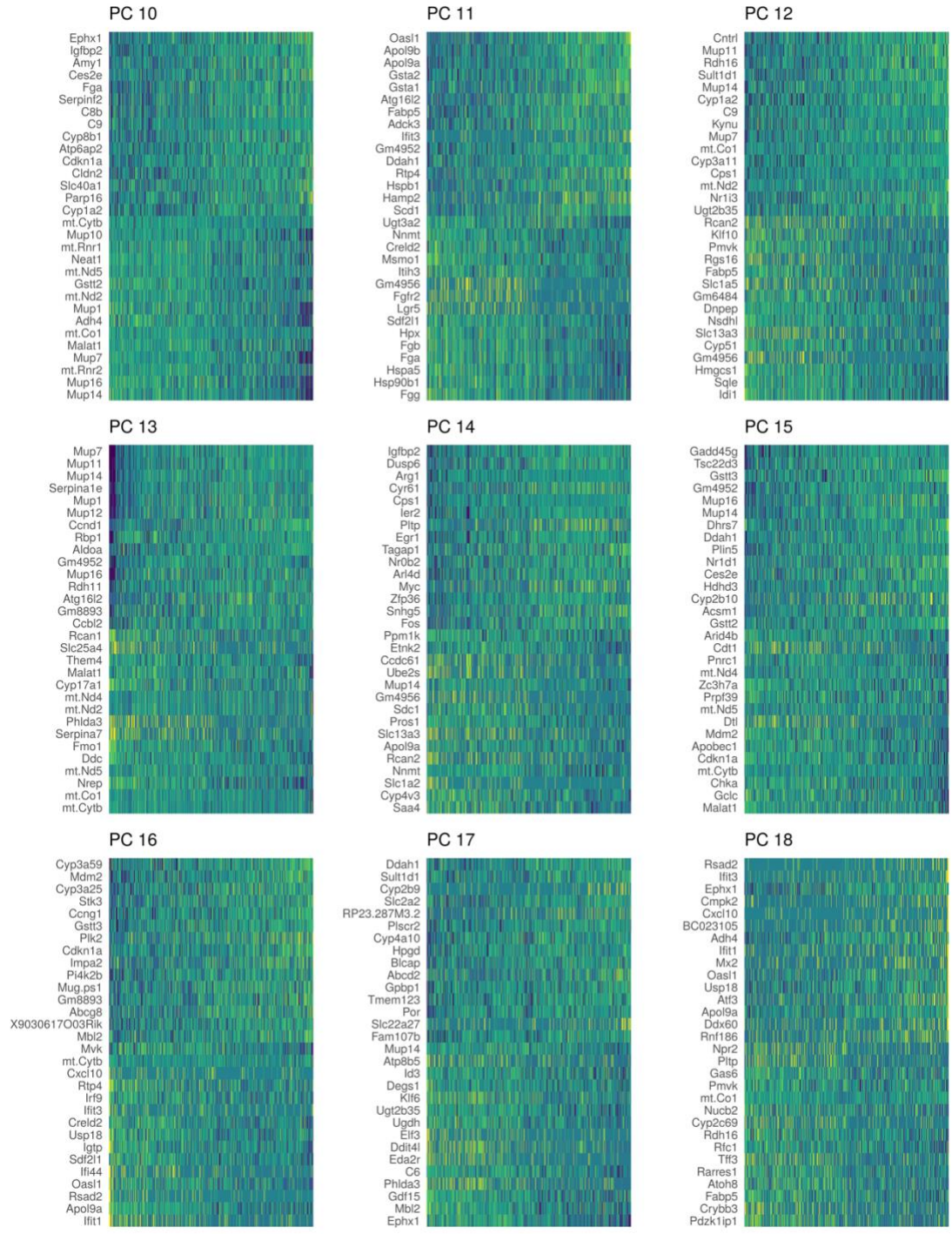
Understanding the function of these proteins and their mechanism of action following PHx may prove valuable to future studies. It is clear that liver regeneration is a complex, heterogeneous response and high-resolution techniques are required to uncover the finer details of this process. Only once the complete picture has been discovered, will we be able to fully control this marvellous biological system.

7 Appendix

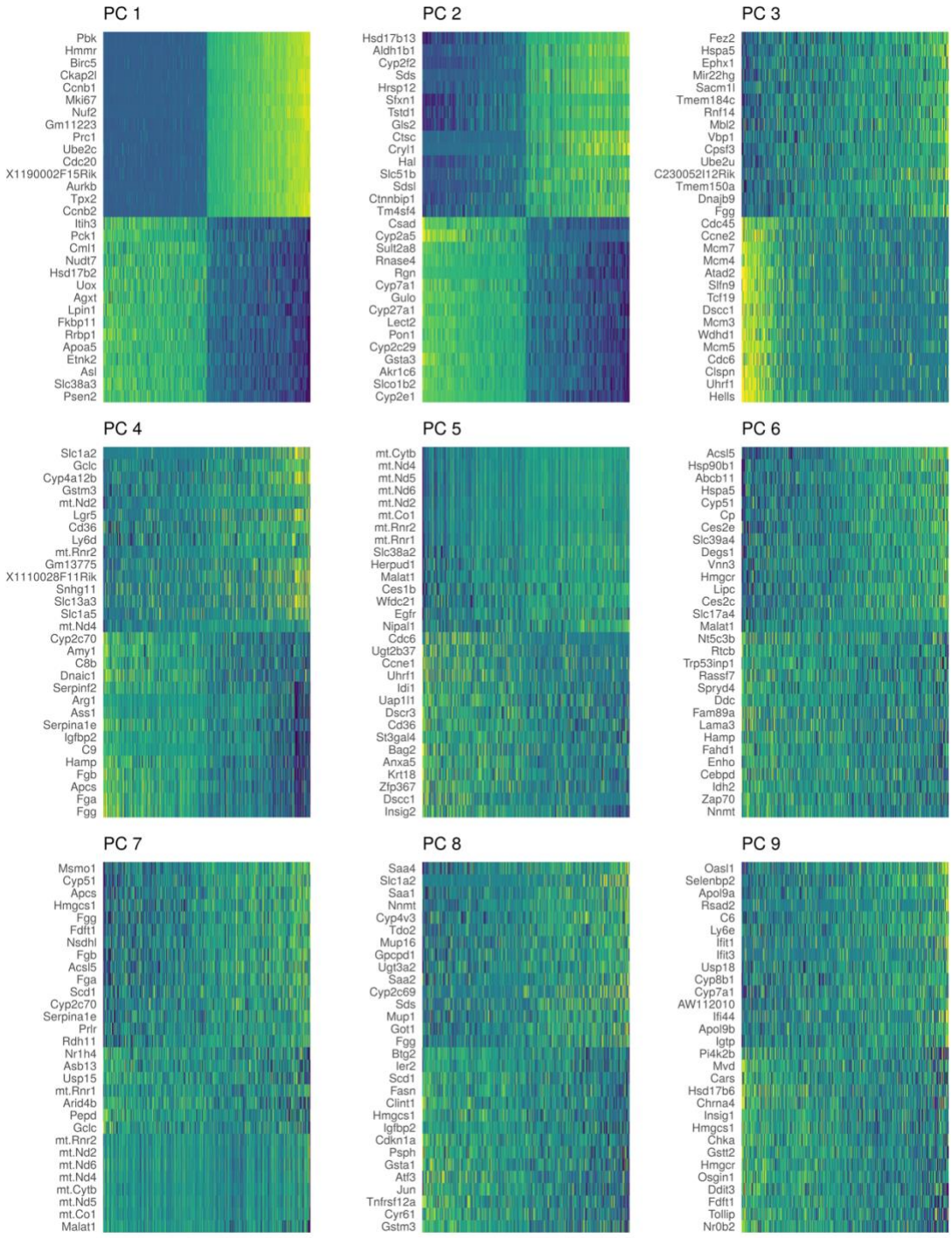
7.1 PCA Heatmaps

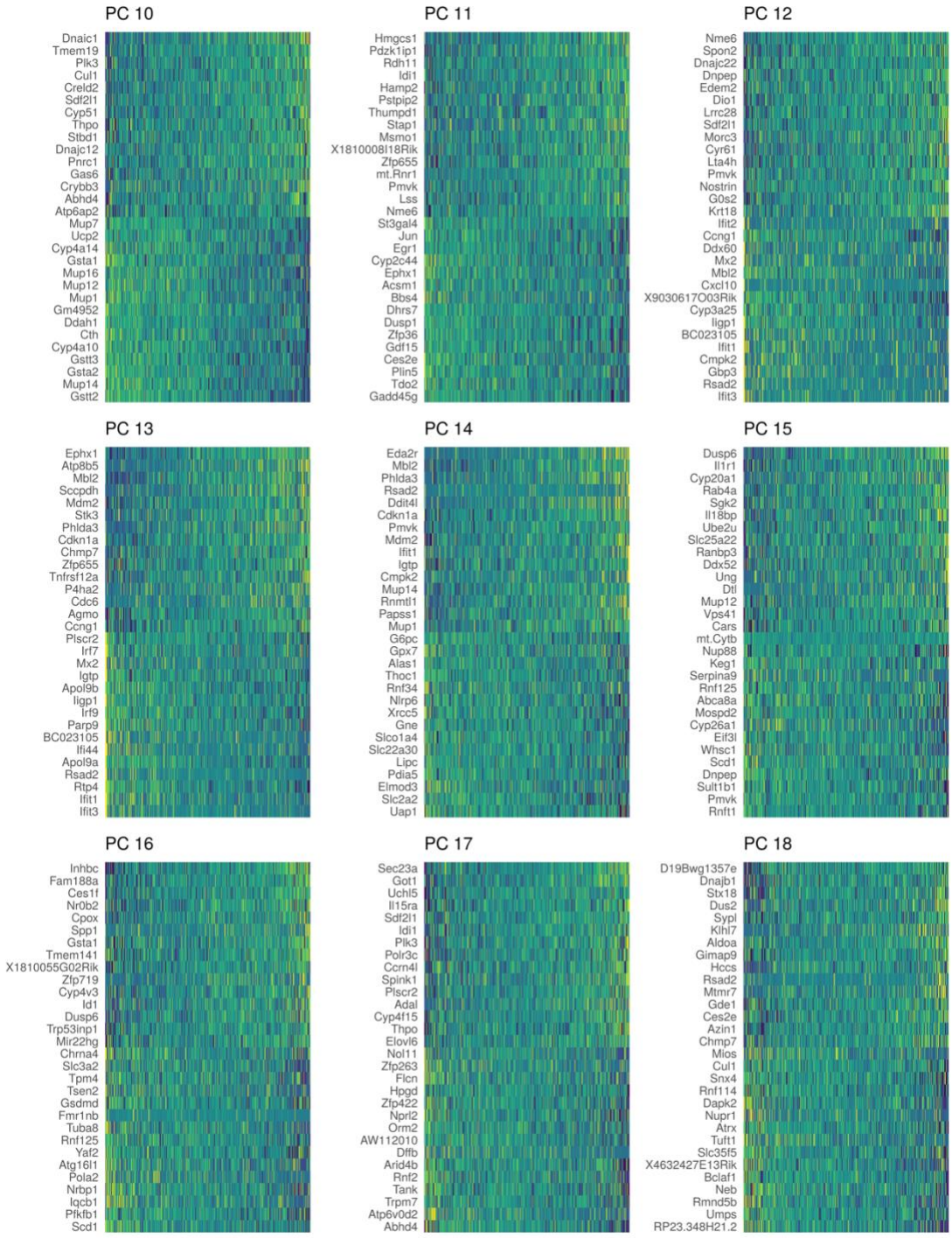
Heatmap of gene scaled expression, where cells and genes are sorted by their principle/harmony component scores. The scaled expression of the top 15 genes with a positive principle/harmony component score and the 15 genes with the biggest negative principle/harmony component score are shown. The top 500 cells are used for the heatmap visualisation. All data shown were generated from scRNA-seq of hepatocytes following a partial hepatectomy time course (Uninjured, 24h, 48h (plate 1), 48h (plate 2), 72h). Data were normalised prior to analysis. Dark blue = Low expression, Yellow = High expression.


PC 1
Pbk
Ube2c
Ckap2l
Birc5
Cdc20
Prc1
Mki67
Tpx2
Nuf2
Ect2
Racgap1
Aurkb
Plk1
Ccnb1
Hmmr
Egfr
Hsd17b2
Ces3a
Hist1h2bc
Adck3
Gm4951
Khk
Ces3b
Tdo2
Cebpa
Slc38a3
Cml1
Mug1
Psen2
Nudt7
PC 2
Sfxn1
Aldh1b1
Sdsl
Gls2
Hsd17b13
Cyp2f2
Sds
Hal
Ctsc
Etnppl
Tstd1
Cdh1
Cryl1
Ctnnbip1
Hpx
Aldh1a1
Akr1c6
Cyp2c40
Avpr1a
Cyp2c50
Cyp2c54
Pon1
Cyp2c38
Slco1b2
Blvrb
Cyp2a5
Rgn
Cyp2c37
Cyp2c29
Cyp2e1
PC 3
Wfdc21
Malat1
Lgals1
Prtn3
Rbp1
Apoc2
Mt2
Hamp2
Mt1
Wbp5
Lpl
Ifi27l2b
Spink1
Adh6.ps1
Fam25c
Tubb2a
Angptl4
Cml5
Acbd4
Crat
Apoa5
Pja1
Acsm1
Slc22a1
Sgk2
Pctp
Cryz
X4833439L19Rik
Tsc22d1
Cyp4a12a
PC 4
Lcn2
Etnk2
Apcs
Saa2
Saa1
Alas1
Cyp4a10
Orm2
Dnajc12
Ccnd1
Nnmt
Mt1
Mt2
Nr1i3
Socs3
Cyp2c70
Ces1d
Hnmt
Ces1e
Acsl5
Ugt2b1
Mup21
Mir22hg
Car3
X1810008I18Rik
Mbl2
Cyp4v3
Inmt
Rnf125
Hyi
PC 5
Pck1
Serpina4.ps1
Jun
Gltp
Gadd45g
Hspb1
Dusp1
Clspn
Mafb
Irf7
Slc38a2
Hamp
Mfsd2a
Trp53inp1
Fabp4
Gsta1
Insig1
Mrap
Smoc2
Gstm3
Idi1
Acsl5
Lss
Hmgcr
Cth
Fdft1
Bhmt
Dnajb9
Msmo1
Hmgcs1
PC 6
Lepr
Crat
Acot1
Aldoa
Tuba1c
Hspb1
X2010003K11Rik
Pif1
Sh2d3c
Anxa5
Spon2
Tmem106a
Cdc42ep5
Cdc25b
Ccnf
Wdhd1
Cdt1
Timeless
Mcm4
Cdc45
Dtl
Pole2
Slfn9
Mcm3
Clspn
Ccne2
Dscc1
Hells
Uhrf1
Cdc6
PC 7
Slc1a2
Fasn
Insig1
Gstm3
Cryl1
Acly
Thrsp
Lss
Hmgcs1
Mvd
Osgin1
Cyp2a5
Cyp4a12b
Sgk2
Cdc6
Hamp
Serpina1e
Cth
Cyp3a13
Dnaic1
Igfbp2
Enho
Pdia4
Wfdc21
Apcs
Itih3
Hsp90b1
Fga
Fgb
Fgg
PC 8
Akr1c14
Fgg
Dhrs7
Slc25a17
Cyp3a25
Gm4952
Hrsp12
Atl2
Abcg2
Blvra
Mup3
Nmt2
Cdadc1
Fam69a
Cyp3a13
Spata2l
Mvk
Slc25a22
Mvd
Zfp36
Hmgcs1
Gale
Slc38a3
Ass1
Jun
Lss
Osgin1
Junb
Etnk2
Ube2s
PC 9
Asl
Cth
Ass1
Got1
Arg1
Hamp
Upp2
Kynu
Enho
Ddc
Nnmt
Uox
Fahd1
Agxt
Mtmr7
Abcb11
Cd9
Cd63
Cp
Acsl5
Slc17a4
Nid1
Ces2c
Vnn3
Vnn1
Mfge8
Hspa5
Slc2a2
Ces2e
Hsp90b1

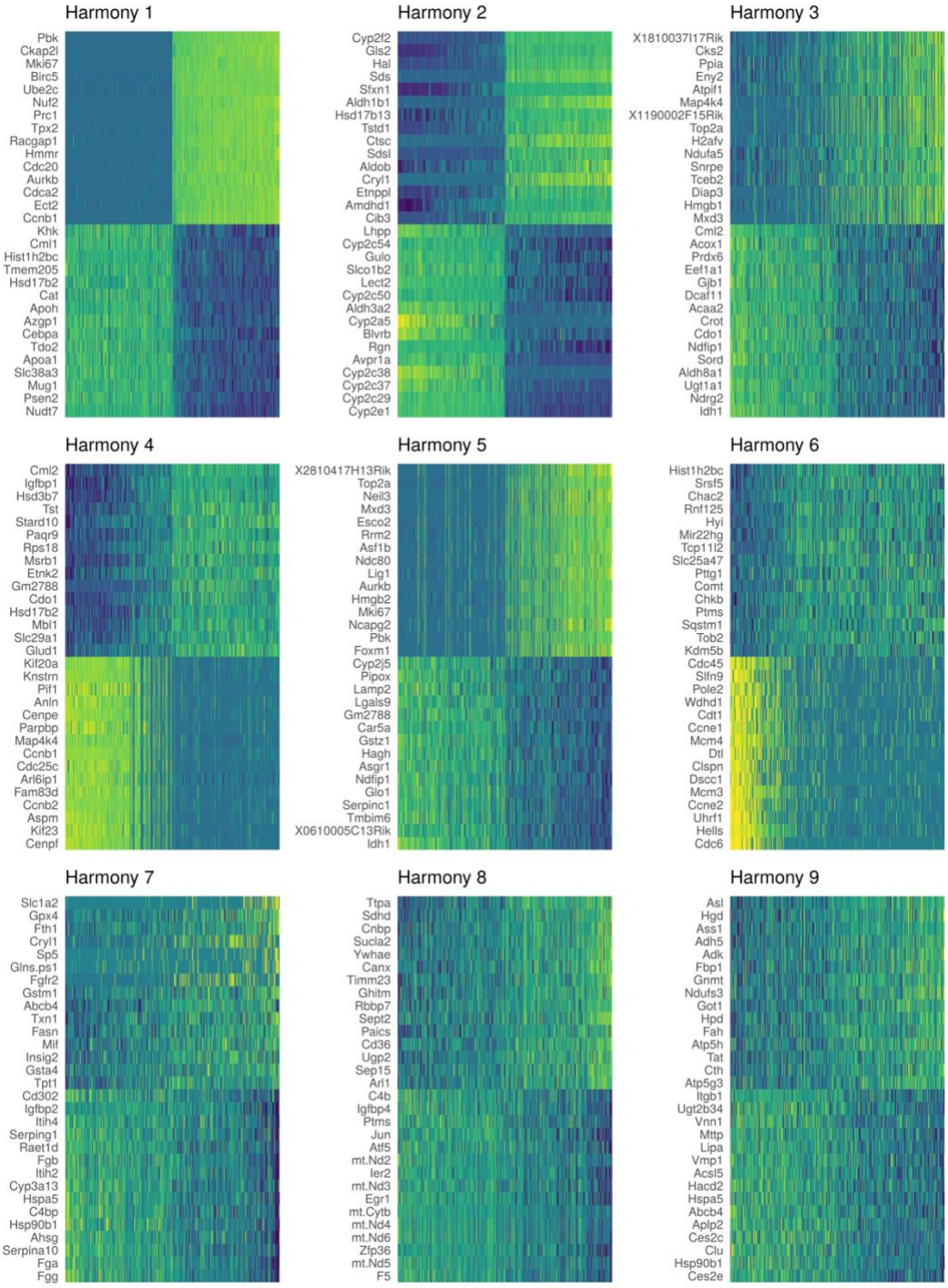


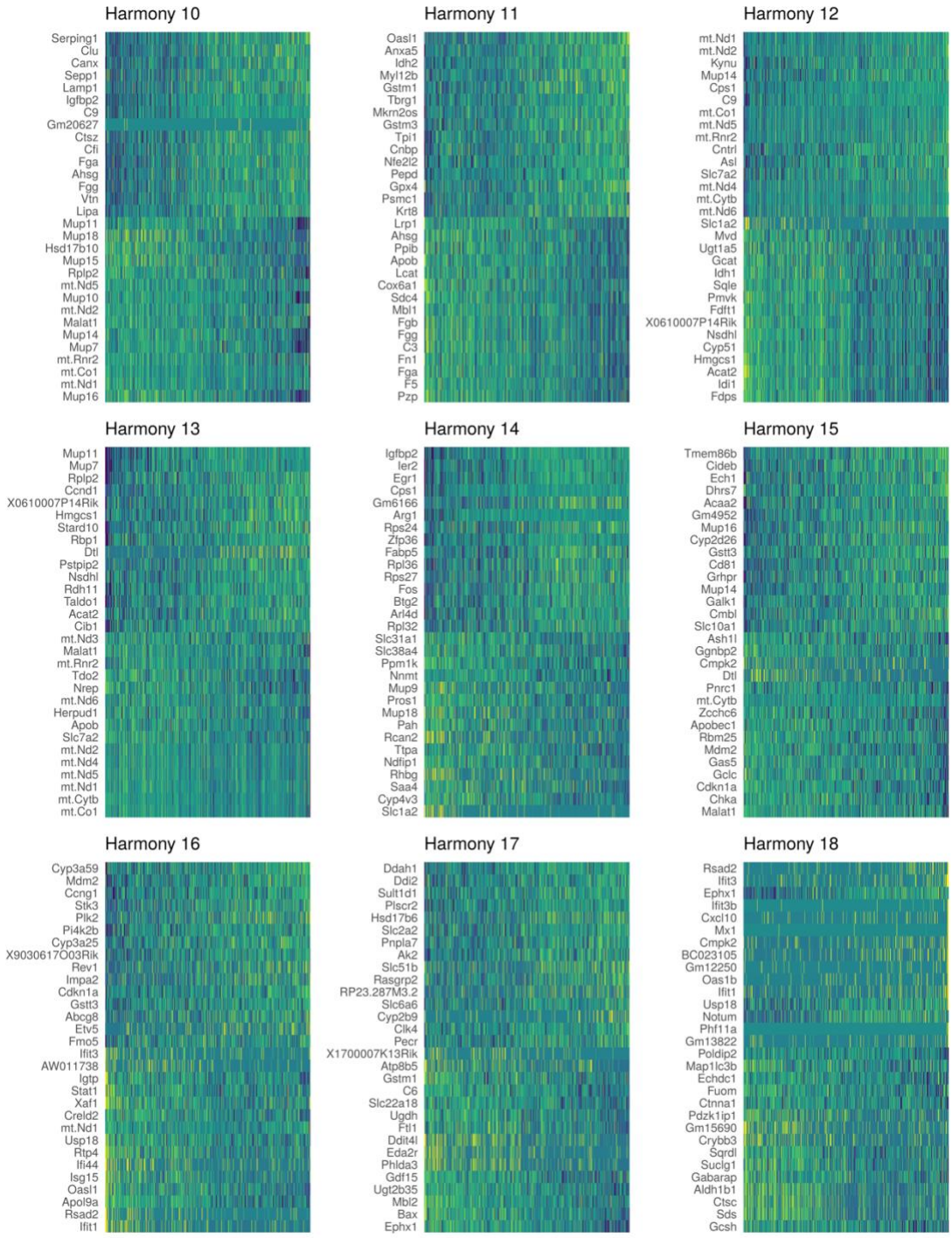
Appendix 1 – Heatmap of Top PCA Genes from Data without Batch Correction


PC 1
Pbk
Hmmr
Birc5
Ckap2l
Ccnb1
Mki67
Nuf2
Gm11223
Prc1
Ube2c
Cdc20
X1190002F15Rik
Aurkb
Tpx2
Ccnb2
Itih3
Pck1
Cml1
Nudt7
Hsd17b2
Uox
Agxt
Lpin1
Fkbp11
Rrbp1
Apoa5
Etnk2
Asl
Slc38a3
Psen2
PC 2
Hsd17b13
Aldh1b1
Cyp2f2
Sds
Hrsp12
Sfxn1
Tstd1
Gls2
Ctsc
Cryl1
Hal
Slc51b
Sdsl
Ctnnbip1
Tm4sf4
Csad
Cyp2a5
Sult2a8
Rnase4
Rgn
Cyp7a1
Gulo
Cyp27a1
Lect2
Pon1
Cyp2c29
Gsta3
Akr1c6
Slco1b2
Cyp2e1
PC 3
Fez2
Hspa5
Ephx1
Mir22hg
Sacm1l
Tmem184c
Rnf14
Mbl2
Vbp1
Cpsf3
Ube2u
C230052I12Rik
Tmem150a
Dnajb9
Fgg
Cdc45
Ccne2
Mcm7
Mcm4
Atad2
Slfn9
Tcf19
Dscc1
Mcm3
Wdhd1
Mcm5
Cdc6
Cispn
Uhrf1
Hells
PC 4
Slc1a2
Gclc
Cyp4a12b
Gstm3
mt.Nd2
Lgr5
Cd36
Ly6d
mt.Rnr2
Gm13775
X1110028F11Rik
Snhg11
Slc13a3
Slc1a5
mt.Nd4
Cyp2c70
Amy1
C8b
Dnaic1
Serpinf2
Arg1
Ass1
Serpina1e
Igfbp2
C9
Hamp
Fgb
Apcs
Fga
Fgg
PC 5
mt.Cytb
mt.Nd4
mt.Nd5
mt.Nd6
mt.Nd2
mt.Co1
mt.Rnr2
mt.Rnr1
Slc38a2
Herpud1
Malat1
Ces1b
Wfdc21
Egfr
Nipal1
Cdc6
Ugt2b37
Ccne1
Uhrf1
Idi1
Uap1l1
Dscr3
Cd36
St3gal4
Bag2
Anxa5
Krt18
Zfp367
Dscc1
Insig2
PC 6
Acsl5
Hsp90b1
Abcb11
Hspa5
Cyp51
Cp
Ces2e
Slc39a4
Degs1
Vnn3
Hmgcr
Lipc
Ces2c
Slc17a4
Malat1
Nt5c3b
Rtcb
Trp53inp1
Rassf7
Spryd4
Ddc
Fam89a
Lama3
Hamp
Fahd1
Enho
Cebpd
Idh2
Zap70
Nnmt
PC 7
Msmo1
Cyp51
Apcs
Hmgcs1
Fgg
Fdft1
Nsdhl
Fgb
Acsl5
Fga
Scd1
Cyp2c70
Serpina1e
Prlr
Rdh11
Nr1h4
Asb13
Usp15
mt.Rnr1
Arid4b
Pepd
Gclc
mt.Rnr2
mt.Nd2
mt.Nd6
mt.Nd4
mt.Cytb
mt.Nd5
mt.Co1
Malat1
PC 8
Saa4
Slc1a2
Saa1
Nnmt
Cyp4v3
Tdo2
Mup16
Gpcpd1
Ugt3a2
Saa2
Cyp2c69
Sds
Mup1
Got1
Fgg
Btg2
Ier2
Scd1
Fasn
Clint1
Hmgcs1
Igfbp2
Cdkn1a
Psph
Gsta1
Atf3
Jun
Tnfrsf12a
Cyr61
Gstm3
PC 9
Oasl1
Selenbp2
Apol9a
Rsad2
C6
Ly6e
Ifit1
Ifit3
Usp18
Cyp8b1
Cyp7a1
AW112010
Ifi44
Apol9b
Igtp
Pi4k2b
Mvd
Cars
Hsd17b6
Chrna4
Insig1
Hmgcs1
Chka
Gstt2
Hmgcr
Osgin1
Ddit3
Fdft1
Tollip
Nr0b2



Appendix 2 – Heatmap of Top PCA Genes from Data with *Seurat Integration* Batch Correction


Harmony 1
Pbk
Ckap2l
Mki67
Birc5
Ube2c
Nuf2
Prc1
Tpx2
Racgap1
Hmmr
Cdc20
Aurkb
Cdca2
Ect2
Ccnb1
Khk
Cml1
Hist1h2bc
Tmem205
Hsd17b2
Cat
Apoh
Azgp1
Cebpa
Tdo2
Apoa1
Slc38a3
Mug1
Psen2
Nudt7
Harmony 2
Cyp2f2
Gls2
Hal
Sds
Sfxn1
Aldh1b1
Hsd17b13
Tstd1
Ctsc
Sdsl
Aldob
Cryl1
Etnppl
Amdhd1
Cib3
Lhpp
Cyp2c54
Gulo
Slco1b2
Lect2
Cyp2c50
Aldh3a2
Cyp2a5
Blvrb
Rgn
Avpr1a
Cyp2c38
Cyp2c37
Cyp2c29
Cyp2e1
Harmony 3
X1810037I17Rik
Cks2
Ppia
Eny2
Atpif1
Map4k4
X1190002F15Rik
Top2a
H2afv
Ndufa5
Snrpe
Tceb2
Diap3
Hmgb1
Mxd3
Cml2
Acox1
Prdx6
Eef1a1
Gjb1
Dcaf11
Acaa2
Crot
Cdo1
Ndfip1
Sord
Aldh8a1
Ugt1a1
Ndrg2
Idh1
Harmony 4
Cml2
Igfbp1
Hsd3b7
Tst
Stard10
Paqr9
Rps18
Msrb1
Etnk2
Gm2788
Cdo1
Hsd17b2
Mbl1
Slc29a1
Glud1
Kif20a
Knstrn
Pif1
Anln
Cenpe
Parpbp
Map4k4
Ccnb1
Cdc25c
Arl6ip1
Fam83d
Ccnb2
Aspm
Kif23
Cenpf
Harmony 5
X2810417H13Rik
Top2a
Neil3
Mxd3
Esco2
Rrm2
Asf1b
Ndc80
Lig1
Aurkb
Hmgb2
Mki67
Ncapg2
Pbk
Foxm1
Cyp2j5
Pipox
Lamp2
Lgals9
Gm2788
Car5a
Gstz1
Hagh
Asgr1
Ndfip1
Glo1
Serpinc1
Tmbim6
X0610005C13Rik
Idh1
Harmony 6
Hist1h2bc
Srsf5
Chac2
Rnf125
Hyi
Mir22hg
Tcp11l2
Slc25a47
Pttg1
Comt
Chkb
Ptms
Sqstm1
Tob2
Kdm5b
Cdc45
Slfn9
Pole2
Wdhd1
Cdt1
Ccne1
Mcm4
Dtl
Clspn
Dscc1
Mcm3
Ccne2
Uhrf1
Hells
Cdc6
Harmony 7
Slc1a2
Gpx4
Fth1
Cryl1
Sp5
Gins.ps1
Fgfr2
Gstm1
Abcb4
Txn1
Fasn
Mif
Insig2
Gsta4
Tpt1
Cd302
Igfbp2
Itih4
Serping1
Raet1d
Fgb
Itih2
Cyp3a13
Hspa5
C4bp
Hsp90b1
Ahsg
Serpina10
Fga
Fgg
Harmony 8
Ttpa
Sdhd
Cnbp
Sucla2
Ywhae
Canx
Timm23
Ghitm
Rbbp7
Sept2
Paics
Cd36
Ugp2
Sep15
Arl1
C4b
Igfbp4
Ptms
Jun
Atf5
mt.Nd2
Ier2
mt.Nd3
Egr1
mt.Cytb
mt.Nd4
mt.Nd6
Zfp36
mt.Nd5
F5
Harmony 9
Asl
Hgd
Ass1
Adh5
Adk
Fbp1
Gnmt
Ndufs3
Got1
Hpd
Fah
Atp5h
Tat
Cth
Atp5g3
Itgb1
Ugt2b34
Vnn1
Mttp
Lipa
Vmp1
Acsl5
Hacd2
Hspa5
Abcb4
Aplp2
Ces2c
Clu
Hsp90b1
Ces2e



Appendix 3 – Heatmap of Top Harmony Genes from Data with *Harmony* Batch Correction

7.2 Differential Gene Analysis

The following contains tables of differentially expressed genes. The logistic regression framework from Seurat was used to identify differentially expressed genes, with batch and replicate used as covariates. Genes with a log-fold change of at least 0.25 and expression in at least 25% of cells in the cluster of interest were used.

Gene	Average Log Fold Change	pct.1	pct.2	Adjusted p value	Cluster
Sds	2.32584309	0.972	0.343	1.59E-102	0
Cyp2f2	1.72678504	1	0.768	4.61E-99	0
Hal	1.80485369	0.995	0.679	2.69E-94	0
Ctsc	1.90247472	0.889	0.233	8.50E-80	0
Cryl1	1.71207277	0.811	0.206	8.78E-64	0
Cib3	1.50653261	0.977	0.675	1.62E-62	0
Hsd17b6	1.78283189	0.834	0.459	1.33E-48	0
Serpina12	1.55044675	0.935	0.711	9.32E-38	0
Ugt2b38	2.11754053	0.641	0.327	1.14E-27	0
Ptgds	1.99606385	0.323	0.113	2.98E-06	0
Enho	0.73191138	0.977	0.806	2.92E-23	1
Cyp8b1	0.78631676	0.994	0.849	6.58E-19	1
Mug.ps1	0.68273133	0.971	0.811	6.11E-13	1
Csad	0.90626026	0.977	0.842	1.63E-11	1
Elovl3	0.70437286	0.953	0.796	7.17E-11	1
Dnaic1	0.76635866	0.643	0.439	2.03E-10	1
Hamp	0.78635702	0.988	0.904	5.43E-09	1
Aqp8	0.65846678	0.994	0.93	7.48E-07	1
Slc17a8	0.72599462	0.637	0.365	0.00054056	1
Hsd3b5	0.70856178	0.819	0.55	0.00072012	1
Cyp2e1	1.77477478	1	0.941	1.62E-87	3
Cyp2c29	1.56321929	1	0.856	1.11E-66	3
Aldh3a2	1.50343251	0.993	0.713	4.07E-60	3
Cyp2c38	1.51652111	0.935	0.391	1.10E-55	3
Cyp2c40	1.83642839	0.863	0.302	3.31E-43	3
Cyp2c37	1.52287677	1	0.705	7.40E-43	3
Tsc22d1	1.5333566	0.964	0.835	2.39E-26	3
Mup18	1.79518122	0.993	0.914	3.22E-26	3
Cyp2c69	1.66901112	0.626	0.165	1.40E-24	3
Mup17	1.87288917	1	0.983	1.23E-22	3
Ccnb1	2.33618712	1	0.232	9.39E-75	2
Cdc20	2.29774518	1	0.308	3.56E-70	2
Ccnb2	2.13103917	1	0.264	4.43E-69	2
Knstrn	2.18712232	1	0.281	6.14E-69	2
Ect2	2.13545491	1	0.295	2.11E-68	2
Prc1	2.1125315	1	0.255	2.88E-67	2
Tpx2	2.15113329	1	0.293	4.11E-66	2
Cenpf	2.22295755	1	0.197	3.71E-65	2
Kif20a	2.1119878	1	0.241	1.02E-64	2

Cenpe	2.13558955	1	0.222	1.60E-63	2
Oat	1.86756382	1	0.812	1.45E-43	4
Esco2	1.62258413	1	0.314	3.88E-43	4
Hmgb2	1.61796079	1	0.696	1.16E-41	4
Mxd3	1.62806894	1	0.237	1.94E-40	4
Top2a	1.5654348	1	0.345	9.92E-40	4
Aurkb	1.51117353	1	0.272	1.35E-37	4
Cdkn2d	1.49196235	1	0.489	5.68E-37	4
Anxa2	1.79131354	0.933	0.346	3.81E-28	4
Glul	1.99995775	1	0.871	7.79E-26	4
Slc1a2	2.2563931	0.375	0.065	3.72E-17	4
Hells	2.40377153	1	0.432	9.09E-58	5
Cdc6	3.02021964	0.887	0.079	1.59E-57	5
Mcm3	2.22970928	0.986	0.341	2.31E-50	5
Mcm4	2.14153961	1	0.365	4.91E-49	5
Ccne2	2.55327499	0.944	0.298	6.20E-48	5
Mcm5	1.92556458	1	0.379	1.41E-45	5
Dtl	2.02281212	0.958	0.197	4.24E-43	5
Cdt1	2.18850455	0.958	0.242	6.81E-43	5
Uhrf1	2.15734737	0.887	0.195	4.37E-41	5
Ccne1	2.3018098	0.93	0.339	3.94E-38	5

Appendix 4 – Top 10 Differential Expressed Genes per Cluster

The top 10 significantly expressed genes per cluster, weighted by average log fold change. Predicted genes, mitochondrial genes and ribosomal genes were removed. This was to prevent uninformative genes being used to assess each cluster. Pct.1 – percentage of cells in the labelled cluster expressing the gene. Pct.2 – Percentage of cells not in the labelled cluster expressing the gene.

Gene	Average Log Fold Change	pct.1	pct.2	Adjusted p value	Cluster
Itga5	1.03262	0.993	0.774	5.96E-32	2
Il33	1.43654533	0.841	0.199	6.13E-29	2
Cklf	0.83304514	0.979	0.373	9.03E-18	2
Il15	0.83312636	0.793	0.234	9.68E-14	2
Ifngr1	0.7331484	0.986	0.708	2.15E-13	2
Raet1e	0.58215256	0.938	0.614	8.46E-09	2
Fam3c	0.42137676	1	0.954	1.15E-08	2
Tnfrsf12a	0.94862122	0.993	0.799	6.35E-07	2
Notch2	0.2779186	0.979	0.833	2.21E-05	2
Ephb4	0.35694469	0.828	0.425	4.53E-05	2
Dll1	0.42735471	0.71	0.249	0.01406146	2
Nrg1	0.50075477	0.469	0.107	0.01950231	2
Cklf	1.22783867	0.99	0.405	3.61E-34	4
Lgr5	1.25006802	0.885	0.226	5.55E-28	4
Avpr1a	0.62342999	0.981	0.658	1.08E-16	4
Itga5	0.68198654	1	0.785	3.42E-13	4
Npr2	0.34896779	0.721	0.28	1.15E-10	4
Eda2r	0.89300297	0.5	0.176	1.04E-06	4
Nrp2	0.51841675	0.346	0.079	1.09E-06	4
Slc1a5	0.7636495	0.558	0.167	1.64E-06	4
Il15	0.67569603	0.75	0.271	2.63E-05	4
Bmp7	0.64207938	0.375	0.121	6.58E-05	4
Igf2r	0.29441997	0.981	0.896	0.00522662	4
Dll1	0.57713867	0.654	0.283	0.02224936	4

Appendix 5 – Differentially expressed Ligand and Receptor Genes for Cluster 2 and 4

All significantly expressed genes for clusters 2 and 4 that were also present in the *CellphoneDB* database. Pct.1 – percentage of cells in the labelled cluster expressing the gene. Pct.2 – Percentage of cells not in the labelled cluster expressing the gene.

7.3 Pathway Analysis

The following contains the table of results obtained from the pathway analysis. All significantly expressed genes (<0.05 adjusted p value) with an average log fold change of >0.5 were used in the pathway analysis. The logistic regression framework from Seurat was used to identify differentially expressed genes, with batch and replicate used as covariates. Only those genes expressed in at least 25% of cells in the cluster of interest were used during the differential expression test. A hypergeometric model in the ReactomePA package was used to perform the enrichment test against the Reactome database, and the p value was adjusted using the Benjamini & Hochberg method. Only the top 8 terms are shown, weighted by adjusted p value. Gene ratio – number of genes in query that overlapped with the term / total genes in query that overlap with all terms. Term ratio – number of genes associated with the term in Reactome database / total number of unique genes associated to all terms in Reactome database.

Appendix 6 – Cluster 0 Pathway Analysis

Term ID	Description	Gene Ratio	Term Ratio	Adjusted pvalue
R-MMU-71291	Metabolism of amino acids and derivatives	14/91	251/8828	7.14E-05

Appendix 7 – Cluster 1 Pathway Analysis

Term ID	Description	Gene Ratio	Term Ratio	Adjusted pvalue
R-MMU-8957322	Metabolism of steroids	5/16	127/8828	9.41E-05
R-MMU-193807	Synthesis of bile acids and bile salts via 27-hydroxycholesterol	3/16	18/8828	9.41E-05
R-MMU-193368	Synthesis of bile acids and bile salts via 7alpha-hydroxycholesterol	3/16	27/8828	0.00016699
R-MMU-211976	Endogenous sterols	3/16	27/8828	0.00016699
R-MMU-192105	Synthesis of bile acids and bile salts	3/16	36/8828	0.00032287
R-MMU-194068	Bile acid and bile salt metabolism	3/16	44/8828	0.00049467
R-MMU-193993	Mineralocorticoid biosynthesis	2/16	11/8228	0.00115055
R-MMU-211897	Cytochrome P450 - arranged by substrate type	3/16	71/8828	0.00155397

Appendix 8 – Cluster 2 Pathway Analysis

Term ID	Description	Gene Ratio	Term Ratio	Adjusted pvalue
R-MMU-68877	Mitotic Prometaphase	93/446	190/8828	1.41E-68
R-MMU-68886	M Phase	120/446	383/8828	1.81E-63
R-MMU-2500257	Resolution of Sister Chromatid Cohesion	65/446	119/8828	2.82E-51
R-MMU-69618	Mitotic Spindle Checkpoint	61/446	108/8828	2.66E-49
R-MMU-141424	Amplification of signal from the kinetochores	54/446	92/8828	7.62E-45
R-MMU-141444	Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	54/446	92/8828	7.62E-45
R-MMU-2555396	Mitotic Metaphase and Anaphase	74/446	194/8828	7.62E-45
R-MMU-68882	Mitotic Anaphase	73/446	193/8828	5.89E-44

Appendix 9 – Cluster 3 Pathway Analysis

Term ID	Description	Gene Ratio	Term Ratio	Adjusted pvalue
R-MMU-211859	Biological oxidations	17/73	203/8828	8.23E-11
R-MMU-211945	Phase I - Functionalization of compounds	13/73	108/8828	3.10E-10
R-MMU-211897	Cytochrome P450 - arranged by substrate type	10/73	71/8828	1.64E-08
R-MMU-8957322	Metabolism of steroids	10/73	127/8828	3.86E-06
R-MMU-193807	Synthesis of bile acids and bile salts via 27-hydroxycholesterol	5/73	18/8828	1.07E-05
R-MMU-194068	Bile acid and bile salt metabolism	6/73	44/8828	4.79E-05
R-MMU-211958	Miscellaneous substrates	5/73	26/8828	5.55E-05
R-MMU-193368	Synthesis of bile acids and bile salts via 7alpha-hydroxycholesterol	5/73	27/8828	5.92E-05

Appendix 10 – Cluster 4 Pathway Analysis

Term ID	Description	Gene Ratio	Term Ratio	Adjusted pvalue
R-MMU-68877	Mitotic Prometaphase	90/381	190/8828	3.56E-71
R-MMU-68886	M Phase	111/381	383/8828	3.75E-62
R-MMU-2500257	Resolution of Sister Chromatid Cohesion	62/381	119/8828	1.70E-51
R-MMU-69620	Cell Cycle Checkpoints	84/381	282/8828	5.51E-47
R-MMU-141424	Amplification of signal from the kinetochores	51/381	92/8828	4.45E-44
R-MMU-141444	Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	51/381	92/8828	4.45E-44
R-MMU-69618	Mitotic Spindle Checkpoint	54/381	108/8828	1.30E-43
R-MMU-2555396	Mitotic Metaphase and Anaphase	68/381	194/8828	6.25E-43

Appendix 11 – Cluster 5 Pathway Analysis

Term ID	Description	Gene Ratio	Term Ratio	Adjusted pvalue
R-MMU-69306	DNA Replication	40/121	123/8828	1.09E-43
R-MMU-69242	S Phase	38/121	142/8828	6.61E-38
R-MMU-69239	Synthesis of DNA	35/121	115/8828	4.70E-37
R-MMU-68962	Activation of the pre-replicative complex	22/121	30/8828	5.08E-34
R-MMU-69190	DNA strand elongation	19/121	22/8828	7.48E-32
R-MMU-5693532	DNA Double-Strand Break Repair	34/121	157/8828	1.28E-30
R-MMU-453279	Mitotic G1-G1/S phases	31/121	126/8828	1.16E-29
R-MMU-73894	DNA Repair	42/121	314/8828	1.68E-29

8 Bibliography

Adrian, R. (2020) *Liver histology: Structure, cells and characteristics*. [Online] [Accessed on 17th April 2020] https://www.kenhub.com/en/library/anatomy/liver-histology.

Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J. S., Pessaux, P., Baumert, T. F. and Grün, D. (2019) 'A human liver cell atlas reveals heterogeneity and epithelial progenitors.' *Nature*, 572(7768), pp. 199–204.

Albrecht, J. H., Hu, M. Y. and Cerra, F. B. (1995) 'Distinct patterns of cyclin D1 regulation in models of liver regeneration and human liver.' *Biochemical and Biophysical Research Communications*, 209(2), pp. 648–655.

Aldeguer, X., Debonera, F., Shaked, A., Krasinkas, A. M., Gelman, A. E., Que, X., Zamir, G. A., Hiroyasu, S., Kovalovich, K. K., Taub, R. and Olthoff, K. M. (2002) 'Interleukin-6 from intrahepatic cells of bone marrow origin is required for normal murine liver regeneration.' *Hepatology*, 35(1), pp. 40–48.

Allen, W. E. (2009) 'Terminologia anatomica: International anatomical terminology and terminologia histologica: International terms for human cytology and histology.' *Journal of Anatomy*, 215(2), pp. 221–221.

Alves-Guerra, M.-C., Capobianco, A. J. and Demarest, R. M. (2011) 'Notch/Jagged signaling.' *In Encyclopedia of Cancer*, pp. 2559–2565.

Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., Waldron, L., Pagès, H., Smith, M. L., Huber, W., Morgan, M., Gottardo, R. and Hicks, S. C. (2020) 'Orchestrating single-cell analysis with Bioconductor.' *Nature Methods*, 17(2), pp. 137–145.

Andrews, T. S. and Hemberg, M. (2018) 'Identifying cell populations with scRNASeq.' *Molecular Aspects of Medicine*, 59, pp. 114–122.

Andrews, T. S. and Hemberg, M. (2019) 'False signals induced by single-cell imputation.' *F1000Research*, 7(1740), pp. 1–35.

Apte, U., Gkretsi, V., Bowen, W. C., Mars, W. M., Luo, J. H., Donthamsetty, S., Orr, A., Monga, S. P. S., Wu, C. and Michalopoulos, G. K. (2009) 'Enhanced liver regeneration following changes induced by hepatocyte-specific genetic ablation of integrin-linked kinase.' *Hepatology*, 50(3), pp. 844–851.

Arias, E. E. and Walter, J. C. (2007) 'Strength in numbers: Preventing rereplication via multiple mechanisms in eukaryotic cells.' *Genes and Development*, 21(5), pp. 497–518.

Asrani, S. K., Devarbhavi, H., Eaton, J. and Kamath, P. S. (2019) 'Burden of liver diseases in the world.' *Journal of Hepatology*, 70(1), pp. 151–171.

Asrani, S. K., Larson, J. J., Yawn, B., Therneau, T. M. and Kim, W. R. (2013) 'Underestimation of liver-related mortality in the United States.' *Gastroenterology*, 145(2), pp. 375–382.

Bacher, R., Chu, L. F., Leng, N., Gasch, A. P., Thomson, J. A., Stewart, R. M., Newton, M. and Kendziorski, C. (2017) 'SCnorm: Robust normalization of single-cell RNA-seq data.' *Nature Methods*, 14(6), pp. 584–586.

Baker, C. (2019) 'Obesity Statistics.' *House of Commons Briefing Paper*, (3336).

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M. and Hamilton, P. W. (2017) 'QuPath: Open source software for digital pathology image analysis.' *Scientific Reports*, 7(1), pp. 1–7.

Barnes, J. W., Tischkau, S. A., Barnes, J. A., Mitchell, J. W., Burgoon, P. W., Hickok, J. R. and Gillette, M. U. (2003) 'Requirement of mammalian timeless for circadian rhythmicity.' *Science*, 302(5644), pp. 439–442.

Bartels, H., Vogt, B. and Jungermann, K. (1987) 'Glycogen synthesis from pyruvate in the periportal and from glucose in the perivenous zone in perfused livers from fasted rats.' *FEBS Letters*, 221(2), pp. 277–283.

Benhamouche, S., Decaens, T., Godard, C., Chambrey, R., Rickman, D. S., Moinard, C., Vasseur-Cognet, M., Kuo, C. J., Kahn, A., Perret, C. and Colnot, S. (2006) 'Apc tumor auppressor gene is the "Zonation-Keeper" of mouse liver.' *Developmental Cell*, 10(6), pp. 759–770.

Berasain, C., García-Trevijano, E. R., Castillo, J., Erroba, E., Lee, D. C., Prieto, J. and Avila, M. A. (2005) 'Amphiregulin: An early trigger of liver regeneration in mice.' *Gastroenterology*, 128(2), pp. 424–432.

Bernal, W. and Wendon, J. (2013) 'Acute liver failure.' *New England Journal of Medicine*, 369(26), pp. 2525–2534.

Bertero, A. and Vallier, L. (2015) 'Fucci2a mouse upgrades live cell cycle imaging.' *Cell Cycle*, 14(7), pp. 948–949.

Bhushan, B. and Apte, U. (2019) 'Liver regeneration after acetaminophen hepatotoxicity: Mechanisms and therapeutic opportunities.' *American Journal of Pathology*, 189(4), pp. 719–729.

Bhushan, B., Walesky, C., Manley, M., Gallagher, T., Borude, P., Edwards, G., Monga, S. P. S. and Apte, U. (2014) 'Pro-regenerative signaling after acetaminophen-induced acute liver injury in mice identified using a novel incremental dose model.' *American Journal of Pathology*, 184(11), pp. 3013–3025.

Biolabs, N. E. (2018) *Increased transcription detection with the NEBNext ® Single Cell/Low Input RNA Library Prep Kit Highly sensitive, robust generation of high quality libraries*.

Borowiak, M., Garratt, A. N., Wüstefeld, T., Strehle, M., Trautwein, C. and Birchmeier, C. (2004) 'Met provides essential signals for liver regeneration.' *Proceedings of the National Academy of Sciences of the United States of America*, 101(29), pp. 10608–10613.

Boyer, J. L. (2013) 'Bile formation and secretion.' *Comprehensive Physiology*, 3(3), pp. 1035–1078.

Bray, S. J. (2016) 'Notch signalling in context.' *Nature Reviews Molecular Cell Biology*, 17, pp. 722–735.

Brosch, M., Kattler, K., Herrmann, A., von Schönfels, W., Nordström, K., Seehofer, D., Damm, G., Becker, T., Zeissig, S., Nehring, S., Reichel, F., Moser, V., Thangapandi, R. V., Stickel, F., Baretton, G., Röcken, C., Muders, M., Matz-Soja, M., Krawczak, M., Gasparoni, G., Hartmann, H., Dahl, A., Schafmayer, C., Walter, J. and Hampe, J. (2018) 'Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control.' *Nature Communications*, 9(1), pp. 1–11.

Brosnan, M. E. and Brosnan, J. T. (2009) 'Hepatic glutamate metabolism: a tale of 2 hepatocytes.' *The American Journal of Clinical Nutrition*, 90(3), pp. 857–861.
Browaeys, R., Saelens, W. and Saeys, Y. (2020) 'NicheNet: modeling intercellular communication by linking ligands to target genes.' *Nature Methods*, 17(2), pp. 159–162.
Bucher, N. L., Patel, U. and Cohen, S. (1977) 'Hormonal factors concerned with liver regeneration.' *Ciba Foundation symposium*, 55, pp. 95–107.
Buermans, H. P. J. and den Dunnen, J. T. (2014) 'Next generation sequencing technology: Advances and applications.' *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1842(10), pp. 1932–1941.
Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C. and Shendure, J. (2019) 'The single-cell transcriptional landscape of mammalian organogenesis.' *Nature*, 566(7745), pp. 496–502.
Carrière, V., Arshad, M. I., Le Seyec, J., Lefevre, B., Farooq, M., Jan, A., Manuel, C., Touami-Bernard, L., Lucas-Clerc, C., Genet, V., Gascan, H., Girard, J.-P., Chalmel, F., Lamontagne, L., Piquet-Pellorce, C. and Samson, M. (2017) 'Endogenous IL-33 deficiency exacerbates liver injury and increases hepatic influx of neutrophils in acute mrine viral hepatitis.' *Mediators of inflammation*, 2017(1359064), pp. 1–15.
Castro, A., Jaumot, M., Vergés, M., Agell, N. and Bachs, O. (1994) 'Microsomal localization of cyclin A and CDK2 in proliferating rat liver cells.' *Biochemical and Biophysical Research Communications*, 201(3), pp. 1072–1078.
Chart, R. S., Price, D. T., Sue, S. R., Meyers, W. C. and Jirtle, R. L. (1995) 'Down-regulation of transforming growth factor beta receptor type I, II, and III during liver regeneration.' *The American Journal of Surgery*, 169(1), pp. 126–132.
Chembazhi, U. V., Bangru, S., Hernaez, M. and Kalsotra, A. (2020) 'Cellular plasticity balances the metabolic and proliferation dynamics of a regenerating liver.' *bioRxiv*, 124263, pp. 1–60.
Chen, G., Ning, B. and Shi, T. (2019) 'Single-cell RNA-seq technologies and related computational data analysis.' *Frontiers in Genetics*, 10(317), pp. 1–13.
Chen, Y., Hata, T., Rehman, F., Kang, L., Yang, L., Kim, B. Y. S. and Nguyen, J. H. (2019) 'Visualization of hepatocellular regeneration in mice after partial hepatectomy.' *Journal of Surgical Research*, 235, pp. 494–500.
Cole, M. B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., Dudoit, S. and Yosef, N. (2019) 'Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq.' *Cell Systems*, 8(4), pp. 315–328.
Collin De L'hortet, A., Ene Gilgenkrantz, H. and Guidotti, J.-E. (2012) 'EGFR: A master piece in G1/S phase transition of liver regeneration.' *International Journal of Hepatology*, 2012(476910), pp. 1–9.
Corlu, A. and Loyer, P. (2012) 'Regulation of the G1/S transition in hepatocytes: Involvement of the cyclin-dependent kinase Cdk1 in the DNA replication.' *International Journal of Hepatology*, 2012, pp. 1–17.
Cossarizza, A., Chang, H. D., Radbruch, A., Akdis, M., Andrä, I., Annunziato, F., Bacher, P., Barnaba, V., Battistini, L., Bauer, W. M., Baumgart, S., Becher, B., Beisker, W., Berek, C., Blanco, A., Borsellino, G., Boulais, P. E., Brinkman, R. R., Büscher, M., Busch, D. H., Bushnell, T. P., Cao, X., Cavani, A.,

Chattopadhyay, P. K., Cheng, Q., Chow, S., Clerici, M., Cooke, A., Cosma, A., Cosmi, L., Cumano, A., Dang, V. D., Davies, D., De Biasi, S., Del Zotto, G., Della Bella, S., Dellabona, P., Deniz, G., Dessing, M., Diefenbach, A., Di Santo, J., Dieli, F., Dolf, A., Donnenberg, V. S., Dörner, T., Ehrhardt, G. R. A., Endl, E., Engel, P., Engelhardt, B., Esser, C., Everts, B., Dreher, A., Falk, C. S., Fehniger, T. A., Filby, A., Fillatreau, S., Follo, M., Förster, I., Foster, J., Foulds, G. A., Frenette, P. S., Galbraith, D., Garbi, N., García-Godoy, M. D., Geginat, J., Ghoreschi, K., Gibellini, L., Goettlinger, C., Goodyear, C. S., Gori, A., Grogan, J., Gross, M., Grützkau, A., Grummitt, D., Hahn, J., Hammer, Q., Hauser, A. E., Haviland, D. L., Hedley, D., Herrera, G., Herrmann, M., Hiepe, F., Holland, T., Hombrink, P., Houston, J. P., Hoyer, B. F., Huang, B., Hunter, C. A., Iannone, A., Jäck, H. M., Jávega, B., Jonjic, S., Juelke, K., Jung, S., Kaiser, T., Kalina, T., Keller, B., Khan, S., Kienhöfer, D., Kroneis, T., Kunkel, D., Kurts, C., Kvistborg, P., Lannigan, J., Lantz, O., Larbi, A., LeibundGut-Landmann, S., Leipold, M. D., Levings, M. K., Litwin, V., Liu, Y., Lohoff, M., Lombardi, G., Lopez, L., Lovett-Racke, A., Lubberts, E., Ludewig, B., Lugli, E., Maecker, H. T., Martrus, G., Matarese, G., Maueröder, C., McGrath, M., McInnes, I., Mei, H. E., Melchers, F., Melzer, S., Mielenz, D., Mills, K., Mirrer, D., Mjösberg, J., Moore, J., Moran, B., Moretta, A., Moretta, L., Mosmann, T. R., Müller, S., Müller, W., Münz, C., Multhoff, G., Munoz, L. E., Murphy, K. M., Nakayama, T., Nasi, M., Neudörfl, C., Nolan, J., Nourshargh, S., O'Connor, J. E., Ouyang, W., Oxenius, A., Palankar, R., Panse, I., Peterson, P., Peth, C., Petriz, J., Philips, D., Pickl, W., Piconese, S., Pinti, M., Pockley, A. G., Podolska, M. J., Pucillo, C., Quataert, S. A., Radstake, T. R. D. J., Rajwa, B., Rebhahn, J. A., Recktenwald, D., Remmerswaal, E. B. M., Rezvani, K., Rico, L. G., Robinson, J. P., Romagnani, C., Rubartelli, A., Ruckert, B., Ruland, J., Sakaguchi, S., Sala-de-Oyanguren, F., Samstag, Y., Sanderson, S., Sawitzki, B., Scheffold, A., Schiemann, M., Schildberg, F., Schimisky, E., Schmid, S. A., Schmitt, S., Schober, K., Schüler, T., Schulz, A. R., Schumacher, T., Scotta, C., Shankey, T. V., Shemer, A., Simon, A. K., Spidlen, J., Stall, A. M., Stark, R., Stehle, C., Stein, M., Steinmetz, T., Stockinger, H., Takahama, Y., Tarnok, A., Tian, Z. G., Toldi, G., Tornack, J., Traggiai, E., Trotter, J., Ulrich, H., van der Braber, M., van Lier, R. A. W., Veldhoen, M., Vento-Asturias, S., Vieira, P., Voehringer, D., Volk, H. D., von Volkmann, K., Waisman, A., Walker, R., Ward, M. D., Warnatz, K., Warth, S., Watson, J. V., Watzl, C., Wegener, L., Wiedemann, A., Wienands, J., Willimsky, G., Wing, J., Wurst, P., Yu, L., Yue, A., Zhang, Q., Zhao, Y., Ziegler, S. and Zimmermann, J. (2017) 'Guidelines for the use of flow cytometry and cell sorting in immunological studies.' *European Journal of Immunology*, 47(10), pp. 1584–1797.

Cressman, D. E., Greenbaum, L. E., DeAngelis, R. A., Ciliberto, G., Furth, E. E., Poli, V. and Taub, R. (1996) 'Liver failure and defective hepatocyte regeneration in interleukin-6- deficient mice.' *Science*, 274(5291), pp. 1379–1383.

Croquelois, A., Blindenbacher, A., Terracciano, L., Wang, X., Langer, I., Radtke, F. and Heim, M. H. (2005) 'Inducible inactivation of Notch1 causes nodular regenerative hyperplasia in mice.' *Hepatology*, 41(3), pp. 487–496.

Cruise, J., Houck, K. and Michalopoulos, G. (1985) 'Induction of DNA synthesis in cultured rat hepatocytes through stimulation of alpha 1 adrenoreceptor by norepinephrine.' *Science*, 227(4688), pp. 749–751.

Cruise, J. L., Knechtle, S. J., Bollinger, R. R., Kuhn, C. and Michalopoulos, G. (1987) 'α 1-Adrenergic effects and liver regeneration.' *Hepatology*, 7(6), pp. 1189–1194.

Ding, B. Sen, Nolan, D. J., Butler, J. M., James, D., Babazadeh, A. O., Rosenwaks, Z., Mittal, V., Kobayashi, H., Shido, K., Lyden, D., Sato, T. N., Rabbany, S. Y. and Rafii, S. (2010) 'Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration.' *Nature*, 468(7321), pp. 310–315.

Ding, J., Adiconis, X., Simmons, S. K., Kowalczyk, M. S., Hession, C. C., Marjanovic, N. D., Hughes, T. K., Wadsworth, M. H., Burks, T., Nguyen, L. T., Kwon, J. Y. H., Barak, B., Ge, W., Kedaigle, A. J., Carroll, S., Li, S., Hacohen, N., Rozenblatt-Rosen, O., Shalek, A. K., Villani, A.-C., Regev, A. and Levin, J. Z. (2019) 'Systematic comparative analysis of single cell RNA-sequencing methods.' *bioRxiv*, pp. 1–61.

Dixon, L. J., Barnes, M., Tang, H., Pritchard, M. T. and Nagy, L. E. (2013) 'Kupffer cells in the liver.' *Comprehensive Physiology*, 3(2), pp. 785–797.

Dobie, R. and Henderson, N. C. (2016) 'Homing in on the hepatic scar: Recent advances in cell-specific targeting of liver fibrosis.' *F1000Research*, 5(1749), pp. 1–9.

Dobie, R., Wilson-Kanamori, J. R., Henderson, B. E. P., Smith, J. R., Matchett, K. P., Portman, J. R., Wallenborg, K., Picelli, S., Zagorska, A., Pendem, S. V., Hudson, T. E., Wu, M. M., Budas, G. R., Breckenridge, D. G., Harrison, E. M., Mole, D. J., Wigmore, S. J., Ramachandran, P., Ponting, C. P., Teichmann, S. A., Marioni, J. C. and Henderson, N. C. (2019) 'Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme during Liver Fibrosis.' *Cell Reports*, 29(7), pp. 1832–1847.

Donne, R., Saroul-Aïnama, M., Cordier, P., Celton-Morizur, S. and Desdouets, C. (2020) 'Polyploidy in liver development, homeostasis and disease.' *Nature Reviews Gastroenterology & Hepatology*, 17, pp. 391–405.

Duncan, A. W., Newell, A. E. H., Bi, W., Finegold, M. J., Olson, S. B., Beaudet, A. L. and Grompe, M. (2012) 'Aneuploidy as a mechanism for stress-induced liver adaptation.' *The Journal of Clinical Investigation*, 122(9), pp. 3307–3315.

Duò, A., Robinson, M. D. and Soneson, C. (2018) 'A systematic performance evaluation of clustering methods for single-cell RNA-seq data.' *F1000Research*, 7(1141), pp. 1–21.

Efremova, M., Vento-Tormo, M., Teichmann, S. A. and Vento-Tormo, R. (2020) 'CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes.' *Nature Protocols*, 15(4), pp. 1484–1506.

Ewels, P. A., Peltzer, A., Fillinger, S., Alneberg, J., Patel, H., Wilm, A., Garcia, M. U., Tommaso, P. Di and Nahnsen, S. (2020) 'The nf-core framework for community-curated bioinformatics pipelines.' *Nature Biotechnology*, 38(3), pp. 276–278.

Fazel Modares, N., Polz, R., Haghighi, F., Lamertz, L., Behnke, K., Zhuang, Y., Kordes, C., Häussinger, D., Sorg, U. R., Pfeffer, K., Floss, D. M., Moll, J. M., Piekorz, R. P., Ahmadian, M. R., Lang, P. A. and Scheller, J. (2019) 'IL-6 Trans-signaling Controls Liver Regeneration After Partial Hepatectomy.' *Hepatology*, 70(6), pp. 2075–2091.

Garnier, D., Loyer, P., Ribault, C., Guguen-Guillouzo, C. and Corlu, A. (2009) 'Cyclin-

dependent kinase 1 plays a critical role in DNA replication control during rat liver regeneration.' *Hepatology*, 50(6), pp. 1946–1956.
Garnier, S. (2018) 'viridis: Default colour maps from "matplotlib."'
Gebhardt, R. and Hovhannisyan, A. (2009) 'Organ patterning in the adult stage: The role of Wnt/β-catenin signaling in liver zonation and beyond.' *Developmental Dynamics*, 239(1), pp. 45–55.
Gebhardt, R. and Matz-Soja, M. (2014) 'Liver zonation: Novel aspects of its regulation and its impact on homeostasis.' *World Journal of Gastroenterology*, 20(26), pp. 8491–8504.
Gebhardt, R. and Mecke, D. (1983) 'Heterogeneous distribution of glutamine synthetase among rat liver parenchymal cells in situ and in primary culture.' *The EMBO Journal*, 2(4), pp. 567–570.
Giacinti, C. and Giordano, A. (2006) 'RB and cell cycle progression.' *Oncogene*, 25(38), pp. 5220–5227.
Giotti, B., Joshi, A. and Freeman, T. C. (2017) 'Meta-analysis reveals conserved cell cycle transcriptional network across multiple human cell types.' *BMC Genomics*, 18(30), pp. 1–12.
Gkretsi, V., Apte, U., Mars, W. M., Bowen, W. C., Luo, J. H., Yang, Y., Yu, Y. P., Orr, A., St.-Arnaud, R., Dedhar, S., Kaestner, K. H., Wu, C. and Michalopoulos, G. K. (2008) 'Liver-specific ablation of integrin-linked kinase in mice results in abnormal histology, enhanced cell proliferation, and hepatomegaly.' *Hepatology*, 48(6), pp. 1932–1941.
Glaser, S. S., Gaudio, E., Miller, T., Alvaro, D. and Alpini, G. (2009) 'Cholangiocyte proliferation and liver fibrosis.' *Expert Reviews in Molecular Medicine*, 11, pp. 1–22.
Goh, Y. P. S., Henderson, N. C., Heredia, J. E., Eagle, A. R., Odegaard, J. I., Lehwald, N., Nguyen, K. D., Sheppard, D., Mukundan, L., Locksley, R. M., Chawla, A. and Mangelsdorf, D. J. (2013) 'Eosinophils secrete IL-4 to facilitate liver regeneration.' *Proceedings of the National Academy of Sciences*, 110(24), pp. 9914–9919.
Greenhalgh, S. N., Matchett, K. P., Taylor, R. S., Huang, K., Li, J. T., Saeteurn, K., Donnelly, M. C., Simpson, E. E. M., Pollack, J. L., Atakilit, A., Simpson, K. J., Maher, J. J., Iredale, J. P., Sheppard, D. and Henderson, N. C. (2019) 'Loss of integrin αvβ8 in murine hepatocytes accelerates liver regeneration.' *American Journal of Pathology*, 189(2), pp. 258–271.
Gross, I. C. on V. (2005) *Nomina anatomica veterinaria*.
Haghverdi, L., Lun, A. T. L., Morgan, M. D. and Marioni, J. C. (2018) 'Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors.' *Nature Biotechnology*, 36(5), pp. 421–427.
Halpern, K. B., Shenhav, R., Massalha, H., Toth, B., Egozi, A., Massasa, E. E., Medgalia, C., David, E., Giladi, A., Moor, A. E., Porat, Z., Amit, I. and Itzkovitz, S. (2018) 'Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells.' *Nature Biotechnology*, 36(10), pp. 1–29.
Halpern, K. B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., Massasa, E. E., Baydatch, S., Landen, S., Moor, A. E., Brandis, A., Giladi, A., Stokar-Avihail, A., David, E., Amit, I. and Itzkovitz, S. (2017) 'Single-cell spatial reconstruction reveals global division of labour in the mammalian liver.' *Nature*, 16(542), pp. 352–356.

Han, W., Lou, Y., Tang, J., Zhang, Y., Chen, Y., Li, Y., Gu, W., Huang, J., Gui, L., Tang, Y., Li, F., Song, Q., Di, C., Wang, L., Shi, Q., Sun, R., Xia, D., Rui, M., Tang, J. and Ma, D. (2001) 'Molecular cloning and characterization of chemokine-like factor 1 (CKLF1), a novel human cytokine with unique structure and potential chemotactic activity.' *The Biochemical journal*, 357(1), pp. 127–135.

Hanawa, N., Shinohara, M., Saberi, B., Gaarde, W. A., Han, D. and Kaplowitz, N. (2008) 'Role of JNK translocation to mitochondria leading to inhibition of mitochondria bioenergetics in acetaminophen-induced liver injury.' *Journal of Biological Chemistry*, 283(20), pp. 13565–13577.

Haque, A., Engel, J., Teichmann, S. A. and Lönnberg, T. (2017) 'A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.' *Genome Medicine*, 9(1), pp. 1–12.

Hempel, M., Schmitz, A., Winkler, S., Kucukoglu, O., Brückner, S., Niessen, C. and Christ, B. (2015) 'Pathological implications of cadherin zonation in mouse liver.' *Cellular and Molecular Life Sciences*, 72(13), pp. 2599–2612.

Henderson, N. C., Pollock, K. J., Frew, J., Mackinnon, A. C., Flavell, R. A., Davis, R. J., Sethi, T. and Simpson, K. J. (2007) 'Critical role of c-jun (NH2) terminal kinase in paracetamol-induced acute liver failure.' *Gut*, 56(7), pp. 982–990.

Hicks, S. C., Townes, F. W., Teng, M. and Irizarry, R. A. (2017) 'Missing data and technical variability in single-cell RNA-sequencing experiments.' *Biostatistics*, 19(4), pp. 562–578.

Higgins, G. M. and Anderson, R. M. (1931) 'Experimental pathology of the livre I. Restoration of the liver of the white rate following partial surgical removal.' *Archives of Pathology & Laboratory Medicine*, 12, pp. 186–202.

Horner, R., Gassner, J. G. M. V., Kluge, M., Tang, P., Lippert, S., Hillebrandt, K. H., Moosburner, S., Reutzel-Selke, A., Pratschke, J., Sauer, I. M. and Raschzok, N. (2019) 'Impact of percoll purification on isolation of primary human hepatocytes.' *Scientific Reports*, 9(1), pp. 1–10.

Houck, K. A., Cruise, J. L. and Michalopoulos, G. (1988) 'Norepinephrine modulates the growth-inhibitory effect of transforming growth factor-beta in primary rat hepatocyte cultures.' *Journal of Cellular Physiology*, 135(3), pp. 551–555.

Houck, K. A. and Michalopoulos, G. K. (1989) 'Altered responses of regenerating hepatocytes to norepinephrine and transforming growth factor type β.' *Journal of Cellular Physiology*, 141(3), pp. 503–509.

Hu, P., Zhang, W., Xin, H. and Deng, G. (2016) 'Single cell isolation and analysis.' *Frontiers in Cell and Developmental Biology*, 4(116), pp. 1–12.

Hu, V. W., Black, G. E., Torres-Duarte, A. and Abramson, F. P. (2002) '3H-thymidine is a defective tool with which to measure rates of DNA synthesis.' *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 16(11), pp. 1456–1457.

Hu, W., Nevzorova, Y. A., Haas, U., Moro, N., Sicinski, P., Geng, Y., Barbacid, M., Trautwein, C. and Liedtke, C. (2014) 'Concurrent deletion of cyclin E1 and cyclin-dependent kinase 2 in hepatocytes inhibits DNA replication and liver regeneration in mice.' *Hepatology*, 59(2), pp. 651–660.

Huang, J., Liang, X., Xuan, Y., Geng, C., Li, Y., Lu, H., Qu, S., Mei, X., Chen, H., Yu, T., Sun, N., Rao, J., Wang, J., Zhang, W., Chen, Y., Liao, S., Jiang, H., Liu, X., Yang, Z., Mu, F. and Gao, S. (2017) 'Clustering trees: a visualisation for

evaluating clusterings at multiple resolutions.' *GigaScience*, 6(5), pp. 1–18.
Huang, W., Ma, K., Zhang, J., Qatanani, M., Cuvillier, J., Liu, J., Dong, B., Huang, X. and Moore, D. D. (2006) 'Nuclear receptor-dependent bile acid signaling is required for normal liver regeneration.' *Science*, 312(5771), pp. 233–236.
Hwang, B., Lee, J. H. and Bang, D. (2018) 'Single-cell RNA sequencing technologies and bioinformatics pipelines.' *Experimental and Molecular Medicine*, 50(8), pp. 1–14.
Ichikawa, T., Zhang, Y. Q., Kogure, K., Hasegawa, Y., Takagi, H., Mori, M. and Kojima, I. (2001) 'Transforming growth factor β and activin tonically inhibit DNA synthesis in the rat liver.' *Hepatology*, 34(5), pp. 918–925.
Ikeda, H., Nagoshi, S., Ohno, A., Yanase, M., Maekawa, H. and Fujiwara, K. (1998) 'Activated rat stellate cells express c-met and respond to hepatocyte growth factor to enhance transforming growth factor β1 expression and DNA synthesis.' *Biochemical and Biophysical Research Communications*, 250(3), pp. 769–775.
Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C. and Teichmann, S. A. (2016) 'Classification of low quality cells from single-cell RNA-seq data.' *Genome Biology*, 17(29), pp. 1–15.
Ishibashi, H., Nakamura, M., Komori, A., Migita, K. and Shimoda, S. (2009) 'Liver architecture, cell function, and disease.' *Seminars in Immunopathology*, 31(3), pp. 399–409.
Jaeschke, H., McGill, M. R. and Ramachandran, A. (2012) 'Oxidant stress, mitochondria, and cell death mechanisms in drug-induced liver injury: Lessons learned from acetaminophen hepatotoxicity.' *Drug Metabolism Reviews*, 44(1), pp. 88–106.
Jaeschke, H., Xie, Y. and McGill, M. R. (2014) 'Acetaminophen-induced liver injury: From animal models to humans.' *Journal of Clinical and Translational Hepatology*, 2(3), pp. 153–161.
Jakowlew, S. B., Mead, J. E., Danielpour, D., Wu, J., Roberts, A. B. and Fausto, N. (1991) 'Transforming growth factor-β (TGF-β) isoforms in rat liver regeneration: Messenger RNA expression and activation of latent TGF-β.' *Molecular Biology of the Cell*, 2(7), pp. 535–548.
Jaumot, M., Estanyol, J.-M., Serratosa, J., Agell, N. and Bachs, O. (1999) 'Activation of Cdk4 and Cdk2 during rat liver regeneration is associated with intranuclear rearrangements of cyclin-Cdk complexes.' *Hepatology*, 29(2), pp. 385–395.
Jiao, J., Ooka, K., Fey, H., Fiel, M. I., Rahmman, A. H., Kojima, K., Hoshida, Y., Chen, X., de Paula, T., Vetter, D., Sastre, D., Lee, K. H., Lee, Y., Bansal, M., Friedman, S. L., Merad, M. and Aloman, C. (2016) 'Interleukin-15 receptor α on hepatic stellate cells regulates hepatic fibrogenesis in mice.' *Journal of hepatology*, 65(2), pp. 344–353.
Jo, M., Stolz, D. B., Esplen, J. E., Dorko, K., Michalopoulos, G. K. and Strom, S. C. (2000) 'Cross-talk between epidermal growth factor receptor and c-Met signal pathways in transformed cells.' *Journal of Biological Chemistry*, 275(12), pp. 8806–8811.
Johansson, M. W., Mccarty, O., Fulkerson, P. C., Bryce, P. J. and Johnston, L. K. (2017) 'Understanding interleukin 33 and its roles in eosinophil development.' *Frontiers in Medicine*, 4(51), pp. 1–7.
Jungermann, K. and Keitzmann, T. (1996) 'Zonation of Parenchymal and

Nonparenchymal Metabolism in Liver.' *Annual Review of Nutrition*, 16(1), pp. 179–203.
Jungermann, K. and Kietzmann, T. (1997) 'Role of oxygen in the zonation of carbohydrate metabolism and gene expression in liver.' *Kidney international*, 51(2), pp. 402–412.
Katoonizadeh, A. (2017) 'Liver regeneration.' *In Liver Pathophysiology: Therapies and Antioxidants*, pp. 113–123.
Kaylan, K. B., Berg, I. C., Biehl, M. J., Brougham-Cook, A., Jain, I., Jamil, S. M., Sargeant, L. H., Cornell, N. J., Raetzman, L. T. and Underhill, G. H. (2018) 'Spatial patterning of liver progenitor cell differentiation me diated by cellular contractility and Notch signaling.' *eLife*, 7, pp. 1–23.
Kharchenko, P. V., Silberstein, L. and Scadden, D. T. (2014) 'Bayesian approach to single-cell differential expression analysis.' *Nature Methods*, 11(7), pp. 740–742.
Kietzmann, T. (2017) 'Metabolic zonation of the liver: The oxygen gradient revisited.' *Redox Biology*, 11, pp. 622–630.
Kim, T., Mars, W. M., Stolz, D. B., Petersen, B. E. and Michalopoulos, G. K. (1997) 'Extracellular matrix remodeling at the early stages of liver regeneration in the rat.' *Hepatology*, 26(4), pp. 896–904.
Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R. and Hemberg, M. (2017) 'SC3: Consensus clustering of single-cell RNA-seq data.' *Nature Methods*, 14(5), pp. 483–486.
Kiso, S., Kawata, S., Tamura, S., Higashiyama, S., Ito, N., Tsushima, H., Taniguchi, N. and Matsuzawa, Y. (1995) 'Role of heparin-binding epidermal growth factor–like growth factor as a hepatotrophic factor in rat liver regeneration after partial hepatectomy.' *Hepatology*, 22(5), pp. 1584–1590.
Kiso, S., Kawata, S., Tamura, S., Inui, Y., Yoshida, Y., Sawai, Y., Umeki, S., Ito, N., Yamada, A., Miyagawa, J. I., Higashiyama, S., Iwawaki, T., Saito, M., Taniguchi, N., Matsuzawa, Y. and Kohno, K. (2003) 'Liver regeneration in heparin-binding EGF-like growth factor transgenic mice after partial hepatectomy.' *Gastroenterology*, 124(3), pp. 701–707.
Köhler, C., Bell, A. W., Bowen, W. C., Monga, S. P., Fleig, W. and Michalopoulos, G. K. (2004) 'Expression of Notch-1 and its Ligand Jagged-1 in Rat Liver during Liver Regeneration.' *Hepatology*, 39(4), pp. 1056–1065.
Kono, S., Nagaike, M., Matsumoto, K. and Nakamura, T. (1992) 'Marked induction of hepatocyte growth factor mRNA in intact kidney and spleen in response to injury of distant organs.' *Biochemical and biophysical research communications*, 186(2), pp. 991–998.
Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru and Raychaudhuri, S. (2019) 'Fast, sensitive and accurate integration of single-cell data with Harmony.' *Nature Methods*, 16(12), pp. 1289–1296.
Kron, P., Linecker, M., Limani, P., Schlegel, A., Kambakamba, P., Lehn, J. M., Nicolau, C., Graf, R., Humar, B. and Clavien, P. A. (2016) 'Hypoxia-driven Hif2a coordinates mouse liver regeneration by coupling parenchymal growth to vascular expansion.' *Hepatology*, 64(6), pp. 2198–2209.
Kruepunga, N., Hakvoort, T. B. M., Hikspoors, J. P. J. M., Köhler, S. E. and Lamers,

W. H. (2019) ‘Anatomy of rodent and human livers: What are the differences?’ *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1865(5), pp. 869–878.

L. Lun, A. T., Bach, K. and Marioni, J. C. (2016) ‘Pooling across cells to normalize single-cell RNA sequencing data with many zero counts.’ *Genome Biology*, 17(1), p. 75.

LaMarre, J., Hayes, M. A., Wollenberg, G. K., Hussaini, I., Hall, S. W. and Gonias, S. L. (1991) ‘An α2-macroglobulin receptor-dependent mechanism for the plasma clearance of transforming growth factor-β1 in mice.’ *Journal of Clinical Investigation*, 87(1), pp. 39–44.

LaMarre, J., Wollenberg, G. K., Gauldie, J. and Hayes, M. A. (1990) ‘Alpha 2-macroglobulin and serum preferentially counteract the mitoinhibitory effect of transforming growth factor-beta 2 in rat hepatocytes.’ *Laboratory investigation; a journal of technical methods and pathology*, 62(5), pp. 545–51.

Lambotte, L., Saliez, A., Triest, S., Maiter, D., Baranski, A., Barker, A. and Li, B. (1997) ‘Effect of sialoadenectomy and epidermal growth factor administration on liver regeneration after partial hepatectomy.’ *Hepatology*, 25(3), pp. 607–612.

Le, D. C., Sunnarborg, S. W., Hinkle, C. L., Myers, T. J., Stevenson, M., Russell, W. E., Castner, B. J., Gerhart, M. J., Paxton, R. J., Black, R. A., Chang, A. and Jackson, L. F. (2003) ‘TACE/ADAM17 processing of EGFR ligands indicates a role as a physiological convertase.’ *Annals of the New York Academy of Sciences*, 995(1), pp. 22–38.

LeCouter, J., Moritz, D. R., Li, B., Phillips, G. L., Liang, X. H., Gerber, H. P., Hillan, K. J. and Ferraral, N. (2003) ‘Angiogenesis-independent endothelial protection of liver: Role of VEGFR-1.’ *Science*, 299(5608), pp. 890–893.

Lesurtel, M., Graf, R., Aleil, B., Walther, D. J., Tian, Y., Jochum, W., Gachet, C., Bader, M. and Clavien, P. A. (2006) ‘Platelet-derived serotonin mediates liver regeneration.’ *Science*, 312(5770), pp. 104–107.

Li, L., Krantz, I. D., Deng, Y., Genin, A., Banta, A. B., Collins, C. C., Qi, M., Trask, B. J., Kuo, W. L., Cochran, J., Costa, T., Pierpont, M. E. M., Rand, E. B., Piccoli, D. A., Hood, L. and Spinner, N. B. (1997) ‘Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for notch1.’ *Nature Genetics*, 16(3), pp. 243–251.

Li, W., Liang, X., Leu, J. I., Kovalovich, K., Ciliberto, G. and Taub, R. (2001) ‘Global changes in interleukin-6-dependent gene expression patterns in mouse livers after partial hepatectomy.’ *Hepatology*, 33(6), pp. 1377–1386.

Lindroos, P. M., Zarnegar, R. and Michalopoulos, G. K. (1991) ‘Hepatocyte growth factor (hepatopoietin A) rapidly increases in plasma before DNA synthesis and liver regeneration stimulated by partial hepatectomy and carbon tetrachloride administration.’ *Hepatology*, 13(4), pp. 743–750.

Lindros, K. O. (1997) ‘Zonation of cytochrome P450 expression, drug metabolism and toxicity in liver.’ *General Pharmacology*, 28(2), pp. 191–196.

Liu, M. and Chen, P. (2017) ‘Proliferation-inhibiting pathways in liver regeneration.’ *Molecular Medicine Reports*, 16(1), pp. 23–35.

Liu, Y., Chen, S., Wang, S., Soares, F., Fischer, M., Meng, F., Du, Z., Lin, C., Meyer, C., DeCaprio, J. A., Brown, M., Liu, X. S. and He, H. H. (2017) ‘Transcriptional landscape of the human cell cycle.’ *Proceedings of the National Academy of*

Sciences of the United States of America, 114(13), pp. 3473–3478.
López-Luque, J., Caballero-Díaz, D., Martinez-Palacián, A., Roncero, C., Moreno-Càceres, J., García-Bravo, M., Grueso, E., Fernández, A., Crosas-Molist, E., García-Álvaro, M., Addante, A., Bertran, E., Valverde, A. M., González-Rodríguez, Á., Herrera, B., Montoliu, L., Serrano, T., Segovia, J. C., Fernández, M., Ramos, E., Sánchez, A. and Fabregat, I. (2016) 'Dissecting the role of epidermal growth factor receptor catalytic activity during liver regeneration and hepatocarcinogenesis.' *Hepatology*, 63(2), pp. 604–619.
Love, M. I., Anders, S., Kim, V. and Huber, W. (2015) 'RNA-Seq workflow: gene-level exploratory analysis and differential expression.' *F1000Research*, 4(1070), pp. 1–41.
Lu, J., Zhou, Y., Hu, T., Zhang, Hui, Shen, M., Cheng, P., Dai, W., Wang, F., Chen, K., Zhang, Y., Wang, C., Li, J., Zheng, Y., Yang, J., Zhu, R., Wang, Jianrong, Lu, W., Zhang, Huawei, Wang, Junshan, Xia, Y., De Assuncao, T. M., Jalan-Sakrikar, N., Huebert, R. C., Zhou, B. and Guo, C. (2016) 'Notch signaling coordinates progenitor cell-mediated biliary regeneration following partial hepatectomy.' *Scientific reports*, 6(22754), pp. 1–16.
Lu, W. Y., Bird, T. G., Boulter, L., Tsuchiya, A., Cole, A. M., Hay, T., Guest, R. V., Wojtacha, D., Man, T. Y., Mackinnon, A., Ridgway, R. A., Kendall, T., Williams, M. J., Jamieson, T., Raven, A., Hay, D. C., Iredale, J. P., Clarke, A. R., Sansom, O. J. and Forbes, S. J. (2015) 'Hepatic progenitor cells of biliary origin with liver repopulation capacity.' *Nature Cell Biology*, 17(8), pp. 973–983.
Luecken, M. D. and Theis, F. J. (2019) 'Current best practices in single-cell RNA-seq analysis: a tutorial.' *Molecular Systems Biology*, 15(6), pp. 1–23.
Lun, A. T. L., Calero-Nieto, F. J., Haim-Vilmovsky, L., Göttgens, B. and Marioni, J. C. (2017) 'Assessing the reliability of spike-in normalization for analyses of single-cell RNA sequencing data.' *Genome Research*, 27(11), pp. 1795–1806.
Lun, A. T. L. and Marioni, J. C. (2017) 'Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data.' *Biostatistics*, 18(3), pp. 451–464.
Lun, A. T. L., McCarthy, D. J. and Marioni, J. C. (2016) 'A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor.' *F1000Research*, 5(2122), pp. 1–64.
Macdonald, R. A. (1961) '"Lifespan" of liver cells: Autoradiographic study using tritiated thymidine in normal, cirrhotic, and partially hepatectomized rats.' *Archives of Internal Medicine*, 107(3), pp. 335–343.
MacParland, S. A., Liu, J. C., Ma, X. Z., Innes, B. T., Bartczak, A. M., Gage, B. K., Manuel, J., Khuu, N., Echeverri, J., Linares, I., Gupta, R., Cheng, M. L., Liu, L. Y., Camat, D., Chung, S. W., Seliga, R. K., Shao, Z., Lee, E., Ogawa, S., Ogawa, M., Wilson, M. D., Fish, J. E., Selzner, M., Ghanekar, A., Grant, D., Greig, P., Sapisochin, G., Selzner, N., Winegarden, N., Adeyi, O., Keller, G., Bader, G. D. and McGilvray, I. D. (2018) 'Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations.' *Nature Communications*, 9(1), pp. 1–21.
Maroni, L., Haibo, B., Ray, D., Zhou, T., Wan, Y., Meng, F., Marzioni, M. and Alpini, G. (2015) 'Functional and structural features of cholangiocytes in health and disease.' *CMGH*, 1(4), pp. 368–380.
Mars, W. M., Liu, M.-L., Kitson, R. P., Goldfarb, R. H., Gabauer, M. K. and

Michalopoulos, G. K. (1995) 'Immediate early detection of urokinase receptor after partial hepatectomy and its implications for initiation of liver regeneration.' *Hepatology*, 21(6), pp. 1695–1701.
Mars, W. M., Zarnegar, R. and Michalopoulos, G. K. (1993) 'Activation of hepatocyte growth factor by the plasminogen activators uPA and tPA.' *The American journal of pathology*, 143(3), pp. 949–58.
Marshall, L., Finch, D., Cairncross, L. and Bibby, J. (2019) *Mortality and life expectancy trends in the UK: stalling progress*.
Marubashi, S., Sakon, M., Nagano, H., Gotoh, K., Hashimoto, K., Kubota, M., Kobayashi, S., Yamamoto, S., Miyamoto, A., Dono, K., Nakamori, S., Umeshita, K. and Monden, M. (2004) 'Effect of portal hemodynamics on liver regeneration studied in a novel portohepatic shunt rat model.' *Surgery*, 136(5), pp. 1028–1037.
Matsuo, T., Yamaguchi, S., Mitsui, S., Emi, A., Shimoda, F. and Okamura, H. (2003) 'Control mechanism of the circadian clock for timing of cell division in vivo.' *Science*, 302(5643), pp. 255–259.
McCarthy, D. J., Campbell, K. R., Lun, A. T. L. and Wills, Q. F. (2017) 'Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.' *Bioinformatics*, 33(8), pp. 1179–1186.
McGarry, T. J. and Kirschner, M. W. (1998) 'Geminin, an inhibitor of DNA replication, is degraded during mitosis.' *Cell*, 93(6), pp. 1043–1053.
Mead, J. E. and Fausto, N. (1989) 'Transforming growth factor α may be a physiological regulator of liver regeneration by means of an autocrine mechanism.' *Proceedings of the National Academy of Sciences of the United States of America*, 86(5), pp. 1558–1562.
Mehendale, H. M. (2005) 'Tissue repair: An important determinant of final outcome of toxicant-induced injury.' *Toxicologic Pathology*, 33(1), pp. 41–51.
Metrich, M., Bezdek Pomey, A., Berthonneche, C., Sarre, A., Nemir, M. and Pedrazzini, T. (2015) 'Jagged1 intracellular domain-mediated inhibition of Notch1 signalling regulates cardiac homeostasis in the postnatal heart.' *Cardiovascular Research*, 108(1), pp. 74–86.
Michalopoulos, G. K. (2007) 'Liver regeneration.' *Journal of Cellular Physiology*, 213(2), pp. 286–300.
Michalopoulos, G. K. (2010) 'Liver regeneration after partial hepatectomy.' *The American Journal of Pathology*, 176(1), pp. 2–13.
Michalopoulos, G. K. (2017) 'Hepatostat: Liver regeneration and normal liver tissue maintenance.' *Hepatology*, 65(4), pp. 1384–1392.
Miller, A. M. (2011) 'Role of IL-33 in inflammation and disease.' *Journal of Inflammation*, 8(22), pp. 1–12.
Mitchell, C., Nivison, M., Jackson, L. F., Fox, R., Lee, D. C., Campbell, J. S. and Fausto, N. (2005) 'Heparin-binding epidermal growth factor-like growth factor links hepatocyte priming with cell cycle progression during liver regeneration.' *Journal of Biological Chemistry*, 280(4), pp. 2562–2568.
Mitchell, C. and Willenbring, H. (2008) 'A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice.' *Nature Protocols*, 3(7), pp. 1167–1170.
Mitchell, J. R., Jollow, D. J., Potter, W. Z., Gillette, J. R. and Brodie, B. B. (1973) 'Acetaminophen-induced hepatic necrosis. IV. Protective role of glutathione.' *The Journal of Pharmacology and Experimental Therapeutics*, 187(1), pp.

211–217.
Miyaoka, Y., Ebato, K., Kato, H., Arakawa, S., Shimizu, S. and Miyajima, A. (2012) 'Hypertrophy and Unconventional Cell Division of Hepatocytes Underlie Liver Regeneration.' *Current Biology*, 22(13), pp. 1166–1175.
Miyaoka, Y. and Miyajima, A. (2013) 'To divide or not to divide: revisiting liver regeneration.' *Cell Division*, 8(8), pp. 1–12.
Monga, S. P. S., Pediaditakis, P., Mule, K., Stolz, D. B. and Michalopoulos, G. K. (2001) 'Changes in wnt/β-catenin pathway during regulated growth in rat liver regeneration.' *Hepatology*, 33(5), pp. 1098–1109.
Mort, R. L., Ford, M. J., Sakaue-Sawano, A., Lindstrom, N. O., Casadio, A., Douglas, A. T., Keighren, M. A., Hohenstein, P., Miyawaki, A. and Jackson, I. J. (2014a) 'Fucci2a: A bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice.' *Cell Cycle*, 13(17), pp. 2681–2696.
Mort, R. L., Ford, M. J., Sakaue-Sawano, A., Lindstrom, N. O., Casadio, A., Douglas, A. T., Keighren, M. A., Hohenstein, P., Miyawaki, A. and Jackson, I. J. (2014b) 'Fucci2a: A bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice.' *Cell Cycle*, 13(17), pp. 2681–2696.
Mossanen, J. and Tacke, F. (2015) 'Acetaminophen-induced acute liver injury in mice.' *Laboratory Animals*, 49(1), pp. 30–36.
Nauck, M., Wolfle, D., Katz, N. and Jungermann, K. (1981) 'Modulation of the Glucagon-Dependent Induction of Phosphoenolpyruvate Carboxykinase and Tyrosine Aminotransferase by Arterial and Venous Oxygen Concentrations in Hepatocyte Cultures.' *European Journal of Biochemistry*, 119(3), pp. 657–661.
Nejak-Bowen, K., Orr, A., Bowen, W. C. and Michalopoulos, G. K. (2013) 'Conditional genetic elimination of hepatocyte growth factor in mice compromises liver regeneration after partial hepatectomy.' *PLoS ONE*, 8(3), pp. 1–12.
Nevzorova, Y. A. and Trautwein, C. (2015) 'Regulation of cell cycle during liver regeneration.' *In Liver Regeneration: Basic Mechanisms, Relevant Models and Clinical Applications*, pp. 153–166.
Nguyen, A., Khoo, W. H., Moran, I., Croucher, P. I. and Phan, T. G. (2018) 'Single cell RNA sequencing of rare immune cell populations.' *Frontiers in Immunology*, 9(1553), pp. 1–11.
Ntranos, V., Yi, L., Melsted, P. and Pachter, L. (2018) 'Identification of transcriptional signatures for cell types from single-cell RNA-Seq.' *bioRxiv*, 258566, pp. 1–18.
Oe, S., Lemmer, E. R., Conner, E. A., Factor, V. M., Levéen, P., Larsson, J., Karlsson, S. and Thorgeirsson, S. S. (2004) 'Intact signaling by transforming growth factor β is not required for termination of liver regeneration in mice.' *Hepatology*, 40(5), pp. 1098–1105.
Olle, E. W., Ren, X., McClintock, S. D., Warner, R. L., Deogracias, M. P., Johnson, K. J. and Colletti, L. M. (2006) 'Matrix metalloproteinase-9 is an important factor in hepatic regeneration after partial hepatectomy in mice.' *Hepatology*, 44(3), pp. 540–549.
Olsen, P. S., Boesby, S., Kirkegaard, P., Therkelsen, K., Almdal, T., Poulsen, S. S. and Nexø, E. (1988) 'Influence of epidermal growth factor on liver regeneration after partial hepatectomy in rats.' *Hepatology*, 8(5), pp. 992–996.
Olsen, P. S., Poulsen, S. S. and Kirkegaard, P. (1985) 'Adrenergic effects on secretion of epidermal growth factor from Brunner's glands.' *Gut*, 26(9), pp. 920–927.

Paranjpe, S., Bowen, W. C., Mars, W. M., Orr, A., Haynes, M. M., DeFrances, M. C., Liu, S., Tseng, G. C., Tsagianni, A. and Michalopoulos, G. K. (2016) 'Combined systemic elimination of MET and epidermal growth factor receptor signaling completely abolishes liver regeneration and leads to liver decompensation.' *Hepatology*, 64(5), pp. 1711–1724.

Patil, M., Pabla, N. and Dong, Z. (2013) 'Checkpoint kinase 1 in DNA damage response and cell cycle regulation.' *Cellular and molecular life sciences : CMLS*, 70(21), pp. 4009–4021.

Pediaditakis, P., Lopez-Talavera, J. C., Petersen, B., Monga, S. P. S. and Michalopoulos, G. K. (2001) 'The processing and utilization of hepatocyte growth factor/scatter factor following partial hepatectomy in the rat.' *Hepatology*, 34(4), pp. 688–693.

Perera, P. Y., Lichy, J. H., Waldmann, T. A. and Perera, L. P. (2012) 'The role of interleukin-15 in inflammation and immune responses to infection: Implications for its therapeutic use.' *Microbes and Infection*, 14(3), pp. 247–261.

Planas-Paz, L., Orsini, V., Boulter, L., Calabrese, D., Pikiolek, M., Nigsch, F., Xie, Y., Roma, G., Donovan, A., Mart, P., Beckmann, N., Dill, M. T., Carbone, W., Bergling, S., Isken, A., Mueller, M., Kinzel, B., Yang, Y., Ma, X., Nicholson, T. B., Zamponi, R., Capodieci, P., Valdez, R., Rivera, D., Loew, A., Ukomadu, C., Terracciano, L. M., Bouwmeester, T., Cong, F., Heim, M. H., Forbes, S. J., Runer, H. and Tchorz, J. S. (2016) 'The RSPO-LGR4/5-ZNRF3/RNF43 module controls liver zonation and size.' *Nature Cell Biology*, 18(5), pp. 467–479.

Poisson, J., Lemoinne, S., Boulanger, C., Durand, F., Moreau, R., Valla, D. and Rautou, P. E. (2017) 'Liver sinusoidal endothelial cells: Physiology and role in liver diseases.' *Journal of Hepatology*, 66(1), pp. 212–227.

Pujol, M. J., Jaime, M., Serratosa, J., Jaumot, M., Agell, N. and Bachs, O. (2000) 'Differential association of p21(Cip1) and p27(Cip1) with cyclin E-CDK2 during rat liver regeneration.' *Journal of Hepatology*, 33(2), pp. 266–274.

Qiu, P. (2020) 'Embracing the dropouts in single-cell RNA-seq analysis.' *Nature Communications*, 11(1), pp. 1–9.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A. and Trapnell, C. (2017) 'Reversed graph embedding resolves complex single-cell trajectories.' *Nature Methods*, 14(10), pp. 979–982.

R Core Team (2020) 'R: A language and environment for statistical computing.' *R Foundation for Statistical Computing, Vienna, Austria*.

Rabes, H. M., Wirsching, R., Tuczek, H. V. and Iseler, G. (1976) 'Analysis of cell cycle compartments of hepatocytes after partial hepatecomy.' *Cell Proliferation*, 9(6), pp. 517–532.

Racine-Samson, L., Scoazec, J., D'Errico, A., Fiorentino, M., Christa, L., Moreau, A., Roda, C., Grigioni, W. F. and Feldmann, G. (1996) 'The metabolic organization of the adult human liver: A comparative study of normal, fibrotic, and cirrhotic liver tissue.' *Hepatology*, 24(1), pp. 104–113.

Ramachandran, P., Dobie, R., Wilson-Kanamori, J. R., Dora, E. F., Henderson, B. E. P., Luu, N. T., Portman, J. R., Matchett, K. P., Brice, M., Marwick, J. A., Taylor, R. S., Efremova, M., Vento-Tormo, R., Carragher, N. O., Kendall, T. J., Fallowfield, J. A., Harrison, E. M., Mole, D. J., Wigmore, S. J., Newsome, P. N., Weston, C. J., Iredale, J. P., Tacke, F., Pollard, J. W., Ponting, C. P.,

Marioni, J. C., Teichmann, S. A. and Henderson, N. C. (2019) 'Resolving the fibrotic niche of human liver cirrhosis at single-cell level.' *Nature*, 575(7783), pp. 512–518.

Rana, B., Mischoulon, D., Xie, Y., Bucher, N. L. and Farmer, S. R. (1994) 'Cell-extracellular matrix interactions can regulate the switch between growth and differentiation in rat hepatocytes: reciprocal expression of C/EBP alpha and immediate-early growth response transcription factors.' *Molecular and Cellular Biology*, 14(9), pp. 5858–5869.

Rappaport, A. M., Borowy, Z. J., Lougheed, W. M. and Lotto, W. N. (1954) 'Subdivision of hexagonal liver lobules into a structural and functional unit. Role in hepatic physiology and pathology.' *The Anatomical Record*, 119(1), pp. 11–33.

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., Hacohen, N., Haniffa, M., Hemberg, M., Kim, S., Klenerman, P., Kriegstein, A., Lein, E., Linnarsson, S., Lundberg, E., Lundeberg, J., Majumder, P., Marioni, J. C., Merad, M., Mhlanga, M., Nawijn, M., Netea, M., Nolan, G., Pe'er, D., Phillipakis, A., Ponting, C. P., Quake, S., Reik, W., Rozenblatt-Rosen, O., Sanes, J., Satija, R., Schumacher, T. N., Shalek, A., Shapiro, E., Sharma, P., Shin, J. W., Stegle, O., Stratton, M., Stubbington, M. J. T., Theis, F. J., Uhlen, M., Van Oudenaarden, A., Wagner, A., Watt, F., Weissman, J., Wold, B., Xavier, R. and Yosef, N. (2017) 'The human cell atlas.' *eLife*, 6(27041), pp. 1–30.

Rmilah, A. A., Zhou, W., Nelson, E., Lin, L., Amiot, B., Nyberg, S. L., Anan, C. and Rmilah, A. (2019) 'Understanding the marvels behind liver regeneration.' *Wiley Interdisciplinary Reviews: Developmental Biology*, 8(3), p. 340.

Roberts, A. B., McCune, B. K. and Sporn, M. B. (1992) 'TGF-β: Regulation of extracellular matrix.' *Kidney International*, 41(3), pp. 557–559.

Robinson, M. D., Mccarthy, D. J. and Smyth, G. K. (2010) 'edgeR: A bioconductor package for differential expression analysis of digital gene expression data.' *Bioinformatics*, 26(1), pp. 139–140.

Ronchini, C. and Capobianco, A. J. (2001) 'Induction of Cyclin D1 Transcription and CDK2 Activity by Notchic: Implication for Cell Cycle Disruption in Transformation by Notchic.' *Molecular and Cellular Biology*, 21(17), pp. 5925–5934.

Russell, W. E., Kaufmann, W. K., Sitaric, S., Luetteke, N. C. and Lee, D. C. (1996) 'Liver regeneration and hepatocarcinogenesis in transforming growth factor-α-targeted mice.' *Molecular Carcinogenesis*, 15(3), pp. 183–189.

Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y. (2019) 'A comparison of single-cell trajectory inference methods.' *Nature Biotechnology*, 37(5), pp. 547–554.

Sakamoto, T., Liu, Z., Murase, N., Ezure, T., Yokomuro, S., Poli, V. and Demetris, A. J. (1999) 'Mitosis and apoptosis in the liver of interleukin-6-deficient mice after partial hepatectomy.' *Hepatology*, 29(2), pp. 403–411.

Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R. E., Neild, A., Farbehi, N., Bartonicek, N. and Gallego-Ortega, D. (2019) 'Droplet-based single cell RNAseq tools: A practical guide.' *Lab on a Chip*, 19(10), pp. 1706–1727.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors.' *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463–5467.
Sasse, D., Spornitz, U. M. and Maly, I. P. (1992) 'Liver architecture.' *Enzyme*, 46(1–3), pp. 8–32.
Schleicher, J., Tokarski, C., Marbach, E., Matz-Soja, M., Zellmer, S., Gebhardt, R. and Schuster, S. (2015) 'Zonation of hepatic fatty acid metabolism - The diversity of its regulation and the benefit of modeling.' *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1851(5), pp. 641–656.
Seitz, H. K. and Stickel, F. (2006) 'Risk factors and mechanisms of hepatocarcinogenesis with special emphasis on alcohol and oxidative stress.' *Biol. Chem*, 387, pp. 349–360.
Shendure, J. and Ji, H. (2008) 'Next-generation DNA sequencing.' *Nature Biotechnology*, 26(10), pp. 1135–1145.
Shimizu, H., Miyazaki, M., Wakabayashi, Y., Mitsuhashi, N., Kato, A., Ito, H., Nakagawa, K., Yoshidome, H., Kataoka, M. and Nakajima, N. (2001) 'Vascular endothelial growth factor secreted by replicating hepatocytes induces sinusoidal endothelial cell proliferation during regeneration after partial hepatectomy in rats.' *Journal of hepatology*, 34(5), pp. 683–689.
Silverman, J. D., Roche, K., Mukherjee, S. and David, L. A. (2020) 'Naught all zeros in sequence count data are the same.' *bioRxiv*, 477794, pp. 1–17.
Six, E., Ndiaye, D., Laâbi, Y., Brou, C., Gupta-Rossi, N., Israël, A. and Logeat, F. (2003) 'The Notch ligand Delta1 is sequentially cleaved by an ADAM protease and γ-secretase.' *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), pp. 7638–7643.
Sokabe, T., Yamamoto, K., Ohura, N., Nakatsuka, H., Qin, K., Obi, S., Kamiya, A. and Ando, J. (2004) 'Differential regulation of urokinase-type plasminogen activator expression by fluid shear stress in human coronary artery endothelial cells.' *American Journal of Physiology - Heart and Circulatory Physiology*, 287(5 56-5), pp. 2027–2034.
Stanger, B. Z. (2015) 'Cellular homeostasis and repair in the mammalian liver.' *Annual Review of Physiology*, 77(1), pp. 179–200.
Steiling, H., Wüstefeld, T., Bugnon, P., Brauchle, M., Fässler, R., Teupser, D., Thiery, J., Gordon, J. I., Trautwein, C. and Werner, S. (2003) 'Fibroblast growth factor receptor signalling is crucial for liver homeostasis and regeneration.' *Oncogene*, 22(28), pp. 4380–4388.
Stolz, D. B., Mars, W. M., Petersen, B. E., Kim, T. H. and Michalopoulos, G. K. (1999) 'Growth factor signal transduction immediately after two-thirds partial hepatectomy in the rat.' *Cancer research*, 59(16), pp. 3954–3960.
Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019) 'Comprehensive integration of single-cell data.' *Cell*, 177(7), pp. 1888–1902.
Sugimoto, H., Yang, C., LeBleu, V. S., Soubasakos, M. A., Giraldo, M., Zeisberg, M. and Kalluri, R. (2007) 'BMP-7 functions as a novel hormone to facilitate liver regeneration.' *The FASEB Journal*, 21(1), pp. 256–264.
Sun, T., Pikiolek, M., Orsini, V., Bergling, S., Holwerda, S., Morelli, L., Hoppe, P. S., Planas-Paz, L., Yang, Y., Ruffner, H., Bouwmeester, T., Lohmann, F., Terracciano, L. M., Roma, G., Cong, F. and Tchorz, J. S. (2020) 'AXIN2+

pericentral hepatocytes have limited contributions to liver homeostasis and regeneration.' *Cell Stem Cell*, 26(1), pp. 97–107.
Sun, Z., Chang, B., Gao, M., Zhang, J. and Zou, Z. (2017) 'IL-33-ST2 axis in liver disease: Progression and challenge.' *Mediators of inflammation*, 2017(5314213), pp. 1–8.
Suzuki, A., McCall, S., Choi, S. S., Sicklick, J. K., Huang, J., Qi, Y., Zdanowicz, M., Camp, T., Li, Y. X. and Diehl, A. M. (2006) 'Interleukin-15 increases hepatic regenerative activity.' *Journal of Hepatology*, 45(3), pp. 410–418.
Svensson, V. (2020) 'Droplet scRNA-seq is not zero-inflated.' *Nature Biotechnology*, 38(2), pp. 147–150.
Tachibana, S., Zhang, X., Ito, K., Ota, Y., Cameron, A. M., Williams, G. M. and Sun, Z. (2014) 'Interleukin-6 is required for cell cycle arrest and activation of DNA repair enzymes after partial hepatectomy in mice.' *Cell and Bioscience*, 4(6), pp. 1–10.
Tan, X., Behari, J., Cieply, B., Michalopoulos, G. K. and Monga, S. P. S. (2006) 'Conditional deletion of β-Catenin reveals Its role in liver growth and regeneration.' *Gastroenterology*, 131(5), pp. 1561–1572.
Tanaka, M., Okada, K., Ueshima, S., Imano, M., Ohyanagi, H., Carmeliet, P. and Matsuo, O. (2001) 'Impaired liver regeneration after partial hepatectomy in plasminogen deficient mice.' *Fibrinolysis and Proteolysis*, 15(1), pp. 2–8.
Taniguchi, K., Wu, L. W., Grivennikov, S. I., De Jong, P. R., Lian, I., Yu, F. X., Wang, K., Ho, S. B., Boland, B. S., Chang, J. T., Sandborn, W. J., Hardiman, G., Raz, E., Maehara, Y., Yoshimura, A., Zucman-Rossi, J., Guan, K. L. and Karin, M. (2015) 'A gp130-Src-YAP module links inflammation to epithelial regeneration.' *Nature*, 519(7541), pp. 57–62.
Taub, R. (2004) 'Liver regeneration: from myth to mechanism.' *Nature Reviews Molecular Cell Biology*, 5(10), pp. 836–847.
Trefts, E., Gannon, M. and Wasserman, D. H. (2017) 'The liver.' *Current Biology*, 27(21), pp. 1147–1151.
Ünsal-Kaçmaz, K., Mullen, T. E., Kaufmann, W. K. and Sancar, A. (2005) 'Coupling of human circadian and cell cycles by the timeless protein.' *Molecular and Cellular Biology*, 25(8), pp. 3109–3116.
Vermeulen, K., Van Bockstaele, D. R. and Berneman, Z. N. (2003) 'The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer.' *Cell Proliferation*, 36(3), pp. 131–149.
Verne, J. (2014) *Liver disease: a preventable killer of young adults - Public health matters*. [Online] [Accessed on 14th April 2020] https://publichealthmatters.blog.gov.uk/2014/09/29/liver-disease-a-preventable-killer-of-young-adults/.
Vieth, B., Parekh, S., Ziegenhain, C., Enard, W. and Hellmann, I. (2019) 'A systematic evaluation of single cell RNA-seq analysis pipelines.' *Nature Communications*, 10(1), pp. 1–11.
Wang, B., Kaufmann, B., Engleitner, T., Lu, M., Mogler, C., Olsavszky, V., Öllinger, R., Zhong, S., Geraud, C., Cheng, Z., Rad, R. R., Schmid, R. M., Friess, H., Hüser, N., Hartmann, D. and von Figura, G. (2019) 'Brg1 promotes liver regeneration after partial hepatectomy via regulation of cell cycle.' *Scientific Reports*, 9(1), pp. 1–11.
Wang, H., Rao, B., Lou, J., Li, J., Liu, Z., Li, A., Cui, G., Ren, Z. and Yu, Z. (2020)

‘The function of the HGF/c-Met Axis in hepatocellular carcinoma.’ *Frontiers in Cell and Developmental Biology*, 8(55), pp. 1–15.

Wang, Lin, Wang, X., Xie, G., Wang, Lei, Hill, C. K. and DeLeve, L. D. (2012) ‘Liver sinusoidal endothelial cell progenitor cells promote liver regeneration in rats.’ *Journal of Clinical Investigation*, 122(4), pp. 1567–1573.

Wang, X., DeFrances, M. C., Dai, Y., Pediaditakis, P., Johnson, C., Bell, A., Michalopoulos, G. K. and Zarnegar, R. (2002) ‘A mechanism of cell survival: Sequestration of Fas by the HGF receptor Met.’ *Molecular Cell*, 9(2), pp. 411–421.

Webber, E. M., Wu, J. C., Wang, L., Merlino, G. and Fausto, N. (1994) ‘Overexpression of transforming growth factor-α causes liver enlargement and increased hepatocyte proliferation in transgenic mice.’ *American Journal of Pathology*, 145(2), pp. 398–408.

Wells, R. G. (2014) ‘The portal fibroblast: Not just a poor man’s stellate cell.’ *Gastroenterology*, 147(1), pp. 41–47.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. *Springer-Verlag New York*.

Williams, R., Alexander, G., Armstrong, I., Baker, A., Bhala, N., Camps-Walsh, G., Cramp, M. E., de Lusignan, S., Day, N., Dhawan, A., Dillon, J., Drummond, C., Dyson, J., Foster, G., Gilmore, I., Hudson, M., Kelly, D., Langford, A., McDougall, N., Meier, P., Moriarty, K., Newsome, P., O’Grady, J., Pryke, R., Rolfe, L., Rice, P., Rutter, H., Sheron, N., Taylor, A., Thompson, J., Thorburn, D., Verne, J., Wass, J. and Yeoman, A. (2018) ‘Disease burden and costs from excess alcohol consumption, obesity, and viral hepatitis: fourth report of the Lancet Standing Commission on Liver Disease in the UK.’ *The Lancet*, 391(10125), pp. 1097–1107.

Williams, R., Aspinall, R., Bellis, M., Camps-Walsh, G., Cramp, M., Dhawan, A., Ferguson, J., Forton, D., Foster, G., Gilmore, S. I., Hickman, M., Hudson, M., Kelly, D., Langford, A., Lombard, M., Longworth, L., Martin, N., Moriarty, K., Newsome, P., O’Grady, J., Pryke, R., Rutter, H., Ryder, S., Sheron, N. and Smith, T. (2014) ‘Addressing liver disease in the UK: A blueprint for attaining excellence in health care and reducing premature mortality from lifestyle issues of excess consumption of alcohol, obesity, and viral hepatitis.’ *The Lancet*, 384(9958), pp. 1953–1997.

Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F. and Quake, S. R. (2014) ‘Quantitative assessment of single-cell RNA-sequencing methods.’ *Nature Methods*, 11(1), pp. 41–46.

Xia, D., Li, X., Lou, Y., Han, W., Ding, P., Zhang, Y., Di, C., Song, Q. and Ma, D. (2002) ‘Overexpression of chemokine-like factor 2 promotes the proliferation and survival of C2C12 skeletal muscle cells.’ *Biochimica et Biophysica Acta - Molecular Cell Research*, 1591(1–3), pp. 163–173.

Yamada, Y., Kirillova, I., Peschon, J. J. and Fausto, N. (1997) ‘Initiation of liver growth by tumor necrosis factor: Deficient liver regeneration in mice lacking type I tumor necrosis factor receptor.’ *Proceedings of the National Academy of Sciences of the United States of America*, 94(4), pp. 1441–1446.

Yamada, Y., Webber, E. M., Kirillova, I., Peschon, J. J. and Fausto, N. (1998) ‘Analysis of liver regeneration in mice lacking type 1 or type 2 tumor necrosis

factor receptor: Requirement for type 1 but not type 2 receptor.' *Hepatology*, 28(4), pp. 959–970.

Yanagita, K., Nagaike, M., Ishibashi, H., Niho, Y., Matsumoto, K. and Nakamura, T. (1992) 'Lung may have an endocrine function producing hepatocyte growth factor in response to injury of distal organs.' *Biochemical and Biophysical Research Communications*, 182(2), pp. 802–809.

Yang, J., Cusimano, A., Monga, J. K., Preziosi, M. E., Pullara, F., Calero, G., Lang, R., Yamaguchi, T. P., Nejak-Bowen, K. N. and Monga, S. P. (2015) 'WNT5A inhibits hepatocyte proliferation and concludes β-Catenin signaling in liver regeneration.' *American Journal of Pathology*, 185(8), pp. 2194–2205.

Yang, J., Mowry, L. E., Nejak-Bowen, K. N., Okabe, H., Diegel, C. R., Lang, R. A., Williams, B. O. and Monga, S. P. (2014) 'Beta-catenin signaling in murine liver zonation and regeneration: A Wnt-Wnt situation!' *Hepatology*, 60(3), pp. 964–976.

Yu, G. and He, Q. Y. (2016) 'ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization.' *Molecular BioSystems*, 12(2), pp. 477–479.

Zaher, H., Buters, J. T. M., Ward, J. M., Bruno, M. K., Lucas, A. M., Stern, S. T., Cohen, S. D. and Gonzalez, F. J. (1998) 'Protection against acetaminophen toxicity in CYP1A2 and CYP2E1 double- null mice.' *Toxicology and Applied Pharmacology*, 152(1), pp. 193–199.

Zhang, F., Zhang, J., Li, X., Li, B., Tao, K. and Yue, S. (2018) 'Notch signaling pathway regulates cell cycle in proliferating hepatocytes involved in liver regeneration.' *Journal of Gastroenterology and Hepatology*, 33(8), pp. 1538–1547.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. and Enard, W. (2017) 'Comparative analysis of single-cell RNA sequencing methods.' *Molecular Cell*, 65(4), pp. 631–643.

Zielke, N. and Edgar, B. A. (2015) 'FUCCI sensors: Powerful new tools for analysis of cell proliferation.' *Wiley Interdisciplinary Reviews: Developmental Biology*, 4(5), pp. 469–487.

Zou, Y., Bao, Q., Kumar, S., Hu, M., Wang, G.-Y. and Dai, G. (2012) 'Four waves of hepatocyte proliferation linked with three waves of hepatic fat accumulation during partial hepatectomy-induced liver regeneration.' Singh, S. R. (ed.) *PLoS ONE*, 7(2), pp. 1–8.