Single-cell analysis of hepatic progenitor cells during liver injury and recovery

DETERMINING THE FACTORS RESPONSIBLE FOR HEPATIC REGENERATION AND REPAIR

Annelijn Speel

CENTRE FOR REGENERATIVE MEDICINE | UNIVERSITY OF EDINBURGH AND UTRECHT UNIVERSITY | 25-04-2023

Final report research project: Single-cell analysis of hepatic progenitor cells during liver injury and repair: Determining the factors responsible for hepatic regeneration and repair.

Research project as part of the Selective Utrecht Medical Master, University of Utrecht
Carried out at the Centre for Regenerative Medicine, University of Edinburgh

Supervisors: prof Stuart Forbes and Dr Sofia Ferreira-Gonzalez
Internal supervisor from Utrecht University: Dr Sabine Fuchs

Student: Annelijn Speel
Student number: 5848474
Date: 25-04-2023

Acknowledgements

I would like to thank my supervisors Dr Sofia Ferreira Gonzalez, Dr Alastair Kilpatrick, and prof Stuart Forbes for their guidance and support during this project. I am extremely grateful to have had the opportunity to work together with – and learn from - every member of the Forbes lab and I would like to extend special thanks to Dr Daniel Rodrigo Torres and Dr Candice Ashmore-Harris for their support and guidance during this project.

I would like to thank Dr Scott Waddell and Dr Luke Boulter for the guidance during the CRISPR knock out experiment and Tak Yung (Janet) Man for teaching me many laboratory skills.

I would like to express special thanks to Dr Sabine Fuchs for the guidance in organizing this internship and prof Stuart Forbes for the opportunity to become part of the Forbes group and learn many research skills.

Finally, I would like to mention that this work has made use of the resources provided by the Edinburgh Compute and Data Facility.

Abstract

Chronic liver disease is a major worldwide health concern accounting for approximately two million deaths per year, and this situation is only expected to aggravate in the upcoming years. Currently, liver transplantation is the only curative option for life-threatening end-stage liver disease, which is limited by the shortage of donor livers. It is therefore needed to develop new therapeutic strategies. Chronic liver disease is characterized by compromised hepatocyte regeneration resulting in scarring and compromised liver function. Previous research revealed that biliary epithelial cells (BECs) act as facultative stem cells – hepatic progenitor cells (HPCs) - when hepatocyte regeneration is compromised and can differentiate into hepatocytes to repair the liver parenchyma. Therefore, BECs are a potential appropriate target for the development of new therapies for the treatment of chronic liver disease. Recent research showed a change in the transcriptional profile of BECs from biliary towards hepatocyte-like gene expression and identified heterogeneity within the biliary epithelium during liver injury and recovery. However, the underlying molecular mechanisms through which BECs differentiate into hepatocytes during BEC-mediated liver regeneration and repair remains unknown.

In this project, a single cell RNA sequencing dataset of biliary cells was analyzed during liver injury and recovery, in order to identify the biological pathways responsible for BEC-driven liver regeneration and repair.
This analysis resulted in the selection of 17 genes of interest of which the function of *Serpina1c, Gc,* and *Ccl2* was studied in HPCs performing a CRISPR knock out approach. All lentivirus transductions were successful and I determined the optimal lentivirus dilutions to obtain viable and well expanding cells. The resulting *Serpina1c* KO cell line was further studied with quantitative reverse transcription PCR to determine transcriptional differences compared to control HPCs.
My results suggest that *Serpina1c* might play a role in the differentiation status of BECs as well as in their stemness and proliferative capacities. Future research is needed to confirm these findings and explore the role of the remaining genes of interest in BEC-mediated liver regeneration and repair.

Table of contents

Chapter 1 - Introduction

Liver diseases are highly prevalent conditions with poor long-term clinical outcomes and are therefore a major health problem (Asrani et al., 2019). Worldwide, liver diseases - including viral hepatitis, hepatocellular carcinoma, and cirrhosis – account for approximately two million deaths per year (Asrani et al., 2019; Pimpin et al., 2018). This situation is only expected to aggravate due to the ever-increasing rates of non-alcoholic fatty liver disease (NAFLD) in the Western world, which is associated with an increase in obesity cases (Pimpin et al., 2018). Two-thirds of these patients die before the age of 65 years resulting in a more considerable loss of working life years compared to most other chronic diseases (Pimpin et al., 2018).
Currently, liver transplantation is the only curative treatment for life-threatening end-stage liver disease, and this procedure is severely limited by the current shortage of donor livers (Asrani et al., 2019; P. T. W. Kim & Testa, 2016). Alternative therapeutic options (such as hepatocyte cell therapy) have been proposed, however, with very limited clinical success (Forbes et al., 2015). It is therefore needed to develop new therapeutic strategies to treat liver diseases.

The liver has a high regenerative capacity capable of restoring its mass and function after injury. However, its regenerative function is compromised with repeated, severe liver damage (So et al., 2020).
At cellular level, there are two modes of liver regeneration: (i) hepatocyte- and (ii) hepatic progenitor cell (HPC)-driven liver regeneration (Duncan et al., 2009; Miyajima et al., 2014). Hepatocyte-driven liver regeneration is the primary regeneration source, involving proliferation of preexisting hepatocytes (Michalopoulos & DeFrances, 1997). On the other hand, in HPC-driven liver regeneration, biliary epithelial cells (BECs) or hepatocytes first dedifferentiate into HPCs, followed by HPC proliferation and subsequent differentiation into hepatocytes or BECs (Duncan et al., 2009; Falkowski et al., 2003; Gadd et al., 2020; Lu et al., 2015; Miyajima et al., 2014; Raven et al., 2017; Rodrigo-Torres et al., 2014; Russell et al., 2019; Santoni-Rugiu et al., 2005; Xiao et al., 2003).
During severe liver injury, hepatocytes become senescent which is marked by an irreversible cell cycle arrest driven by p21 and/or p16 (Lu et al., 2015). Senescent hepatocytes are not able to proliferate and hence regeneration is not effective, leading to increased scarring and compromised liver function. In this scenario, the secondary regeneration source, the hepatic progenitor cell, plays a significant role in liver repair (Marshall et al., 2005; Raven et al., 2017; Richardson et al., 2007).

When hepatocyte proliferation is compromised, a ductular reaction (DR) emerges (Gouw et al., 2011; He et al., 2014; Marshall et al., 2005; Petersen et al., 1998), which describes the proliferation and expansion of ductular structures located at the interface of the parenchymal and portal compartments that may derive from HPCs, pre-existing cholangiocytes, or hepatocytes (Aleksieva, 2021; Boulter et al., 2013; Gouw et al., 2011;

Roskams et al., 2004). So et al. (2020) stated that the prevalence of HPC-driven regeneration in severe liver injury settings suggests that this regeneration mode should be clinically beneficial to patients with severe liver diseases. Promoting HPC-driven liver regeneration might therefore be clinically beneficial to these patients (So et al., 2020). For this reason, this project proposes HPCs as a potential source for liver cell replacement as an alternative therapeutic strategy to treat end-stage liver diseases (Shin & Kaestner, 2014).

Preliminary data leading to this project

Dr Niya Aleksieva, a former PhD student in the Forbes lab, showed that differentiation of BECs towards hepatocytes occurred after diet induced liver injury (Aleksieva, 2021). Individual biliary-derived hepatocytes emerged as part of and in close association with the areas of expanding ductular reaction.

Dr Aleksieva studied transcriptional differences between BECs isolated from mouse models in which hepatocytes can regenerate (control), or in which the regenerative capacity of hepatocytes was initially compromised due to cellular senescence and BEC-mediated regeneration occurred (CLD, p21-driven model) (see a description of the dataset in Chapter 2.1). Bulk transcriptomic analysis and single-cell RNA Sequencing (scRNA-Seq) was performed at different time points during liver injury and recovery to study the signaling pathways and genes involved in BEC-derived hepatic regeneration. This study helped to identify transcriptional differences between the groups indicating a BEC profile transition from a biliary phenotype towards a hepatocyte expression program. Additionally, heterogeneity within the biliary epithelium was identified as the BECs in this study contained cell populations that only emerged in the animals with compromised hepatocyte regeneration (Aleksieva, 2021).

In this project, I analyzed the above scRNA-Seq dataset of BECs during liver injury and recovery with the aims of:

(i) Identifying the biological pathways and genes responsible for BEC-mediated liver regeneration and repair.

(ii) Determine the role of these pathways and genes in BECs *in vitro*.

The overarching hypothesis of my project is that clusters of BECs in the p21 condition of the aforementioned dataset can be identified by pathways and genes that are important for the differentiation of BECs - or HPCs - towards hepatocytes during liver injury and repair.

By analyzing this dataset, I aim to contribute to a better understanding of the underlying molecular mechanisms of BEC-mediated hepatocyte regeneration with the long-term goal of developing novel therapies for the treatment of CLD.

Chapter 2 - Materials and methods

1. Bioinformatics analysis

1.1 Experimental design

A scRNA-seq dataset of BECs was created by Dr Niya Aleksieva, a former PhD student in the Forbes group (Aleksieva, 2021). To create this dataset, K19CreERTtdTomato animals - in which a tdTomato reporter is expressed in K19-expressing cells - were used (**Figure 1A**). To mimic chronic liver disease (CLD), hepatocyte regeneration was compromised in the CLD animal group by inducing p21 overexpression in hepatocytes. This was induced by the administration of Adeno-associated viral vectors serotype 8 (AAV8), expressing p21 under the hepatocyte-specific thyroid binding globulin (TBG) promotor (AAV8.TBG.p21). Impaired hepatocyte regeneration characterized by p21-mediated cell cycle arrest is a common feature of liver disease and is correlated to the degree of fibrosis and associated with poor disease outcome (Aleksieva, 2021; Aravinthan et al., 2013; Marshall et al., 2005) . The control animals were injected with an empty AAV8.TBG vector – AAV8.TBG.null. After transgene induction, liver injury was induced in both groups by administering a methionine- and choline-deficient (MCD) diet for 15 days, which reproduces NAFLD induced CLD. The healthy animals did not receive transgene induction, nor liver injury.
The mice were subsequenctly culled at different time points (peak MCD injury, and recovery days 3 [d3], 6 [d6], 9 [d9], and 12 [d12]) to compare gene expression between different conditions at different time points after liver injury.

1.2 Single-cell RNA-seq dataset

Bile ducts were isolated from the mice and tdTomato+EpCAM+ cells of all animals were sorted with fluorescence activated cell sorting (FACS) (**Figure 1B**). Biliary epithelial cells (BECs) were sorted by positive selection for cholangiocytes (EpCAM+ cells), and negative selection for leukocytes (CD45+ cells), endothelial cells (CD31+ cells), and erythrocytes (Ter199+ cells). The sorted cells of the mice culled at recovery days 6 and 9 - 2 male mice per condition per timepoint - were used for scRNA-Seq. The cells were fixated and frozen after which library preparation and sequencing was performed as described by Dr Aleksieva (2021). For this project, the scRNA-seq data of the healthy and day 9 recovery animals was selected for analysis for reasons described in Chapter 3.1.
Alignment of scRNA-seq data and barcode counting of samples was performed on the Eddie compute cluster at the University of Edinburgh, using 10X Genomics Cell Ranger count (v7.0.0) with reference dataset mm10-2020-A (GRCm38.98). Sample IDs for each sample in the healthy or day 9 conditions were passed to Cell Ranger aggr, which combined the results for each sample into a single file, which could be analyzed using the

Loupe browser software (v6.0, www.10xgenomics.com/products/loupe-browser). This file was prepared for me by Dr Alastair Kilpatrick (Centre for Regenerative Medicine, Liver Stem Cells and Regeneration group).

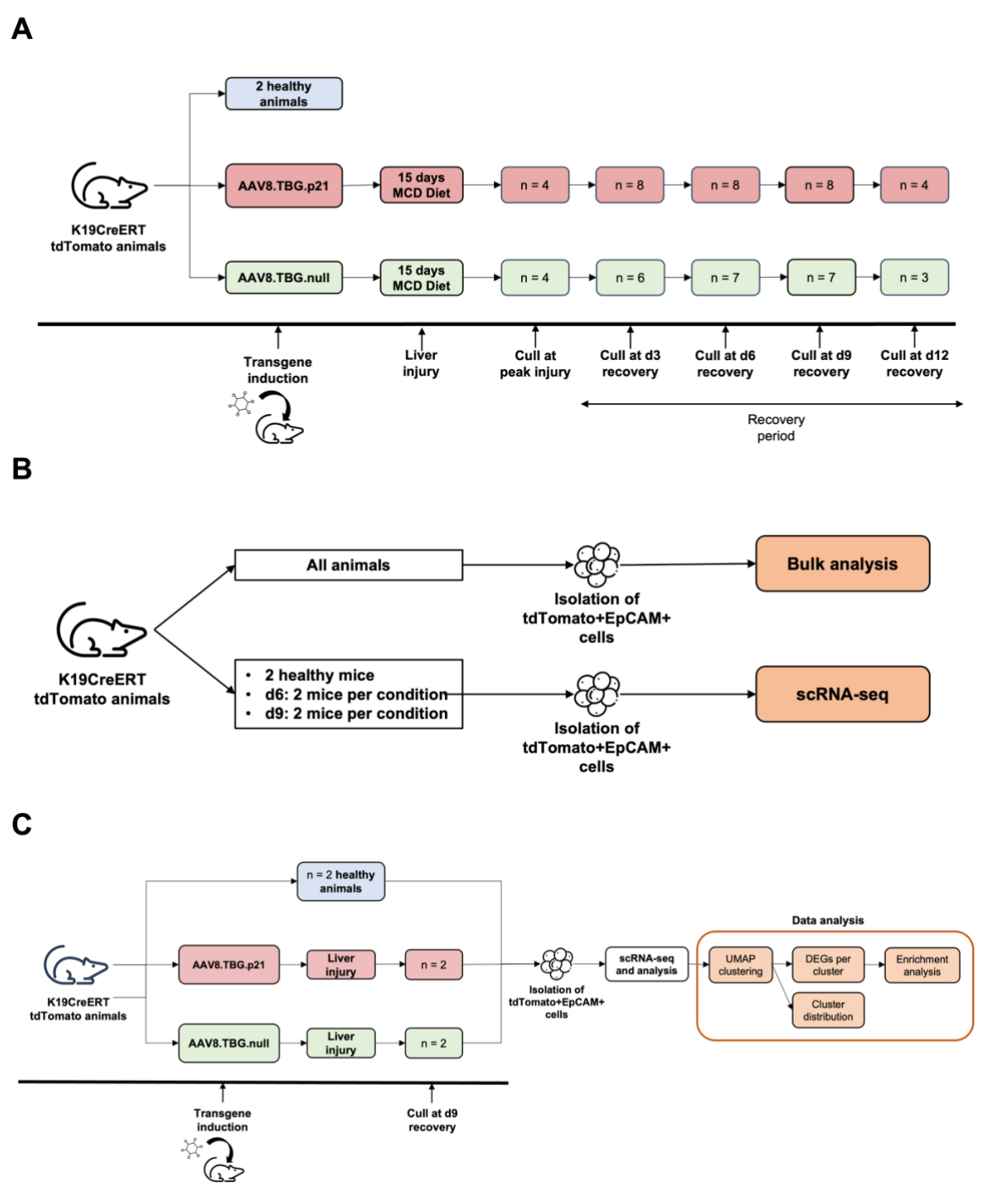


Figure 1: Experimental model for single cell sequencing and analysis.

(**A**) Experimental design showing 3 K19CreERTtdTomato mouse groups: healthy, AAV8.TBG.p21, and AAV8.TBG.null. AAV8.TBG.p21 and AAV8.TBG.null mice were treated with MCD for 15 days to induce liver injury. 'n' numbers represent the number of mice culled at each timepoint in the timeline. (**B**) Flowchart showing the performed bulk and scRNA-seq analysis on tdTomato+EpCAM+ cells derived from mice in (A). (**C**) Represents the tdTomato+EpCAM+ cells derived from mice that were used for scRNA-seq in (B) and for subsequent analysis in this project (red box). Data analysis consisted of Uniform Manifold Approximation and Projection (UMAP) clustering, determination of differentially expressed genes (DEGs) for each condition and cluster, assessment of cluster distribution, and enrichment analysis.

1.3 Single-cell RNA-seq analysis

The scRNA-seq dataset was analyzed using Loupe Browser (v6.0, 10x Genomics, https://www.10xgenomics.com/products/loupe-browser/downloads). When opening the file in Loupe Browser, the data could be viewed as a Uniform Manifold Approximation and Projection (UMAP) plot containing 15 clusters of cells. UMAP is a technique for dimension reduction that can be used for visualization (*Understanding UMAP*, n.d.). A UMAP plot shows clusters of points of which each point represents a cell. The distance between a point and its neighbor is related to how connected these points are (*Understanding UMAP*, n.d.) which, in this case, is based on similarity in gene expression.

scRNA-seq quality control and preprocessing

A Quality control (QC) analysis was performed to verify the FACS sorting, ensuring that the cells were *Epcam*+, *Ptprc*- (CD45), *Pecam1*- (CD31), and *Ly76*- (Ter119); did not express genes corresponding with cell types other than BECs or BEC-derived cells; and did not contain >15% mitochondrial reads (indicating poor sample quality). Gene expression of genes of interest was analyzed to characterize the dataset. Genes of interest were hepatocyte, cholangiocyte, HPC, senescence, proliferation, and stemness (surface) marker genes; genes involved in the Hedgehog, Notch, JAK/STAT, Wnt/β-catenin, Hippo, and transforming growth factor (TGF)-β pathways; and vascular endothelial growth factor (VEGF)-signaling, foetal liver, biliary developmental, DR, and inflammatory genes. Gene expression was analyzed using the "Gene/Feature Expression" function in Loupe Browser and results were viewed as Log2 Expression (Log2E). Cluster 14 and 15 had an overall low expression of any gene of interest, and 5,1% and 41,25% of the cells contained >15% mitochondrial reads respectively. Therefore, these clusters were removed from downstream analysis.

Outlier groups (OGs) were manually identified based on their outlying position in the UMAP plot containing all cells of the dataset. For each OG, the differentially expressed genes (DEGs) were determined using the "Globally Distinguishing" function in Loupe Browser. This function finds the DEGs that distinguish one cluster or group from all cells outside this cluster or group (*Finding Significant Genes -Software -Single Cell Gene Expression -Official 10x Genomics Support*, n.d.). This comparison is based on the exact negative binomial test (Yu et al., 2013), or the fast asymptotic negative binomial test when counts become large (Robinson & Smyth, 2007).

DEGs were defined as genes with a log2 Fold Change (Log2FC) of expression > 0 and a p-value <0.05. To examine the representing cell type for each OG, the DEGs were exported from Loupe and imported into R using the read.csv() function. OG2 had no DEGs. For the other OGs, comparison of the DEGs for each OG suggested that there was little overlap in DEGs. Then, the DEGs for each of the remaining OGs were tested against specific cell types, by analyzing using FACS-sorted and annotated liver data from the Tabula Muris scRNA-seq dataset (n=2,859 cells; Schaum et al., 2018). To perform the analysis, the AddModuleScore() function from the Seurat R package (v4.3.0;Hao et al., 2021) was used.

This function computes the average expression for cluster markers at a single-cell level and compares it across annotated cell types. The function returns a module score for each cell, with higher module scores reflecting higher expression of the input cluster markers. Module scores were visualized as boxplots for each cell type using ggplot2.
Subsequently, enrichment analysis was performed on the list of DEGs for each OG using Gene Ontology (GO) biological processes via the online Panther tool (v17.0, http://geneontology.org/ (Mi et al., 2019; Thomas et al., 2022)). GO terms significantly enriched with a False Discovery Rate (FDR) <0.05 and <500 associated genes (to exclude unspecific GO terms associated with a very high number of genes) were selected to determine the functional state of the cells in each OG.
Based on the above analysis, OG1 was removed from downstream analysis (see Chapter 3.2).

Furthermore, a cluster similarity analysis was performed to examine the overlap of DEGs between the clusters. Lists of DEGs for each cluster (FDR<0.05) were exported from Loupe and imported into R as above. The number of DEGs in common between each pair of clusters was counted and visualized as a heatmap using the Pheatmap R package (v1.0.12), with brighter colours indicating more shared DEGs and darker colours fewer shared DEGs. Pairs of clusters with no DEGs in common are coloured gray. Clusters with a high percentage of shared DEGs (>50%; cluster 1 and 9, and cluster 5 and 13) were further studied by determening the DEGs for these paired clusters using the "Locally Distinguishing" function. This function finds the DEGs of one cluster or condition compared to the other checked cluster(s) or condition(s). The DEGs distinguishing the clusters were manually analyzed for genes of interest to determine if the clusters should be merged or kept as separate clusters. Genes of interest are described in Chapter 2.1.3 and might be important for BEC-driven liver regeneration. As the distinguishing DEGs contained some genes of interest, cluster 1, 9, 5, and 13 were kept as separate clusters.
Removing OG1 and cluster 14 and 15 from the analysis was done using the "Filter" option in Loupe Browser.

Data analysis
After removal of OG1 and cluster 14 and 15, the scRNA-seq dataset was analyzed, of which the different types of analyses are shown in **Figure 1C**. The DEGs for each cluster were determined as above. Each cluster was divided into three subgroups (healthy, d9ctrl, and d9p21) using the "Filter" option in Loupe Browser. Subsequently, the DEGs for each subgroup within every cluster were determined using the 'Globally Distinguishing' and 'Locally Distinguishing" function in Loupe Browser. Using the 'Globaly Distinguishing' function, the DEGs for every subgroup within a cluster were determined by comparing gene expression of a given subgroup with the other 2 subgroups. Using the 'Locally Distinguishing' function, the DEGs for every subgroup within a cluster were determined by comparing gene expression of every subgroup with 1 other subgroup. The DEGs for every condition were analyzed for genes of interest, Log2FC of expression, and frequency among all clusters.

For each cluster, enrichment analysis was performed using the same method as for the OGs. This was performed on the list of DEGs for every cluster as described above. GO terms with FDR <0.05 and a <500 associated genes were selected for further analysis. The selected GO terms were manually analyzed and selected based on functions of interest that potentially play a role in liver injury and BEC-driven liver regeneration (**Table 1**). Subsequently, genes associated with the selected GO terms were analyzed and ranked by the number of GO terms that they were associated with.

Table 1: functions of interest in GOEA

Functions of interest	
Liver functoins	Growth
Signaling pathways	Apoptosis and cell death
Regeneration	Hormone synthesis
Liver cells	Oxidative stress
Immune system responses/functions	Proliferation
Vascular changes	Senescence
Vascular development	Chaperones and refolding proteins

The DEGs of the subgroups within every cluster were analyzed based on Log2FC of expression, containing genes of interest (described in Chapter 2.1.3), and in how many clusters the gene was a DEG for each condition.

The selection of genes that were most of interest when comparing d9p21 to the other conditions was based on pseudo-bulk analysis, Log2FC of expression of DEGs and enrichment analysis for clusters mainly represented by d9p21 BECs (2, 6, 10, 11, 12), and comparing gene expression between d9p21 and the other conditions within the clusters. Due to time restrictions, 3 genes were selected for *in vitro* validation. First, 17 and then 6 genes were selected based on the analyses above. These 6 genes were further studied by analyzing protein interactions using STRING (v11.5; Szklarczyk et al., 2021) and by manually looking up the genes' functions and associated pathways. For these 6 genes and for any other studied gene or protein in this project, this was performed using several tools: National Center for Biotechnology Information (NCBI) (*Home - Gene - NCBI*, n.d.; Sayers et al., 2022), Bgee (v14.1; Bastian et al., 2021), UniProt (release 2023_01; Consortium et al., 2023), and STRING (v11.5; Szklarczyk et al., 2021). Based on this information and the analysis above, 3 genes were selected for *in vitro* validation.

2. Isolation and culture of bile duct organoids and HPCs

Bile duct organoids and hepatic progenitor cells (HPCs) were isolated and cultured using different methods depending on two-dimensional or three-dimensional culture.

2.1 Bile duct organoids

Wildtype female mouse bile ducts were isolated and transferred to Matrigel (Corning, #354230) for culture by members of the Forbes group as described previously by Ferreira-Gonzalez et al. (2022). Isolated bile ducts were cultured until passage 4 after which the organoids were frozen down. Culture medium composition was Advanced DMEM/F12 (Gibco, #12634010) supplemented with A-83-01 (5 uM, Sigma-Aldrich, #909910-43-6), B27 and N2 (Invitrogen, #17504044 and #17502001), N-acetylcysteine (1.25 uM, Sigma-Aldrich, #A9165), forskolin (10 uM, Sigma-Aldrich, #66474029-9), gastrin (10 nM, Sigma-Aldrich), HEPES pH7 (10 mM, Gibco, #15630-056), EGF (50 ng/ml, Peprotech, #315-09), FGF10 (100 ng/mL, Peprotech, #AF-450-33), HGF (50 ng/mL, Peprotech, #315-23), 10% R-spondin1 (Peprotech, #315-32), and nicotinamide (10 mM, Sigma-Aldrich, #N0636).
Bile duct organoids were passaged in 1:4-1:8. The organoids were mechanically dissociated into small fragments and base medium was added to dilute the Matrigel. Then, the organoids were centrifuged at 400G for 5 min, supernatant was discarded, and organoids were resuspended in 50% fresh Matrigel (diluted in base medium). Base medium composition was Advanced DMEM/F12 (Thermo Fisher, #12634010) supplemented with 1% Penicilline/Streptomycin (Gibco, #15140-122) and 1% L-Glutamine (Gibco, #25030-024). The organoids were then added to wells as droplets and were topped up with culture media. Freezing down organoids was done by resuspending in CELLBANKER 2 Cryopreservative Medium (Zenogen Pharma, #210119) and storage at -80°C and subsequently in liquid nitrogen.

2.2 HPCs

EpCAM+/CD24+/CD133+ HPCs were isolated, sorted, and plated on rat tail collagen 1 (25%, Sigma-Aldrich, #C3867) coated plates by members of the Forbes group as described by Lu et al. (2015). At passage 14, the sorted HPCs were frozen down in CELLBANKER Medium as above and stored in liquid nitrogen. Base medium was added to thawed HPCs, cells were centrifuged at 200G for 7 minutes, supernatant was removed, and HPCs were resuspended in culture medium and seeded on collagen coated plates as above. Base medium composition was William's E medium (Gibco, #22551-022) with 10% FCS (APS, #A1010115), 1% L-Glutamine (Gibco, #25030-024), and 1% Penicilline/Streptomycin (Gibco, #15140-122). Culture medium composition was base medium, NaHCO3 (17.6 mM, Sigma-Aldrich, #S876), HEPES pH7.5 (20 mM, Gibco, #15630-056), nicotinamide (10mM, Sigma-Aldrich, #N0636), Sodium Pyruvate (1mM, Gibco, #11360-039), 1x Insulin Transferrin Selenium Solution (ITS) (Gibco, #51300-044),

dexamethasone (100 nM, Sigma-Aldrich, #D4902), Ascorbic Acid (0.2 mM, Sigma-Aldrich, #A4544), Glucose (14 mM, Sigma-Aldrich, #G7021), Il-6 (10ng/mL, Peprotech, #216-16), HGF (100ug/mL, Peprotech, #315-23), and EGF (100ug/mL, Peprotech, #315-09).
Cultured HPCs were passaged by washing with Dulbecco's Phosphate Buffered Saline (PBS) (Sigma-Aldrich, #D8537) and incubation with 700 µL Accutase solution (Sigma-Aldrich, #A6964) for 7-14 minutes at 37°C. Base medium was then added to inhibit the reaction of Accutase and cells were then centrifuged at 200G for 7 minutes. Supernatant was discarded and the cells were resuspended in culture medium and plated on collagen coated plates as above.

3. CRISPR-editing of *Serpina1c*, *Gc*, and *Ccl2*

Target sequence cloning and lentivirus production was performed under supervision of Dr Scott Waddell (Luke Boulter research group, MRC human genetics unit, Institute of Genetics & Cancer, the University of Edinburgh, UK).

3.1 Target sequence cloning

Oligos were selected using the GeCKO v2 mouse library guide RNA (gRNA) sequence catalogue and SnapGene software (v6.2, www.snapgene.com), and were synthesized as single stranded forward (top) and backward (bottom) oligos (Integrated DNA Technologies) (see **Table 2**). The LentiCRISPRv2 (Addgene, #82416) backbone was digested (to produce a linear backbone) and dephosphorylated using Esp3I (New England Biolabs, #R0734), FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher Scientific, #EF0651), 10X FastDigest Buffer (ThermoFisher Scientific, #B64), 100 mM DTT (Takra, #1908A80A), and ddH2O. The digested plasmids were gel purified by gel electrophoresis in a 1.5% agarose gel for 120-180 minutes (100 V, constant) and extracted using the QIAquick Gel Extraction Kit (Qiagen). A successful digestion was visible by a change in the molecular weight compared to the undigested plasmid upon gel visualization (**Supplementary figure 1A**). gRNA oligos were phosphorylated and annealed using 10X T4 Ligation Buffer (New England Biolabs, #B0202S), ddH2O, and T4 PNK (New England Biolabs, #M0201S) for 30 minutes at 37°C, 5 minutes at 95°C, and ramped down to 25°C at 5°C/min. The annealed oligos were then diluted to 1:200 in sterile water. The digested plasmid backbone and the annealed oligos (50 ng) were ligated with T4 DNA ligase (New England Biolabs, #M0202L) and T4 DNA Ligase Buffer (New England Biolabs, #B0202S) at 16°C overnight. Ligation reactions were then heat-inactivated and resulting plasmids were transformed into Stbl3 bacteria (One shot Stbl3 Chemically Competent E. coli; ThermoFisher Scientific, #C737303) as instructed by the manufacturer. The bacteria were grown overnight on Luria Broth (LB) agar with ampicillin at 37°C. A vector-only negative control ligation and transformation was performed by adding water in place of oligos.

For each gRNA, 5 colonies were picked. Small plasmid DNA of the colonies was isolated with the QIAprep Spin Miniprep Kit (Qiagen, #27104 and #27106) and subsequently sequenced (Sanger sequencing, Technical Services at the Institute of Genetics & Cancer, The University of Edinburgh, UK). Sequencing was performed using a customized primer (Integrated DNA Technologies, sequence: GAG GGC CTA TTT CCC ATG ATT CC) that binds to the U6 promotor upstream of the guide insert. Clones with correctly ligated plasmids were picked, cultured and plasmid DNA was then extracted using the Qiagen Plasmid Maxi Kit March 2016 (Qiagen, #12162, #12163, and #12165). DNA concentration was determined by conventional spectrophotometry (Nanodrop). Absorbance measurements at 260 nm were used to determine concentration and 230 and 280 nm values were used to determine DNA purity.

Sanger sequencing repeatedly resulted in nonsense results for a significant part of the plasmids likely due to contamination. Gel electrophoresis on the plasmids was performed as above and confirmed contamination of some of the colonies as the molecular weight of the plasmids was different from the LentiCRISPRv2 plasmid (**Supplementary figure 1B-C**). This means that these colonies had probably taken up a different plasmid during cloning. From then on, the purified plasmid DNA was first analyzed for contamination by gel electrophoresis. Only the plasmids with the correct molecular size were sent for Sanger sequencing as above to identify clones with correctly ligated plasmids (**Supplementary figure XD-E**).

Table 2: Oligo sequences

Name	Mouse GeCKOv2 ID	Oligo sequence
***Serpina1c*_guide1_Top**	MGLibB_47793	CACCGAAATGACTCCCTCCATCTCA
***Serpina1c*_guide1_Btm**	MGLibB_47793	AAACTGAGATGGAGGGAGTCATTTC
***Serpina1c*_guide2_Top**	MGLibB_47794	CACCGTAAAGAGAATGTAATTTGCC
***Serpina1c*_guide2_Btm**	MGLibB_47794	AAACGGCAAATTACATTCTCTTTAC
***Serpina1c*_guide3_Top**	MGLibA_47807	CACCGGCACATCAAGCATTCCCGAG
***Serpina1c*_guide3_Btm**	MGLibA_47807	AAACCTCGGGAATGCTTGATGTGCC
***Ccl2*_guide1_Top**	MGLibA_08788	CACCGGCCGGCAACTGTGAACAGCA
***Ccl2*_guide1_Btm**	MGLibA_08788	AAACGCCGGCAACTGTGAACAGCAC
***Ccl2*_guide2_Top**	MGLibB_08787	CACCGGCTGGTGAATGAGTAGCAGC
***Ccl2*_guide2_Btm**	MGLibB_08787	AAACGCTGCTACTCATTCACCAGCC
***Ccl2*_guide3_Top**	MGLibB_08788	CACCGGTGCTGACCCCAAGAAGGAA
***Ccl2*_guide3_Btm**	MGLibB_08788	AAACTTCCTTCTTGGGGTCAGCACC
***Gc*_guide1_Top**	MGLibA_19677	CACCGTCTAGGCCGAGACTATGAGA
***Gc*_guide1_Btm**	MGLibA_19677	AAACTCTCATAGTCTCGGCCTAGAC
***Gc*_guide2_Top**	MGLibB_19668	CACCGACCCCACCTGCTACGACACC
***Gc*_guide2_Btm**	MGLibB_19668	AAACGGTGTCGTAGCAGGTGGGGTC
***Gc*_guide3_Top**	MGLibB_19670	CACCGTGCCCTCAGTCACCAGCCGC
***Gc*_guide3_Btm**	MGLibB_19670	AAACGCGGCTGGTGACTGAGGGCAC

3.2 Lentivirus production

Lentivirus was produced for *Serpina1c* (guide 1-3), *Gc* (guide 3), *Ccl2* (guide 3), and a non-target gRNA (scramble; guide sequence: 5' ATAACTTATGCGCTTCGGGG 3'), which was produced as a negative control.

HEK293 cells (passage 1+) were passaged. Cells were washed with PBS and lifted with 1.5 mL Trypsin (Gibco, #25300-054) after which medium was added to inactivate Trypsin. Medium consisted of DMEM (Sigma #D5796/Gibco #41965-039) with 10% FBS (Hyclone, #SV30180.03), and Antibiotic-Antimycotic (Invitrogen-Gibco, #15240-062). Cells were then centrifuged at 1000 rpm for 5 minutes, supernatant was discarded. Cells were resuspended in Medium, plated at 4.5 x 10^6 cells in 10 ml on 10cm dishes (Corning, #430167), and cultured overnight at 37 ^oC.

The next day, Medium was removed and 10mL fresh medium was added to each plate. Transfection mix was made combining 425 µL Lentiviral Vector mix and 75 µL PEI mix. 7.5µg of psPAX2 Helper Vector (Addgene, #12260), 2.5ug of pCMV-VSV-G envelope plasmid (Addgene, #8454), 10ug of Lentiviral vector was added to 500ml of DMEM with 10% FBS to make Lentiviral Vector mix. PEI mix consisted of polyethyleneimine (PEI, linear, MW 25K, Polysciences, #23966) that was dissolved in heated water (80 ^oC), cooled down, adjusted to pH 7.0, filtered with 0.2, and frozen). Transfection mix was added to HEK293 cells drop by drop and cells were kept overnight at 37 ^oC. Media were changed the following day and cells were cultured overnight at 37^oC. The next day, media were collected and filtered (0.35uM cellulose, VWR #514-0063 or Fisher #15216869). Lentivirus was then concentrated using Lenti-X Concentrator (Takara, #631232) as per manufacturer's instructions. Virus was kept at 4^oC for 4 hours and centrifuged at 1500G for 45 minutes at 4^oC. Supernatant was removed and pellet containing virus was resuspended in 100 µl PBS. Virus was kept at -80 ^oC.

3.3 Lentiviral transduction of HPCs

HPCs were passaged 1:3 and cultured overnight in 2 mL media at 37^oC (Medium composition described in methods of HPC isolation and culture). Lentiviral transduction was performed in 70% confluent HPCs using 1.5 mL media containing 8 µg/mL Polybrene Infection / Transfection Reagent (Sigma-Aldrich, #TR-1003) and variable lentivirus dilutions listed in **Table 3**. Media were changed after 24h. Additionally, medium containing only polybrene was added to a subset of HPCs to examine Polybrene toxicity.

HPCs were analyzed for GFP+ cells and images were taken on a Nikon Eclipse TS2R microscope fitted with a Photometrics CoolSNAP DYNO camera using NIS-Elements (v.5.21.00). Successfully transduced cells were subsequently sorted for GFP+ cells with FACS.

Table 3: Lentivirus dilutions for transduction

gRNA	1/50	1/100	1/250	1/500
Serpina1c guide1	X	X		X
Serpina1c guide 2	X	X		X
Serpina1c guide 3		X		X
Gc guide 3		X	X	X
Ccl2 guide 3	X	X	X	X
Scramble		X	X	X

3.4 Cell sorting

Transgenic *Serpina1c* KO HPCs and scramble control HPCs were washed with PBS, lifted with 700 µL Accutase solution (Sigma-Aldrich, #A6964) for 7 minutes at 37°C. Base medium was added to inhibit the Accutase reaction and cells were then centrifuged at 200G for 7 minutes. Supernatant was discarded and cells were resuspended in base Medium without FCS. Base medium composition was William's E medium (Gibco, #22551-022) with 1% L-Glutamine (Gibco, #25030-024), and 1% Penicilline/Sterptomycin (Gibco, #15140-122). Cells were filtered (100 µm), counted using Tryphan blue (Corning, #25-900-Cl) and a Neubauer chamber, and centrifuged at 200G for 7 minutes. Supernatant was discarded and HPCs were resuspended in 200 µL of base medium as above. Cells were then filtered (35 µM) and FACS was performed by the staff at the Centre for Regenerative Medicine, Flow and Genomic Cytometry Facility using a BDFACSAria II special order system. GFP+, DRAQ7- (Abcam; #ab109202) cells were selected, collected, seeded on collagen coated plates, and cultured in culture media at 37°C as described previously.

3.5 Zeocin selection

As an alternative for FACS, a Zeocin (Thermo Fisher Scientific, #R25001) selection strategy to select successfully transduced cells was prepared using the Invitrogen Zeocin™ Selection Reagent User Guide (2012). Selection Medium was prepared and consisted of culture medium and 0, 100, 200, 400, 800 or 1000 µg/mL Zeocin. 1.5 mL Selection Medium was added to HPCs. 400 µg/mL killed +/- 70%, and 800µg/mL killed +/- 95% of the HPCs. Therefore, we decided that 600 µg/mL was the right concentration for Zeocin selection in HPCs.
The 1/500 transduced HPCs were treated with selection Medium containing 600µg/mL Zeocin on day 10 after transduction. The effect on the number of viable and GFP+ cells was measured with a Nikon Eclipse TS2R microscope as above on day 14 after transduction.

4. Reverse transcription and Real Time PCR

4.1 RNA extraction from HPCs

RNA extraction was performed using the RNeasy Mini Kit, Part 1 (March 2016) (Qiagen, #74104 and #74105) as per manufacturer's instructions. RNA concentration was assessed using NanoDrop ND-100 Spectrophotometer (ThermoFisher Scientific).

4.2 Reverse transcription and Real Time PCR

Reverse transcription (RT) was performed using Quantitect RT kit (Jan 2011) (Qiagen, #205310, 3205311, #205313, #205314) and a T100 Thermal Cycler (BIO-RAD). The resulting cDNA was diluted 1:8 in sterile RNase and DNase free water (Invitrogen, #10977-035). Real-time PCR (qPCR) was prepared combining 4µL diluted cDNA with 2.5µL Fast SYBR Green Master Mix (Thermo Fisher Scientific, #4385612), 2.5µL RNase free water, and 1µL primer. A list of primers used in this study is provided in **Table 4**.
Positive control qPCRs were performed to test the *Serpina1c, Gc*, and *Ccl2* primers. The *Serpina1c* primer was tested in hepatocytes treated with ethanol or water, the *Gc* primer in injured bulk liver tissue (paracetamol), and the *Ccl2* primer in HPCs treated with paracetamol or water.
Reverse transcription and Real Time PCR (RT-qPCR) was then performed for 4 biological replicates of *Serpina1c* KO and scramble control HPCs, and 3 biological replicates of non-transduced control HPC using LightCycler 480 II (Roche). Samples were run in technical triplicates.
Gene expression was normalized using Glyceraldehyde 3- phosphate dehydrogenase (*Gapdh)* and Peptidylpropyl isomerase A (*Ppia*) as housekeeping genes. Data analysis was performed in Microsoft Excel (v 16.70) using relative quantification method.

Table 4: List of primers for Real Time PCR

Primer	Manufacturer	Cat. No
Alb	Qiagen	QT00115570
Ccl2	Qiagen	QT00167832
Cyp2e1	Qiagen	QT00112539
EpCam	Qiagen	QT02304456
Gapdh	Qiagen	QT01658692
Gc	Qiagen	QT00267799
Hnf1b	Qiagen	QT00103320
Hnf4a	Qiagen	QT00144739
Krt19	Qiagen	QT00156667
Gpr49	Qiagen	QT00123193
Mki67	Qiagen	QT00247667
Onecut1	Qiagen	QT00297815
Pcna	Qiagen	QT00103313
Ppia	Qiagen	QT00247709
Prom1	Qiagen	QT01065162
Serpina1c	Qiagen	QT01046038
Sox9	Qiagen	QT00163765

5. Histology, immunohistochemistry, and immunofluorescence

5.1 Tissue

Tissue collection was performed by members of the Forbes group as described by Raven et al (2017). Liver tissue was fixed in 10% formalin, neutral buffered (Sigma-Aldrich #HT501128) for 8 hours and embedded in paraffin by the Histology service, SuRF, Queen's Medical Research Institute, Edinburgh University, UK.

4 µm thick paraffin slides of liver tissue were cut on a Rotary Microtome 2125RT (Leica) and transferred to a microscope slide. Antibodies used for immunostaining in this study are listed in **Table 5**. Positive control immunostainings of Monocyte Chemoattractant Protein 1 (MCP1, *Ccl2)* were performed on either liver and kidney tissue derived from male mice treated with Thioacetamide to induce liver injury, or from AAV8.TBG.p21 mice treated with MCD diet for 15 days and culled on day 6 recovery. Positive control immunostainings for α-1-antitrypsin 1-3 and GC were performed on tissue derived from untreated 8 weeks old wildtype C57/Bl6 mice.

Immunohistochemistry

For 3, 3'-diaminobenzidine (DAB) staining, paraffin slides were dewaxed in xylene for 2x 5 minutes and rehydrated through alcohols (100%, 100%, 95%, 80%, and 70%) for 30 seconds each and washed in tap water for 1-2 minutes. Then, heat-mediated antigen retrieval was performed in 100mM Tris-EDTA (pH 9.0) using an 800-Watt microwave. Slides were cooled down in running tap water for 5 minutes and blocked for endogenous peroxidase and alkaline phosphatase activity with BLOXALL® Blocking Solution (VECTOR laboratories, #SP-6000-100). The slides were washed 3 times with PBS after each blocking step. Endogenous avidin and biotin expression was blocked by adding three drops of Avidin and Biotin solution respectively (BioLegend, #SIG-31126) after which non-specific protein binding was blocked by adding 3 drops of Protein Block (Abcam, #ab64226) for at least 30 minutes. The slides were incubated overnight at 4°C with at least 120 µl of primary antibody solution consisting of primary antibody diluted in Antibody Diluent (Invitrogen, #003218). Isotype Control antibodies were used as a negative control. After washing 3 times with PBS, slides were incubated with species-specific biotinylated secondary antibodies for 30 minutes to detect primary antibodies, 3 drops of R.T.U Vectastain Kit (VECTOR laboratories, #PK-7100), and 120 µl of ImmPACT DAB EqV Substrate Kit, Peroxidase (HRP) (VECTOR laboratories, #SK-4103) for 1-2 minutes. Slides were washed 3 times with PBS after every incubation step. The slides were then counterstained in haematoxylin for 30-40 seconds, washed in cold running water, placed in Scott's tap water for 20-30 seconds, washed again, dehydrated through alcohols (70, 80%, 95%, 100%, and 100%) for 20 seconds each, and cleared in xylene for 2 times 5 minutes. Slides were mounted with PERTEX Mounting Medium (CellPath, #00801-EX)

5.2 Bile duct organoids

1% Agarose (Sigma-Aldrich, #A95359) was prepared at 55-60°C. culture medium was aspirated, and agarose was added to the plate to be solidified. Agarose was then flipped over, and the bile duct organoids were covered with more agarose. Solidified agarose was subsequently fixed, embedded in paraffin, and cut similar as described for tissue.

Organoid were characterized by staining for EpCAM, PCNA, K19, and Ki67 using immunofluorescence staining.

Immunofluorescence

Dewaxing, rehydration, and antigen retrieval was performed as described for DAB staining. Then, slides were incubated with 3 drops of Protein Block for at least 30 minutes, incubated with 120 µl primary or isotype control antibody as described above for 1 hour at room temperature or overnight at 4°C, and washed 3 times with PBS. The slides were incubated with species-specific fluorescent conjugated secondary antibodies for 30-40 minutes to detect the primary antibody (**Table 5**). After washing with PBS 3 times, DAPI (Sigma-Aldrich, #D9542) was added for 10-15 minutes, the slides were washed again, and mounted with Fluoromount-G Mounting Medium (Invitrogen, #00-4958-02).

5.3 HPCs

Positive control stainings of MCP1 (Abcam, #ab8101) were performed on HPCs. HPC were prepared for staining with a cytospin 4 (ThermoShandon) or in a 6-well plate. 15.000-20.000 resuspended cells in 200 µl were added to each cytofunnel and cytocentrifuged at 300G for 5 minutes. The slides were subsequently fixed with Acetone:Methanol (1:1) for 10 minutes at room temperature and washed 3 times with PBS for 3-5 minutes. To prepare HPCs in a 6-well plate, HPCs were passaged as described previously, seeded on a circular coverslide in a collagen coated plate, and kept in culture at 37°C. Media was then removed, 1 mL Acetone:Methanol (1:1) was added to for 10 minutes at room temperature, and cells were washed 3 times with PBS for 3-5 minutes.

Immunofluorescence

Slides or wells were washed 3 times with PBS and 3 drops of Protein Block was added for at least 30 minutes. HPCs were then incubated with primary antibody for 1 hour at room temperature and washed 3 times with PBS. Subsequent incubation with secondary antibody and DAPI antibody, washing steps, and mounting were performed as described for organoids.

Table 5: List of antibodies for immunohistochemistry

Primary antibody	Manufacturer	Cat. No	Host species	Dilution
EpCAM	Abcam	Ab71914	Rabbit	1/00
Gc	Sigma-Aldrich	HPA019855	Rabbit	1/500
Goat IgG, Control Antibody	VECTOR laboratories	I-5000	Goat	Adjusted to primary antibody
K19 (TROMA III)	DSHB	AB_2133570	Rat	1/50
Ki67	Abcam	Ab16667	Rabbit	1/200
MCP1	Abcam	Ab8101	Rat	1/10 - 1/300
MCP1	Abcam	ab25124	Rabbit	1/100-1/200
Mouse IgG, Control Antibody	VECTOR laboratories	I-2000	Mouse	Adjusted to primary antibody
PCNA	Abcam	Ab29	Mouse	1/500
Rabbit IgG, Control Antibody	VECTOR laboratories	I-1000	Rabbit	Adjusted to primary antibody
Rat IgG, Control Antibody	VECTOR laboratories	I-4000	Rat	Adjusted to primary antibody
α-1-antitrypsin 1-3	R&D Systems	AF2979	Goat	1/500
Secondary antibody	**Manufacturer**	**Cat. No**	**Host species**	**Dilution**
anti-Goat IgG (H+L), Biotinylated	VECTOR laboratories	BA-5000	rabbit	1/500
anti-mouse IgG (H+L), Alexa FluorTM 555	Invitrogen	A31570	donkey	1/200
anti-rabbit IgG (H+L), Alexa FluorTM 488	Invitrogen	A11008	goat	1/200
anti-rabbit IgG (H+L), Alexa FluorTM 555	Invitrogen	A31572	donkey	1/200
anti-rabbit IgG (H+L), Biotinylated	VECTOR laboratories	BA-1100	horse	1/250-1/500
anti-rat IgG (H+L), Alexa FluorTM 488	Invitrogen	A21208	donkey	1/200
anti-Rat IgG (H+L), Biotinylated	VECTOR laboratories	BA-9400	goat	1/250-1/400
anti-rat IgG, Alexa FluorTM 555	Invitrogen	A48270	donkey	1/200

5.4 Imaging

Images of stained slides were taken on a Nikon ECLIPSE H550L microscope fitted with a MiroPublisher 6 CCD camera using Image-Pro Premier (v9.3.1, Media Cybernetics) and edited using ImageJ2 (v2.9.0/1.53t).

Chapter 3 - Bioinformatics analysis

1. Description of the dataset

To study the factors responsible for hepatic regeneration and repair, Dr Niya Aleksieva generated two different groups (biliary epithelial cells (BECs) derived from mice with impaired hepatocyte regeneration (p21 group) and BECs derived from mice with intact hepatocyte regeneration (control animals)). These conditions were compared at a transcriptional level (previously described in Chapter 2.1.1, **Figure 1A**). Both animal groups were exposed to liver injury (MCD dietary damage,) and culled at different time points of recovery. Additionally, a healthy group consisting of 2 healthy mice was created to study transcriptional differences between BECs derived from injured and healthy mice.

Dr Aleksieva performed RT-qPCR on the isolated BECs at multiple timepoints (peak MCD injury, and recovery days 3 [d3], 6 [d6], 9, and 12 [d12] (**Figure 1B**) (Aleksieva, 2021). RT-qPCR was focused on four categories: biliary genes, Notch pathway associated genes, hepatocyte genes, and Wnt pathway associated genes. Notch pathway and Wnt signaling have been associated with BEC and hepatocyte specification respectively (Boulter et al., 2012; Spee et al., 2010; Tchorz et al., 2009; Zong et al., 2009).
The RT-qPCR results suggested that the td+Tomato+EpCAM+ cells displayed the most transcriptional changes between the control and p21 condition from day 6 recovery onwards. Genes associated with (commitment or differentiation towards) a hepatocyte phenotype - *Alb, Krt8, Foxa2*, and *Cebpa* - showed significant or a trend for upregulated expression in the p21 condition compared to the control condition. Additionally, *Dvl1* and *Fzd1* - genes of the canonical Wnt signaling pathway, which have been shown to play a role in differentiation of BECs towards hepatocytes (Boulter et al., 2012; Okabe et al., 2016) – showed a similar pattern of transcriptional differences from day 6 recovery onwards. This suggested that a subset of the BECs were starting to develop hepatocyte characteristics from day 6 recovery onwards. Histological data from Dr Aleksieva confirmed this result, indicating more pronounced differences in abundance of BEC-derived hepatocytes in the liver parenchyma during recovery after MCD injury (Aleksieva, 2021).

Next, a scRNA-seq dataset of mouse HPCs of biliary origin after liver injury was generated to study heterogeneity and to define the signature of the HPCs capable of differentiating towards hepatocytes (Aleksieva, 2021). scRNA-seq was performed on EpCAM+ cells derived from the mice culled at recovery days 6 and 9 (n=2 for both timepoints), and 2 healthy mice (**Figure 1B**). The scRNA-seq data was visualized with Uniform Manifold Approximation and Projection (UMAP) plots for each condition (**Figure 2**). A clear difference in UMAP plot point density is visible on day 9 in which a subset of cells is present in the d9p21 condition and absent in the d9ctrl and healthy conditions. Therefore, further analysis of transcriptional differences between d9p21 and d9ctrl on day 9 might

be informative in order to find biological pathways that may play a role in the differentiation of BECs towards hepatocytes.
Therefore, this project is focused on analyzing transcriptional differences between BECs derived from livers in which hepatocyte regeneration is impaired and BECs from livers with intact hepatocyte regeneration, on day 9 after MCD-mediated liver injury.

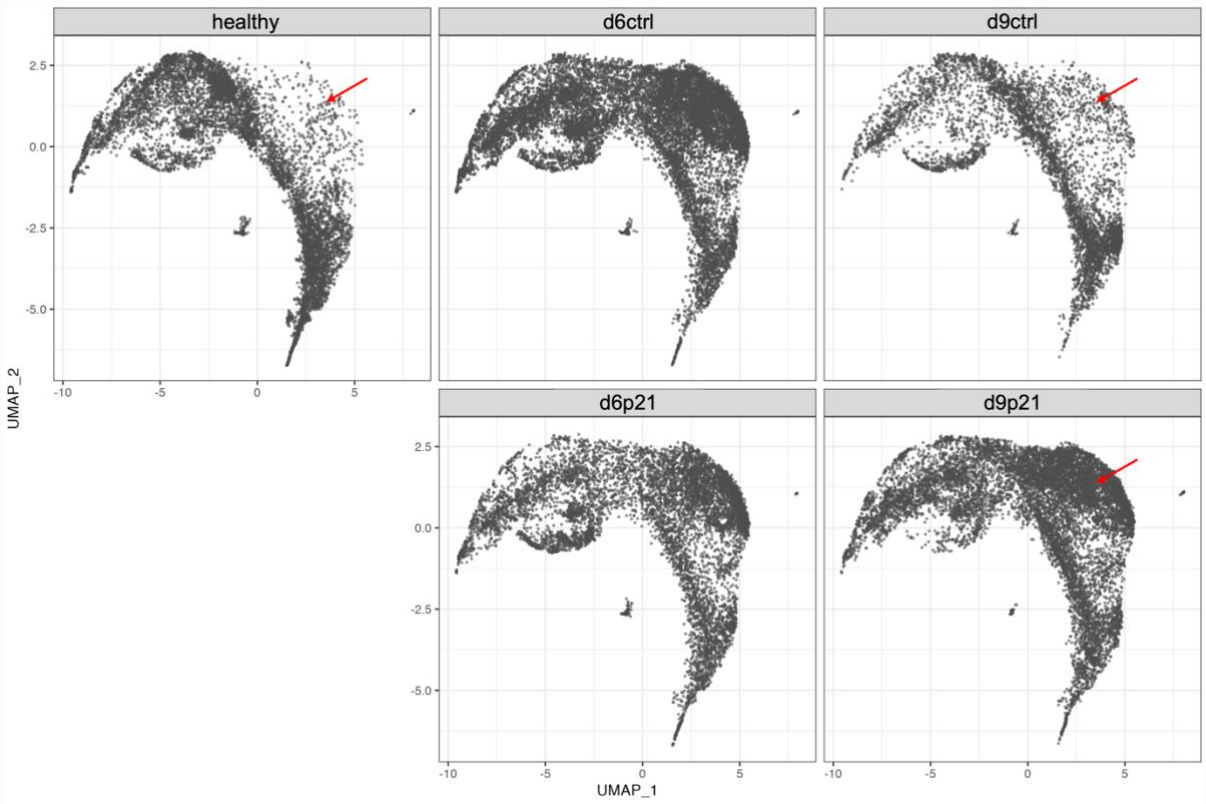


Figure 2: UMAP plots of single cell transcriptomics dataset.

ScRNA-seq dataset visualized in Uniform Manifold Approximation and Projection (UMAP) plots for healthy mice (healthy), recovery day 6 AAV8.TBG.null (d6ctrl) and AAV8.TBG.p21 (d6p21) mice, and recovery day 9 AAV8.TBG.null (d9ctrl) and AAV8.TBG.p21 (d9p21) mice. A clear difference in point density is shown when comparing d9p21 with both d9ctrl and healthy (red arrows). n=2 animals per condition.

2. Quality control analysis

The scRNA-seq dataset for this project consisted of three experimental conditions representing BECs derived from healthy, d9ctrl (AAV8.TBG.null, d9 recovery), and d9p21 (AAV8.TBG.p21, d9 recovery) mice (**Figure 3A**). Two biological replicas (2 mice) were included per condition (**Supplementary figure 2**). The data was split into 15 clusters by Loupe browser (**Figure 3B**).

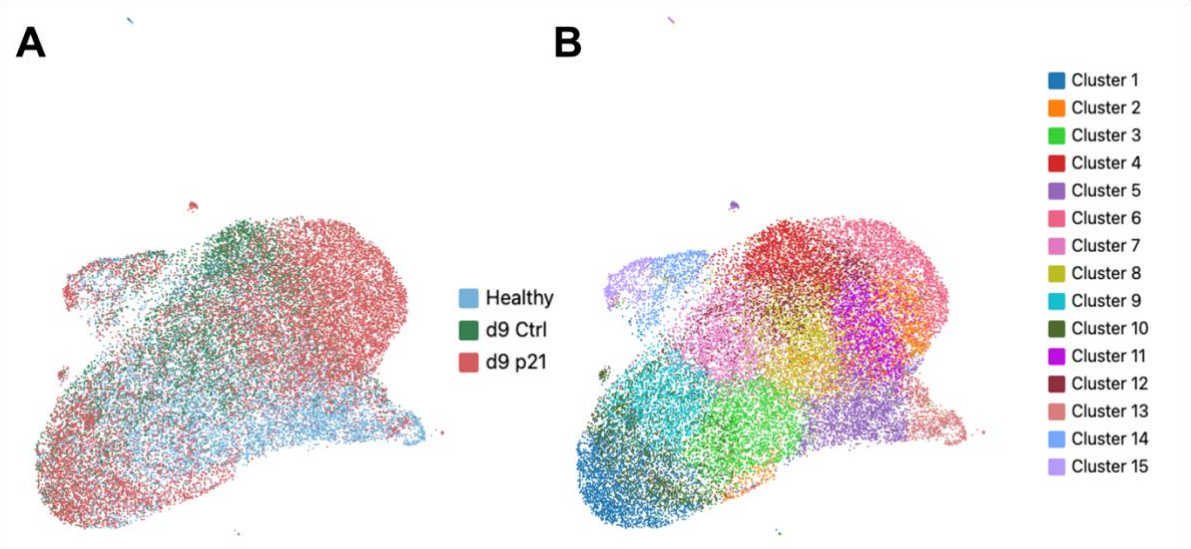


Figure 3: Groups and clusters of the scRNA-seq dataset

Uniform Manifold Approximation and Projection (UMAP) plots showing the conditions (**A**) and clusters (**B**) in the scRNA-seq dataset. Each point represents a cell.

Before analyzing the dataset, a quality control (QC) analysis was carried out. To assess if the cells in the dataset were expressing BEC genes that defined this population during FACS sorting (EpCAM+, CD45-, CD31-, Ter119+), gene expression of *Epcam, Ptprc, Pecam1*, and *Ly76* was analyzed (**Figure 4**).

Data for *Ly76* was not available in the dataset and is therefore not shown. The presence of *Epcam* and absence of *Ptprc* and *Pecam1* expression verifies the FACS BEC selection strategy.

Furthermore, expression of *Cdkn1a* (p21) and of proliferation (*Mki67* and *Pcna*), hepatocyte (*Krt18, Alb,* and *Serpina1c*), cholangiocyte (*Krt19,* and *Sox9*), HPC (*Cd44, Cd24a,* and *Sox9*), and potential DR (*Dmbt1*) markers was confirmed (**Figure 4-5**). This analysis indicates expression of all the above genes in the dataset except for *Mki67*. This might suggest an overall low proliferation rate in the cells on day 9 recovery, however, expression of *Pcna* is visible in a subset of the cells and so this cannot be concluded from this data on its own.

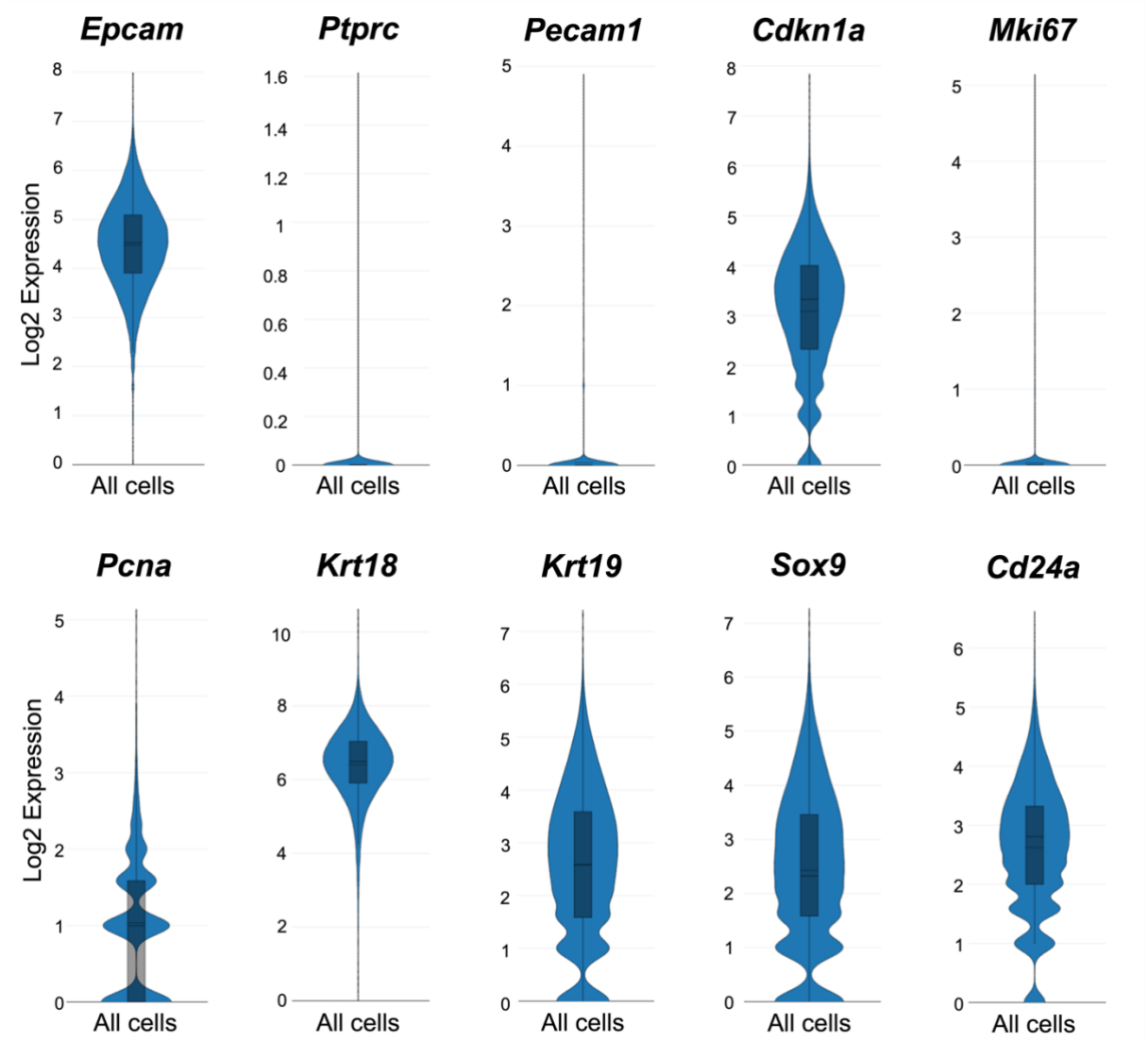


Figure 4: Characterization of the scRNA-seq dataset.

Violin plots showing Log2 Expression of *Epcam*, *Ptprc*, *Pecam1*, *Cdkn1a*, *Mki67*, *Pcna*, *Krt18*, *Krt19*, *Sox9*, and *Cd24a* in the scRNA-seq dataset.

Gene expression analysis of the above genes at a cluster level showed that the expression of *Epcam* and every other gene is relatively low in cluster 14 and 15 (**Figure 5**). To examine if this was technically driven, the percentage of mitochondrial genes was analyzed. 5.1% of cluster 14 and 41.25% of cluster 15 contained >15% mitochondrial genes whereas this was ≤ 0.1% for all other clusters. Based on these findings, cluster 14 and 15 were excluded from downstream analyses.

Looking at cluster 1-13, expression of *Epcam* is relatively constant throughout all clusters. Expression of the hepatocyte markers *Alb* and *Krt18* and the HPC marker *Cd44* slightly varies throughout the clusters. *Serpina1c* (a hepatocyte marker) expression is very high in cluster 2, 6, 10, and 13 compared to the remaining clusters. The potential DR gene *Dmbt1* and the HPC gene *Prom1* are both low in every cluster except for cluster 13.

The above-described patterns of expression confirm expression of hepatocyte, HPC, and BEC markers and transcriptional heterogeneity within the scRNA-seq dataset.

Figure 5: Characterization of the scRNA-seq dataset: Gene expression across clusters.

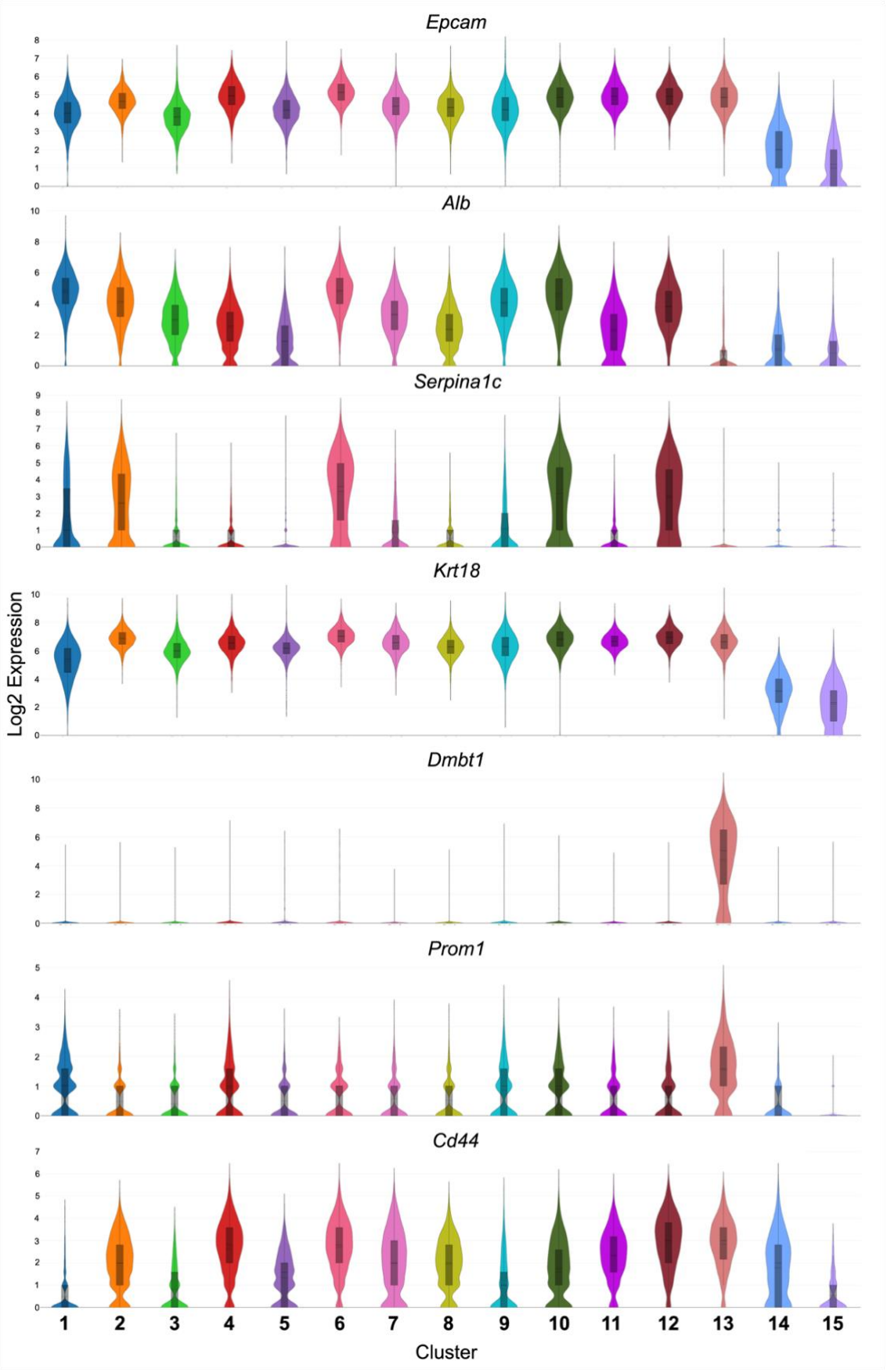


Figure 5: Characterization of the scRNA-seq dataset: Gene expression throughoutt clusters.

Violin plots showing Log2 Expression of *Epcam*, *Alb*, *Serpina1c*, *Krt18*, *Dmbt1*, *Prom1*, *and Cd44* across the 15 clusters (different colours) of the scRNA-seq dataset.

Manually selected outlier groups (OGs) were selected and analyzed to find out if they contained aberrant cells (not representing BECs or hepatocytes) (**Figure 6A**). Information about the number of cells and group distribution in each OG is presented in **Table 6 and Figure 6B**. Gene expression analysis resulted in a list of significant differentially expressed genes (DEGs) for each OG except OG2. Analyzing these DEGs for associations with specific cell types revealed that DEGs of OG1 contained marker genes of endothelial cells of the hepatic sinusoid (**Figure 6C-E**). Additionally, functional enrichment analysis of the DEGs in OG1 showed a high number of GO terms associated with functions of endothelial cells (**figure 7A**), whereas the GO terms for OG3 and OG4 were related to inflammation and tissue remodeling, and epithelial cell functions respectively (**Figure 7B-C**).

To confirm that OG1 represents endothelial cells, gene expression of 3 endothelial cell markers (*Egfl7, Pecam1,* and *Flt1*) was analyzed. This revealed an overall low gene expression throughout the dataset except for OG1 (**Figure 8A-D**).

Based on all findings, OG1 cells were defined as a contaminating population of endothelial cells of the hepatic sinusoid and were therefore excluded from downstream analysis.

Table 6: Number of cells in outlier groups

	Number of cells (% of all cells)
Outlier group 1	49 (0.18%)
Outlier group 2	72 (0.27%)
Outlier group 3	12 (0.05%)
Outlier group 4	55 (0.21%)

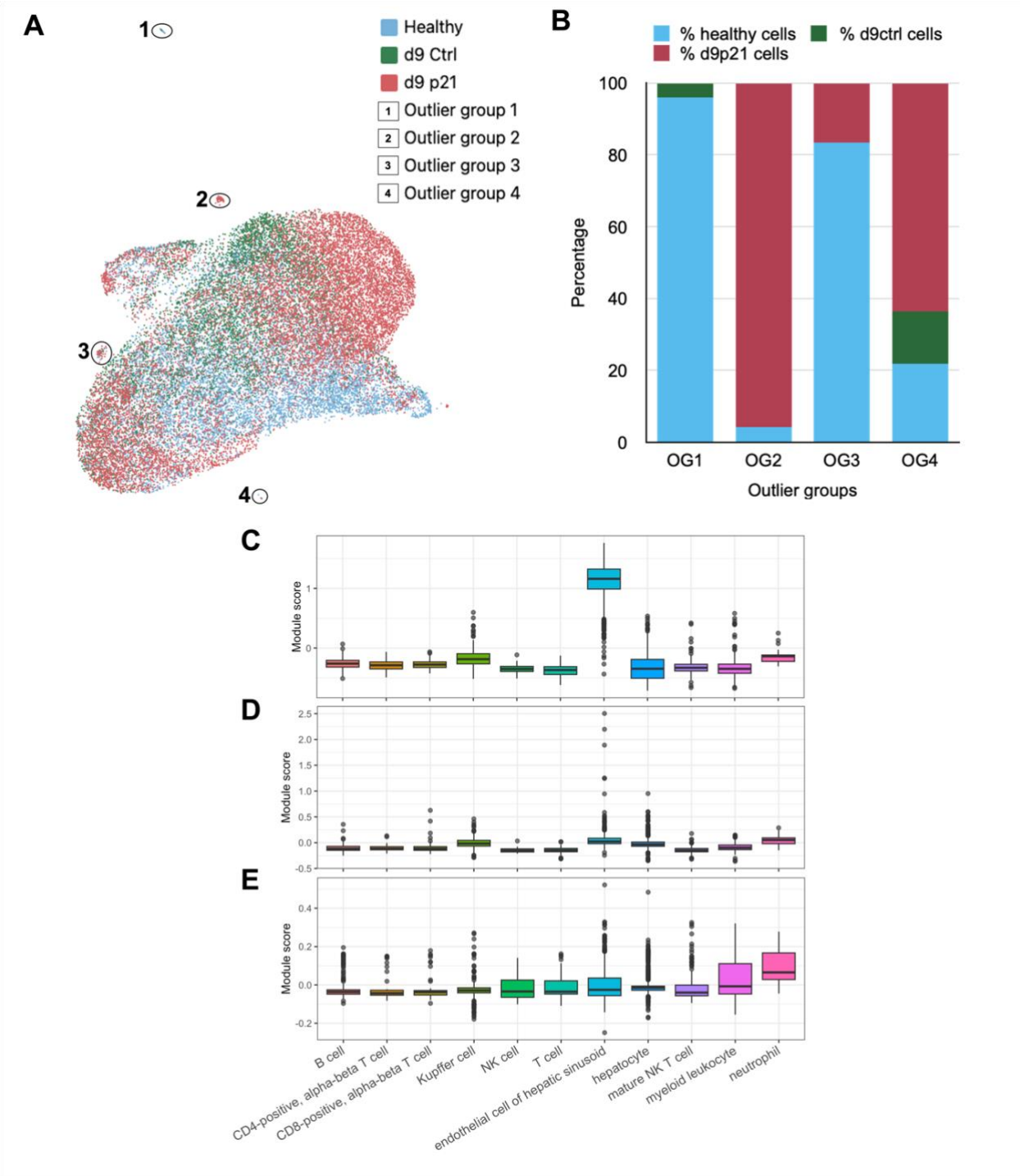


Figure 6: Analysis of outlier groups.
(**A**) Uniform Manifold Approximation and Projection (UMAP) plot showing 3 conditions (healthy (blue), d9ctrl (green), and d9p21 (red)) and 4 manually selected outlier groups (OGs, 1-4). (**B**) Group distribution within the OGs shown as percentage of cells derived from each condition. (**C-E**) Analysis of differentially expressed genes (DEGs) of OG1 (**C**), OG3 (**D**), and OG4 (**E**) for associations with specific cell types presented with a module score for each cell type. Higher module scores reflect higher expression of marker genes of a specific cell type. Module scores are visualized as boxplots.

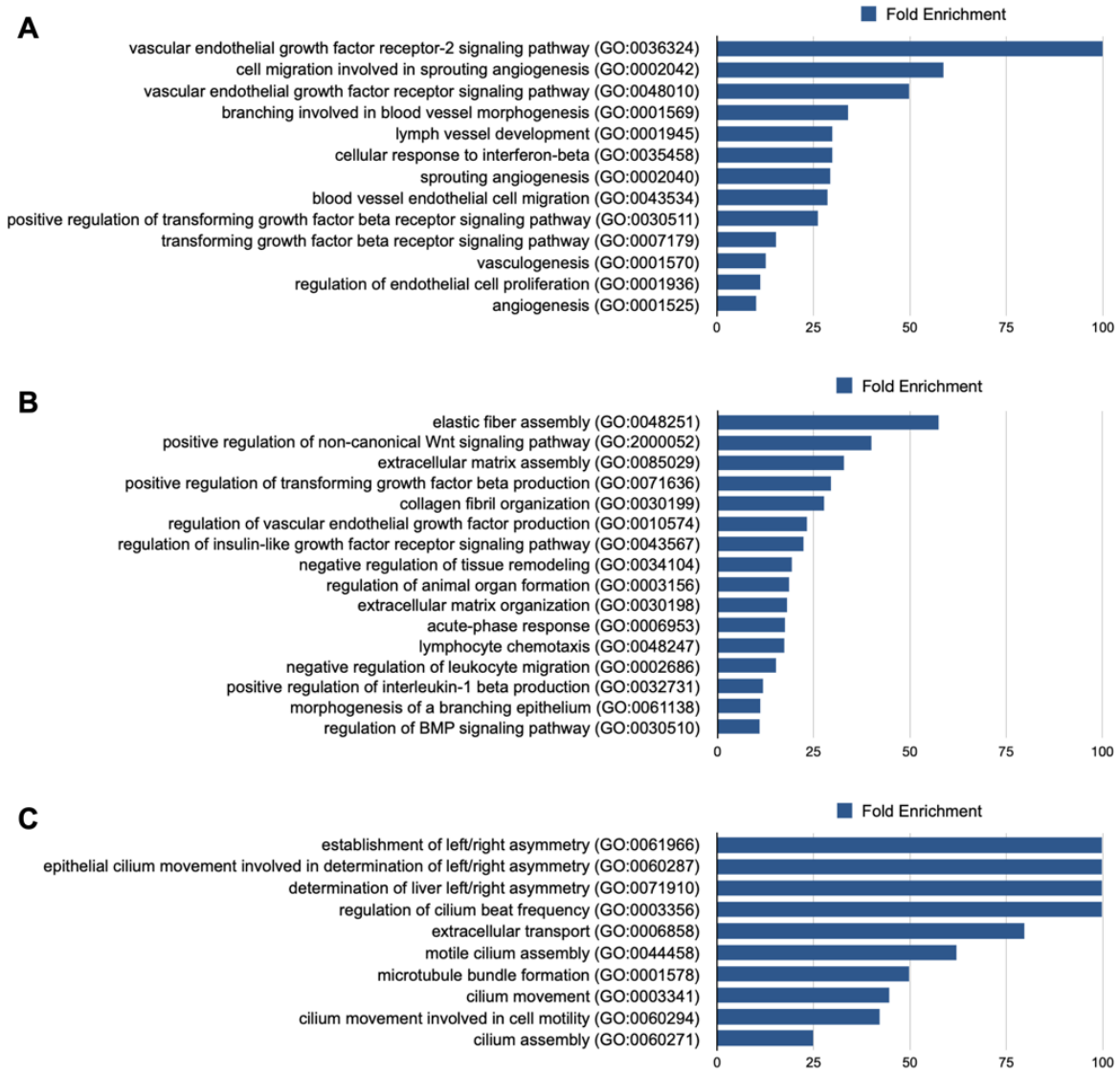


Figure 7: Functional enrichment analysis of outlier group 1, 3, and 4.

(**A**) Gene Ontology (GO) terms associated with the DEGs from outlier group (OG) 1. Blue bars: Fold Enrichment. (**B**) GO terms associated with the DEGs from OG 3. Blue bars: Fold Enrichment. (**C**) GO terms associated with the DEGs from OG 4. Blue bars: Fold Enrichment.

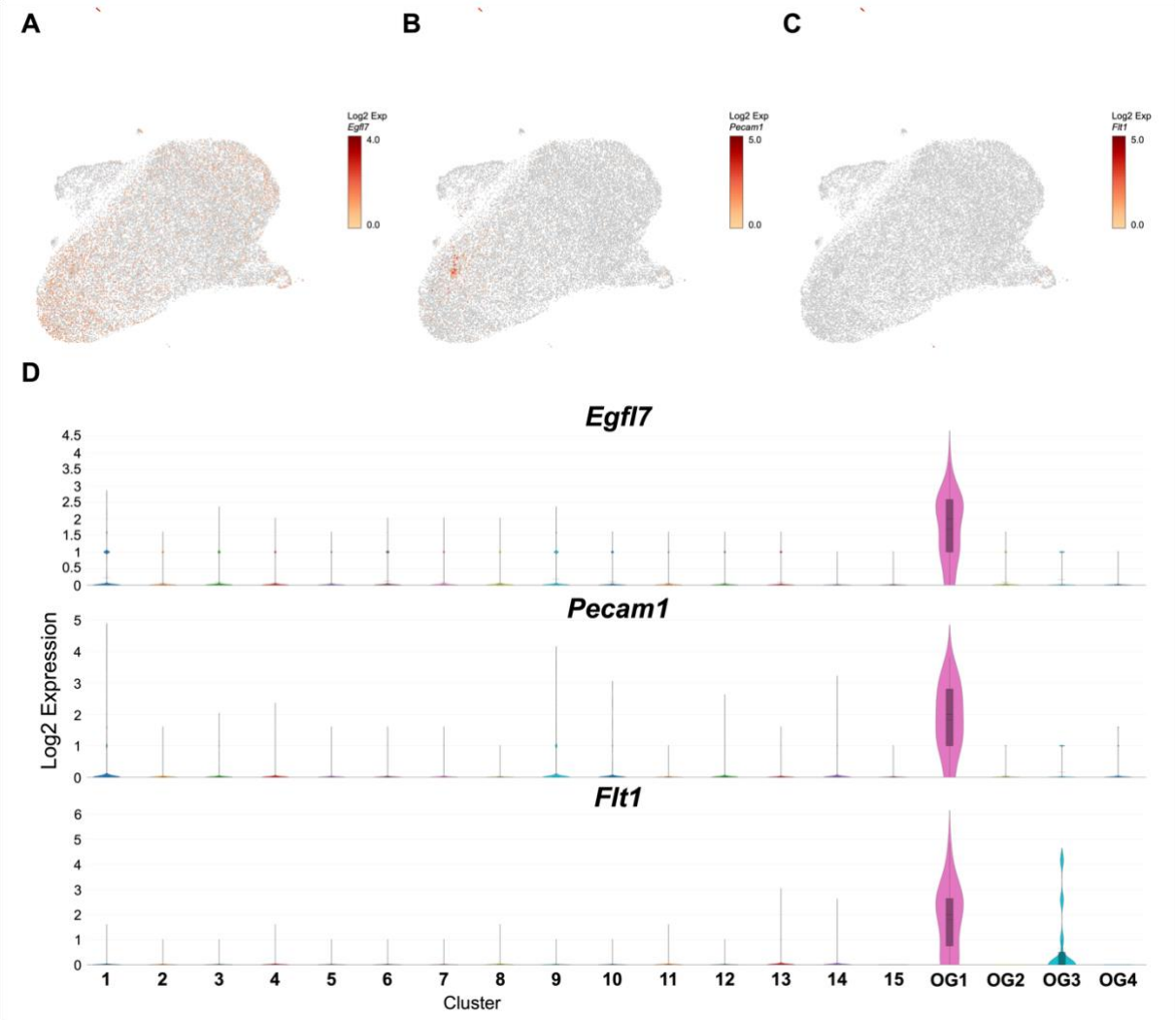


Figure 8: Gene expression of endothelial cell markers in outlier group 1.

Expression of *Egfl7* (**A**), *Pecam1* (**B**), and *Flt1* (**C**) in the scRNA-seq dataset presented in a Uniform Manifold Approximation and Projection (UMAP) plot. Colour bars indicate Log2 Expression. (**D**) Violin plots presenting the Log2 Expression of *Egfl7*, *Pecam1*, and *Flt1* across the clusters and outlier groups (OGs).

A cluster similarity analysis revealed that 55% of the DEGs in cluster 9 were also present in cluster 1. Similarly, 67% of the DEGs in cluster 5 were DEGs in cluster 13. For all other clusters, the overlap was low (**Figure 9**). Based on these similarities, I considered grouping clusters 1 and 9, and 5 and 13 together. However, when comparing gene expression between cluster 9 and 1, the DEGs for cluster 1 included *Gc, Serpina1c, Socs3, Onecut1, Ccnd1, Alb, Bmyc, Hes1, Ttr, Ctsb, Adh5, Anxa4,* and *Psenen*; and the DEGs for cluster 9 included *F3*, *Plaur, Hnf4α, Sprr1, Egf3, Cdkn1a, Cd44, Lgals3, Kitl, Fermt2*. Gene expression comparison between cluster 5 and 13 showed that the DEGs for cluster 5 included *Cxcl1, Ccl2, Cxcl2, Apoe, Icam1, Spp1, Alb, Csf3, Tnfaip2, Cxcl3, Cxcl5, Fermt2, Trp53, Myc,* and *Cdkn1a*; and the DEGs for cluster 13 included *Dmbt1, Gsto1, Gsta4, Ceacam1, Lgals3, Runx1, Prom1, Kitl, Ptk2, Krt7, Egfr*, and *Cd44*.

These DEGs for cluster 1, 5, 9, and 13 included hepatocyte, cholangiocyte, inflammatory, stemness, HPC, potential DR, and senescence marker genes that are relevant for further analysis. Therefore, I kept all 4 clusters separate in the downstream analysis.

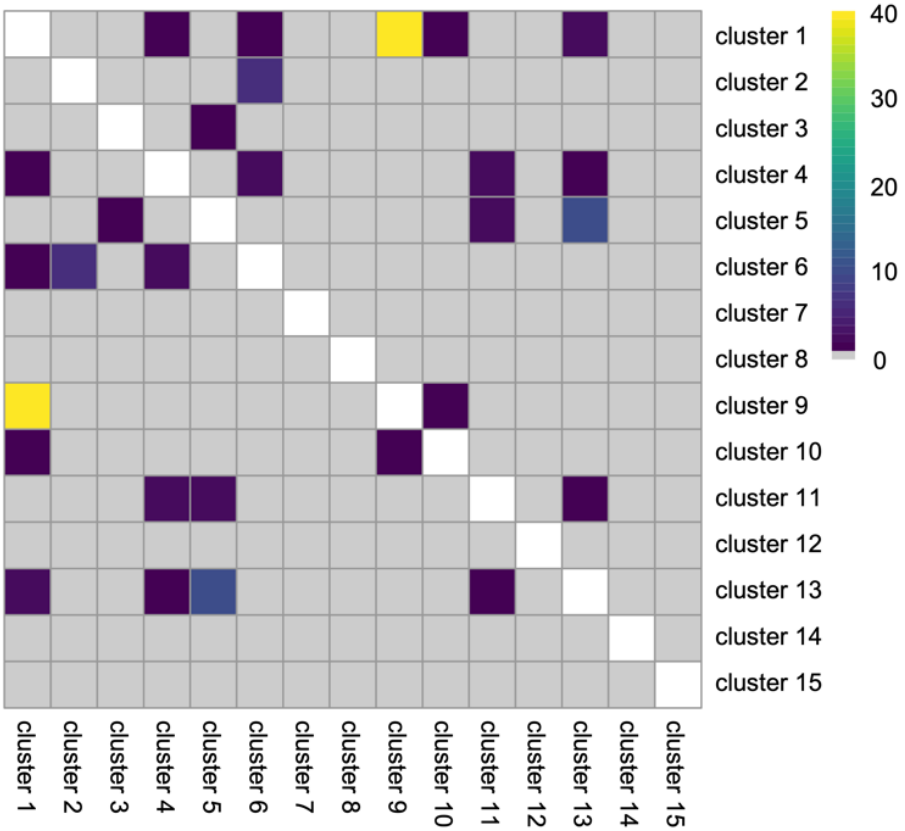


Figure 9: Cluster similarity analysis.
Heatmap showing the number of shared differentially expressed genes (DEGs) between each pair of clusters. Brighter colours indicate more shared DEGs and darker colours fewer shared DEGs. Pairs of clusters with no DEGs in common are coloured gray. Scale 0-40 represents the number of shared genes.

Altogether, cluster 14, 15, and OG1 were removed from the dataset for downstream analysis resulting in the UAMP plot presented in **Figure 10A-B**.

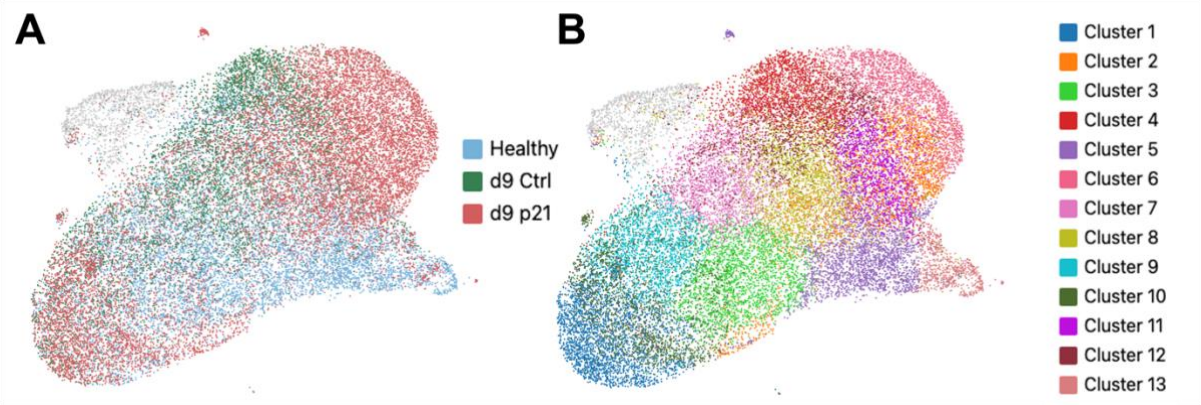


Figure 10: Groups and clusters in the scRNA-seq dataset after quality control analysis. Uniform Manifold Approximation and Projection (UMAP) plots showing the conditions (**A**) and clusters (**B**) in the scRNA-seq dataset after QC analysis. Each point represents a cell. Grey points are removed from downstream analysis.

3. Data analysis

After QC analysis, differences in gene expression between the three conditions were analyzed by determining the cluster distribution of the conditions; the DEGs for each cluster, condition, and condition within every cluster; and enrichment analysis on the DEGs for every cluster. In order to find genes or pathways that might be important for BEC-derived liver regeneration, the analyses were focused on finding significantly upregulated genes in d9p21 compared to d9ctrl.

3.1 Pseudo-bulk analysis

First, a pseudo-bulk analysis was carried out to assess the differences in gene expression between the conditions. In this report, these results are only shown for a selection of 17 genes based on the genes of interest mentioned in Chapter 2.1.3 and 5 factors described in Chapter 3.4. This analysis revealed that the DEGs with the highest Log2 Fold Change (L2FC) of expression in d9p21 compared to both d9ctrl and healthy conditions included the hepatocyte markers *Serpina1c* (encoding α-1-antitrypsin 1-3 (UniProt)) and *Gc* (encoding vitamin D binding protein (UniProt)), *Ccl2* (encoding C-C motif chemokine 2, a protein involved in chemotaxis attracting monocytes (UniProt)); *Rgs5* (encoding regulator of G-protein signaling 5, involved in signal transduction (UniProt)); *Hmox1* (encoding stress protein and metabolic enzyme heme oxygenase 1, known to catabolize free heme (Haines & Tosaki, 2020)); *Thbs1 (*encoding thrombospondin 1; a multifunctional protein implicated in angiogenesis, inflammation, wound healing, and more biological processes (UniProt)*; Cnn2* (encoding Calponin-2, involved in regulating and modulating smooth muscle contraction (UniProt)*; Tnf* (encoding tumor necrosis

factor, a cytokine primarily secreted by macrophages, which can induce cell death, and can sometimes stimulate cell proliferation and differentiation (UniProt)*; and Ccn2* (encoding CCN family member 2, a protein that mediates divalent cation- and heparin-dependent cell adhesion in various cell types including epithelial cells (UniProt)) (**Figure 11**).

For some genes (*Serpina1c, Hmox1, Cnn2, Rgs5*), the Log2FC of expression in d9p21 compared to healthy is higher than compared to d9ctrl. This might suggest that liver injury itself also plays a role in upregulating the expression of these particular genes. However, comparing d9ctrl with healthy, these genes are not significantly upregulated, suggesting that liver injury only is not sufficient to drive these changes. Both *Csf1* and *Cxcl3* are significantly upregulated in d9p21 compared to healthy but not compared to d9ctrl, suggesting that these genes may not be important for BEC-derived liver regeneration after injury and might be a result of injury alone. *Cxcl3* is also significantly upregulated in d9ctrl compared to healthy, reinforcing this hypothesis for *Cxcl3*.
The genes with the highest Log2FC of expression in d9p21 compared to d9ctrl and with no significant upregulation of gene expression in d9ctrl compared to healthy - *Serpina1c*, *Gc, Ccl2, Hmox1,* and *Rgs5* - were of most interest for further analysis.

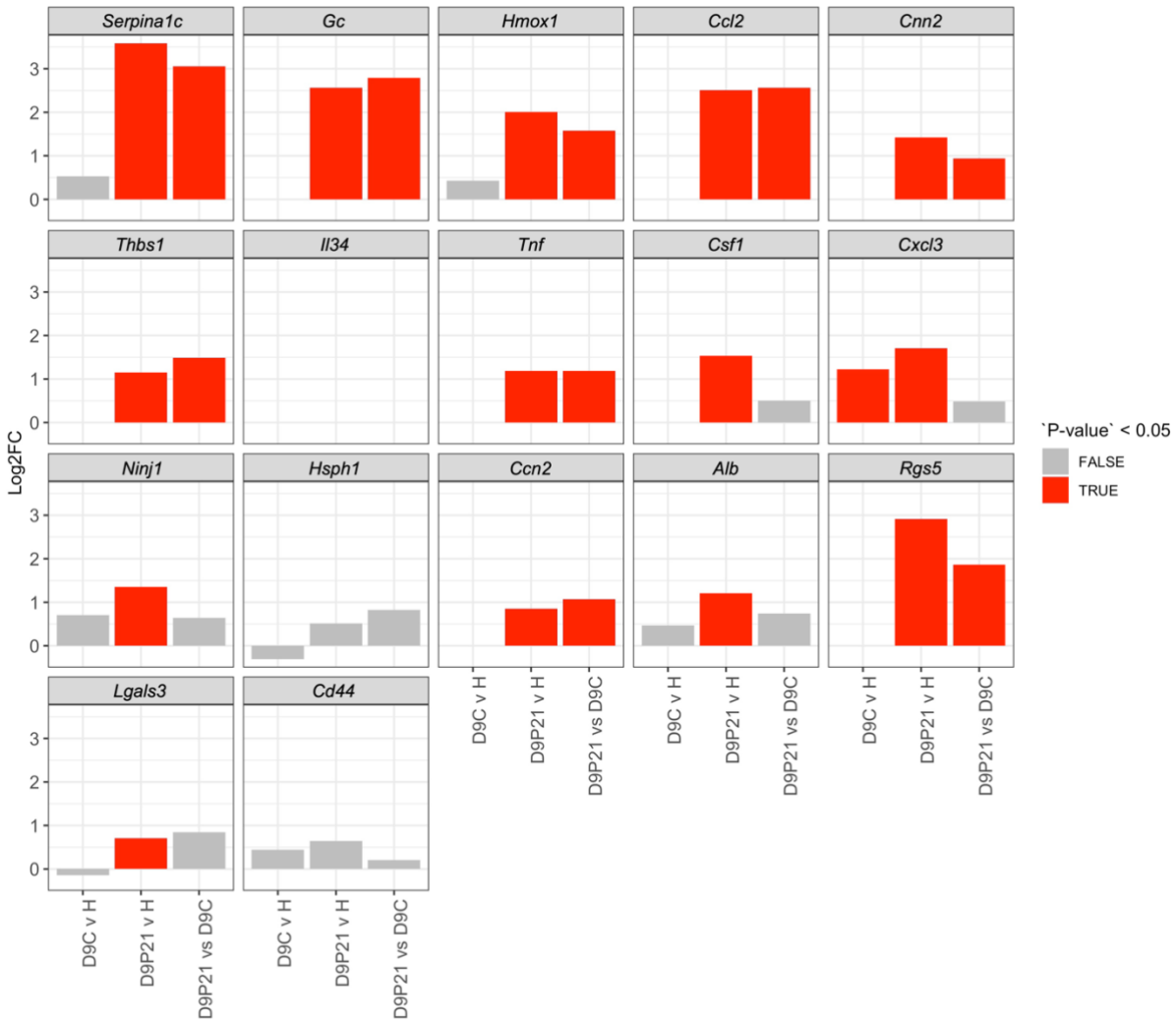


Figure 11: Pseudo-bulk analysis comparing gene expression between conditions.
Pseudo-bulk analysis comparing gene expression between conditions (healthy (H), d9ctrl (D9C), and d9p21 (D9P21) in Log2 Fold Change (Log2FC) of expression.

3.2 Group distribution

Comparison of the 3 condition-specific UMAP plots showed an area of cells (in cluster 2, 6, 10, 11, and 12) that is more abundant in d9p21, compared to healthy and d9ctrl (**Figure 12A-B**). The group distribution was determined for each cluster, to define which clusters are highly represented by each condition (**Figure 12C**). This confirmed the high representation of d9p21 cells in the clusters mentioned above.

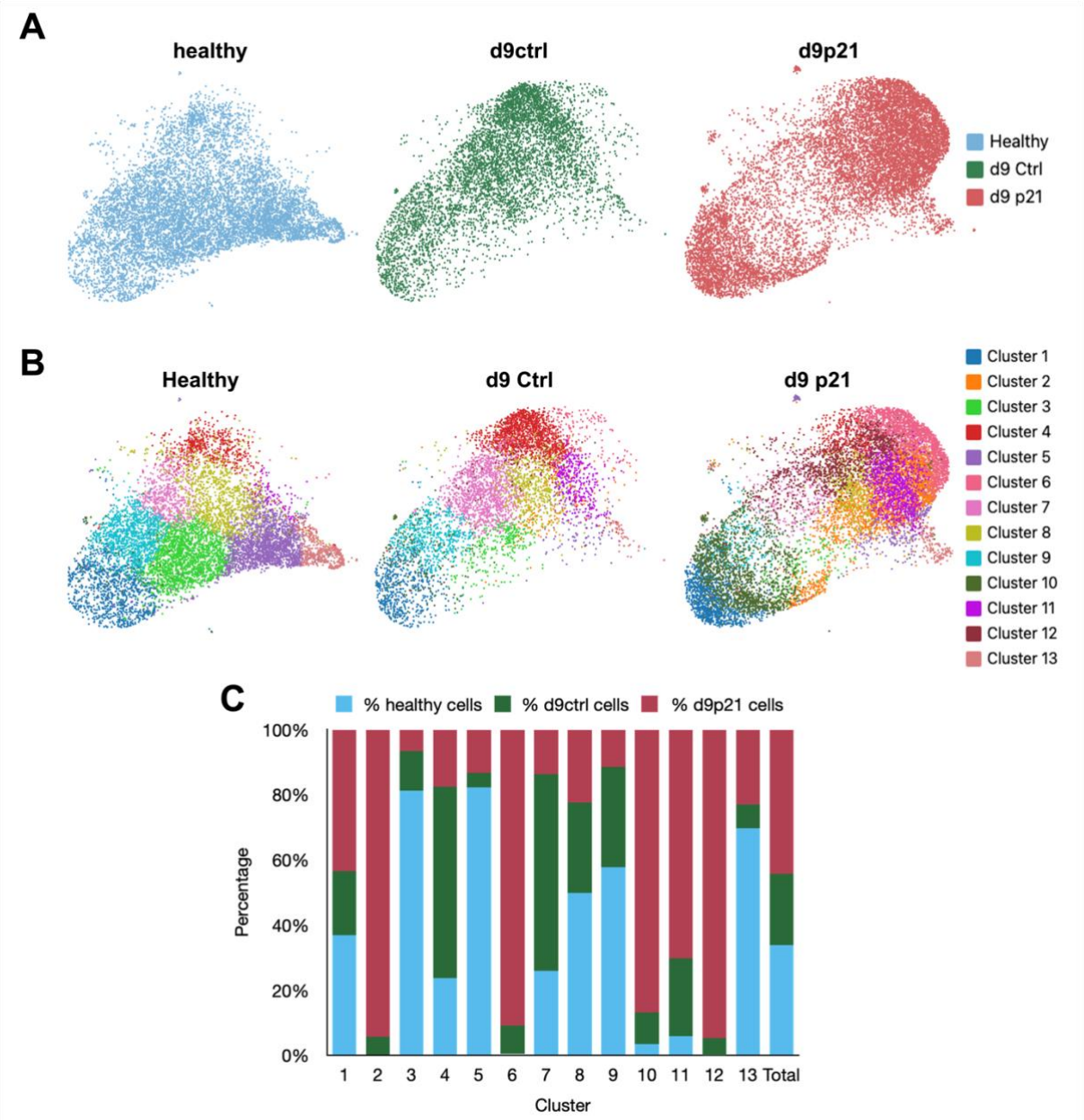


Figure 12: Group distribution in the scRNA-seq dataset.

(**A**) Individual Uniform Manifold Approximation and Projection (UMAP) plots for each condition (healthy, d9ctrl, and d9p21) in the scRNA-seq dataset. Each point represents a cell. (**B**) Individual UMAP plots for each condition (healthy, d9ctrl, and d9p21) in the scRNA-seq dataset presenting the clusters per condition. Each point represents a cell. (**C**) Group distribution within the clusters presented as percentage of cells per condition.

3.3 Gene expression and functional analysis in clusters

To determine which genes were representative of the different conditions, DEGs were determined for each cluster compared to all other clusters and for the conditions within each cluster compared to the other conditions. For a selection of 17 genes (based on the genes of interest mentioned in Chapter 2.1.3 and 5 factors described in Chapter 3.4), the

Log2FC of expression in every cluster compared to all other clusters is shown in **Figure 13A**.
Considering the clusters of interest for d9p21 – cluster 2, 6, 10, 11, and 12 – most of the selected genes are upregulated in these clusters (except for cluster 11), whereas they are mostly downregulated in the other clusters. Cluster 10, 11, and 12 only consisted of 4, 6, and 1 DEG(s) respectively and downstream enrichment analysis described below did not result in significantly enriched GO terms for these clusters. These clusters are therefore not further analyzed. Comparing gene expression of the clusters to all other clusters, *Serpina1c, Ccl2, Cnn2, and Csf1* are significantly upregulated in both cluster 2 and 6; *Hmox1, Thbs1*, and *Tnf* are significantly upregulated in cluster 2; and *Gc, Il34, Cxcl3, Ninj1*, and *Rgs5* are significantly upregulated in cluster 6. These genes are therefore of interest for further analysis. Also, the upregulation of *Serpina1c, Gc, Ccl2, Hmox1*, and *Rgs5* in the clusters of interest correspond with the most upregulated genes for d9p21 in the pseudo-bulk analysis performed above.
Surprisingly, *Serpina1c* shows a significant downregulation in cluster 11, which is in contrast with all other clusters of interest. However, cluster 11 is not included in downstream analysis by reasons explained above which is why this finding is not taken into consideration.

To understand the functional role of each cluster, enrichment analysis was performed on the DEGs of each cluster**.** Considering the clusters of interest for d9p21 (2, 6, 10, 11, 12), this resulted in a list of GO terms for cluster 2, 6, and 10 (14, 31, and 4 DEGs respectively) but not for cluster 11 and 12 (probably due to the low number of DEGs for these clusters). Analysis of cluster 10 resulted in only three GO terms, related to regulation of high voltage-gated calcium channel activity. The GO terms associated with the DEGs in cluster 2 were related to endothelial cells and angiogenesis, endothelial and epithelial cell apoptosis, endothelial cell and macrophage migration, and the immune system (**Figure 13B)**. The GO terms associated with the DEGs of cluster 6 were related to monocyte homeostasis and differentiation, macrophage migration, and cytokine-mediated signaling (**Figure 13C**). As cluster 2 and 6 mostly consist of d9p21 cells (**Figure 12C**), this suggests that these d9p21 cells play a role in vascular changes and are interacting with the immune system. Interestingly, cluster 2 seems to be mostly associated with vascular changes, whereas cluster 6 is mostly functioning in regulating the immune system. This suggests that these d9p21 cells exhibit different functions after liver injury confirming heterogeneity within the biliary epithelium after liver injury.
For the GO terms enriched in clusters 2 and 6, the DEGs associated with these GO terms were analyzed. In cluster 2, the most frequently represented genes were *Thbs1, Ccl2, Hmox1, Tnf,* and *Cnn2.* In cluster 6, these were *Csf1, Ccl2, Il34, Cnn2, Ninj1*, and *Cxcl3*.

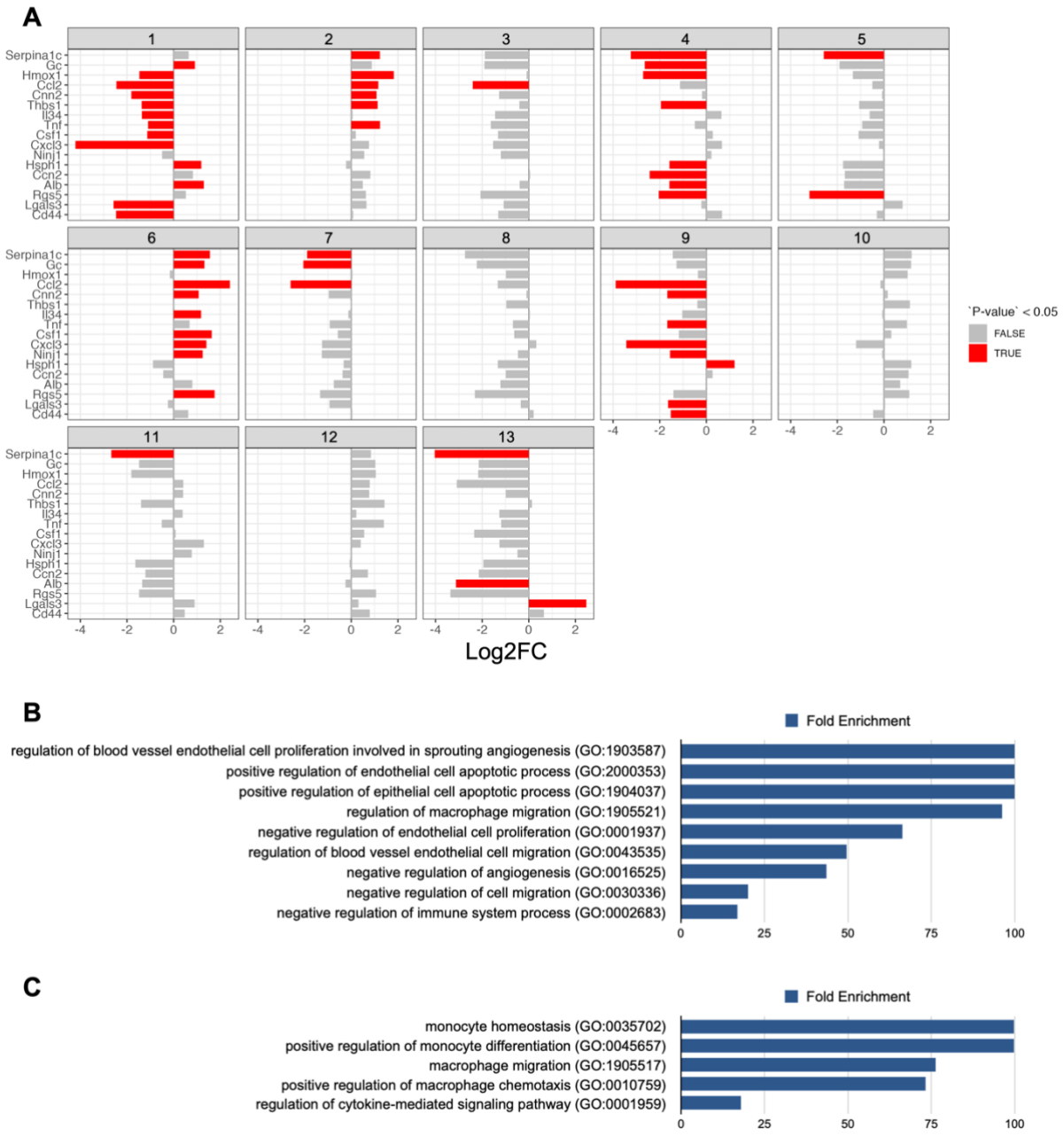


Figure 13: Gene expression of 17 selected genes in all clusters and functional analysis for cluster 2 and 6.

(**A**) Log2 Fold Change (Log2FC) of expression for 17 selected genes in every cluster compared to all other clusters. (**B**) GO terms determined for cluster 2. Blue bars: Fold Enrichment. (**C**) GO terms determined for cluster 6. Blue bars: Fold Enrichment.

Subsequently, the DEGs for each condition in comparison to the other conditions ("globally distinguishing" function in Loupe Browser) and to one other condition ("locally distinguishing" function in Loupe browser) were defined and analyzed for a given cluster. This analysis provides insight into marker genes of the d9p21 subgroups of every cluster.

Combining these results for all clusters, frequently upregulated genes in the healthy condition compared to the other conditions were mostly ribosomal and mitochondrial genes (e.g. *Gm10076, mt-Atp8, mt-Nd4l, Rps27-29, Rpl35,* and *Rpl38-39*). Only 7 genes were significantly upregulated in d9ctrl compared to the other conditions, and in no more than 1 cluster each. For d9p21, the most frequently upregulated genes were *Serpina1c, Gc. Hmox1, Phgdh* (a gene encoding a metabolic enzyme that plays a role in the L-Serine biosynthesis (Grant, 2018), and the heat shock protein (HSP) genes *Hsph1, Hspa1b, Hspa1a*, and *Hspb1*.
HSPs protect the cells from stressors by maintaining cellular protein homeostasis. They function as molecular chaperones, play a role in signaling transduction, the regulation of apoptosis, and in the cell cycle (Hu et al., 2022). It is interesting that these genes are mostly upregulated in d9p21, but not in d9ctrl whereas both conditions were treated with MCD diet. This might suggest that gene expression of these HSP genes is favorable for BEC-derived liver regeneration after liver injury. The upregulation of *Serpina1c* and *Gc* suggests that the d9p21 BECs are adopting hepatocyte characteristics and might be differentiating into hepatocytes.

Interestingly, *Serpina1c, Gc, Hsph1, Ccn2*, and *Rgs5* were significantly upregulated in d9p21 compared to both d9ctrl and healthy within a total of 6, 3, 3, 2, and 1 cluster(s) respectively (**Supplementary figure 3, panels A, B, L, M, and O**). Additionally, *Hmox1* and *Lgals3* (a potential DR gene (Hsieh et al., 2015), and one of the DEGs in cluster 13) were significantly upregulated in d9p21 compared to d9ctrl within 4 and 1 cluster(s) respectively, however, for some clusters these genes were only significantly upregulated in d9p21 compared to healthy but not compared to d9ctrl (**Supplementary figure 3, panels C and P**).
Comparing these results with the enrichment analysis described above, *Thbs1* – a gene frequently associated with the GO terms in cluster 2 - was significantly upregulated in d9p21, compared to both d9ctrl and healthy within cluster 1 (**Supplementary figure 3F**). *Hmox1* is also of interest based on both the enrichment analysis of cluster 2 and the DEG analysis comparing conditions within clusters as described above. *Cxcl3* and *Ninj1* are both interesting genes based on the enrichment analysis of cluster 6; however, comparing the conditions within clusters only showed upregulated gene expression of these genes in d9p21 compared to healthy, and not compared to d9ctrl (**Supplementary figure 3I-J**). Also, the hepatocyte marker *Albumin* - one of the DEGs in cluster 1 - was significantly upregulated in d9p21 compared to healthy within cluster 5 and 9 (**Supplementary figure 3N**). However, there is no significant difference when comparing d9p21 with d9ctrl or d9ctrl with healthy.

Furthermore, *Rgs5* is a DEG in cluster 6 and in d9p21 compared to both d9ctrl and healthy within cluster 1 (**Supplementary figure 3O**). Finally, *Cd44* (a previously proposed HPC marker (H. Y. Kim et al., 2022; Kon et al., 2006), which is involved in Wnt/β-catenin signaling (Schmitt et al., 2015) was significantly upregulated in d9p21 compared to healthy within cluster 1, but not compared to d9ctrl (**Supplementary figure 3Q**).

3.4 Selection of genes for in vitro validation

Due to a limited amount of time for this project, only 3 genes could be selected for *in vitro* validation. First, 17 genes of interest (see **Chapter 3.3.3**) were selected based on the genes of interest defined in Chapter 2.1.3 and 5 determinant factors:

(i) The DEGs in the clusters of interest (2 and 6).
(ii) The DEGs in d9p21 compared to the other conditions, in both the pseudo-bulk analysis and within clusters.
(iii) The level of expression of the genes in the clusters and conditions described above.
(iv) The genes that mostly represented the selected GO terms in the enrichment analysis.
(v) The frequency of the DEGs being of interest when combining the different analyses described here.

Based on these 5 factors alone, the gene selection for *in vitro* validation was narrowed down to 6 genes. The 6 most interesting genes were *Serpina1c, Gc, Hmox1, Ccl2, Cnn2*, and *Thbs1*. For these genes, the expression in the different conditions and clusters is shown in **Figure 14 and 15**. As described above, the difference in gene expression between d9p21 and d9ctrl was most evident for *Serpina1c, Gc,* and *Ccl2*. The difference in gene expression between the clusters of interest (2, 6, 10, 11, and 12) and the other clusters is also most prominent for these 3 genes.

Figure 14: Gene expression pattern of *Serpina1c, Gc, Hmox1, Ccl2, Cnn2*, and *Thbs1* in the experimental groups.

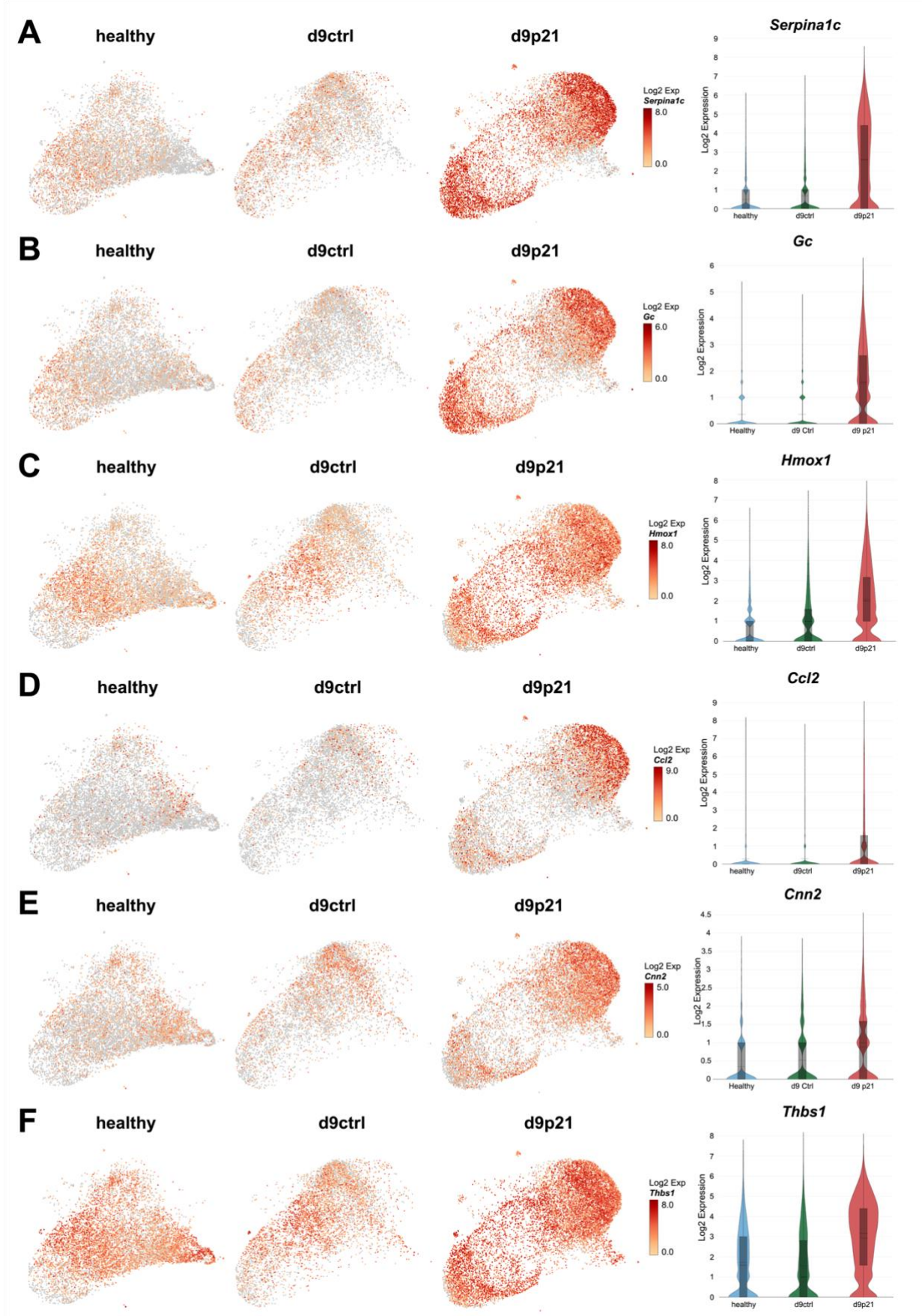


Figure 14: Gene expression pattern of *Serpina1c, Gc, Hmox1, Ccl2, Cnn2,* and *Thbs1* in the experimental groups.

Log2 Expression of *Serpina1c* (**A**), *Gc* (**B**), *Hmox1* (**C**), *Ccl2* (**D**), *Cnn2* (**E**), and *Thbs1* (**F**) in the healthy, d9ctrl, and d9p21 conditions presented in Uniform Manifold Approximation and Projection (UMAP) plots (left) and violin plots (right).

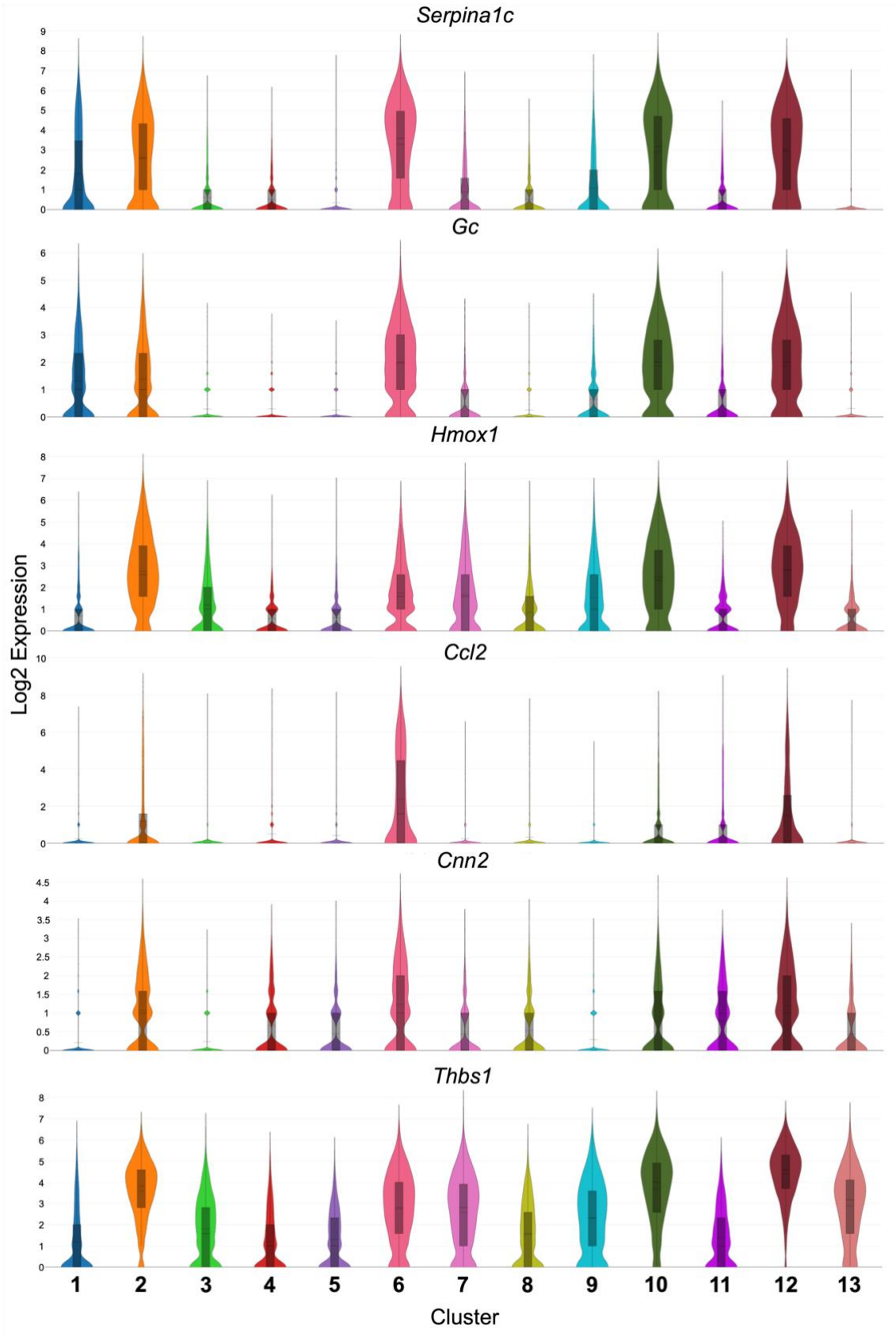


Figure 15: Gene expression pattern of *Serpina1c, Gc, Hmox1, Ccl2, Cnn2,* and *Thbs1* across clusters.

Log2 Expression of *Serpina1c*, *Gc*, *Hmox1*, *Ccl2*, *Cnn2*, and *Thbs1* across the clusters of the scRNA-seq dataset presented in violin plots.

Next, the functions of these genes, their protein interactions, and their biological pathways were explored.

Serine (or cysteine) peptidase inhibitor, clade A, member 1C

Serpina1c or Serine (or cysteine) peptidase inhibitor, clade A, member 1C encodes for α-1-antitrypsin 1-3, which belongs to the serpin family and is known as a hepatocyte marker (NCBI, Bgee). α-1-antitrypsin 1-3 is a serine protease inhibitor which can inhibit trypsin and chymotrypsin and is involved in "cholesterol metabolism", "SNARE interactions in vesicular transport", "autophagy", and "complement and coagulation cascades" (STRING). The human ortholog of *Serpina1c* is SERPINA1, serpin family A member 1 (MGI). The murine α-1-antitrypsin is represented by 6 *Serpina1* genes whereas the human α-1-antitrypsin is only represented by SERPINA1 (Heit et al., 2013).

Human α-1-antitrypsin is mainly expressed in hepatocytes and secreted into the bloodstream to prevent proteolytic degradation of lung tissue caused by neutrophil elastase (S. Janciauskiene et al., 2018; S. M. Janciauskiene et al., 2011). α-1-antitrypsin deficiency (AATD) is a common genetic disorder in humans that is caused by a mutation resulting in misfolding of α-1-antitrypsin (Strnad et al., 2020). AATD patients are therefore predisposed to lung and liver damage (Strnad et al., 2020). This highlights the importance of a well-functioning α-1-antitrypsin in humans and might explain why restoring this function happens early in hepatic repair.

GO terms enriched in cluster 2 and 6 that were associated with *Serpina1c* were "protease binding", "(negative regulation of) serine-type endopeptidase inhibitor activity", "response to peptide hormone", and "response to cytokine".

Protein interaction analysis showed that the proteins of *Serpina1c, Stx17, Scfd1, Bet1, Proc, Apoa1, Alb, Ambp, Akt1,* and *Plg* were the most direct predicted functional partners of α-1-antitrypsin 1-3 (*Serpina1c)* (STRING). This means that these proteins are most likely interacting with each other based on different types of evidence (STRING). In the scRNA-seq dataset, *Proc* (encoding for vitamin K dependent protein C) and *Alb* (Albumin) were both DEGs in cluster 1 compared to all other clusters, and *Alb* is also significantly upregulated in d9p21 compared to healthy in both the pseudo-bulk analysis (**Figure 11**) and within cluster 5 and 9 (**Supplementary figure 3N**).

Vitamin D-binding protein

Gc encodes for vitamin D-binding protein (VDBP) and is part of the albumin family. It is known as a hepatocyte marker (Bgee) and plays a role in in "vitamin D transport and storage", enhancing the chemotaxis of neutrophils in inflammation and activating macrophages, and "scavenging of extracellular G-actin" (UniProt, STRING). The human ortholog of *Gc* is GC, which encodes for a similar protein (MGI).

In humans, not one person with absent VDBP had been identified until the New England Journal of Medicine described one patient with complete VDBP deficiency caused by homozygous *GC* deficiency(Bouillon & Pauwels, 2018; Chun, 2012; Henderson et al., 2019). This resulted in normocalcemia, mild disruption of bone metabolism, and was complicated by severe autoimmune disease (Henderson et al., 2019).

GO terms enriched in cluster 1 and 6 that were associated with *Gc* were "vitamin transport", "vitamin transmembrane transporter activity", "vitamin D metabolic process", and "calcidiol binding". Protein interaction analysis revealed that the proteins encoded by *Lrp2, Vdr, Ahsg, Cubn, Cyp2r1, Fgg, Hrg, Serpinc1* (Serine (or cysteine) peptidase inhibitor, clade c, member 1), and *Hpx* were the most predicted direct functional partners of vitamin D-binding protein (STRING). Of these genes, *Ahsg* (Alpha-2-HS glycoprotein) was also a DEG in cluster 1 compared to all other clusters of the single-cell transcriptomics dataset.

Heme oxygenase 1

Hmox1 encodes for heme oxygenase 1. It is mostly expressed in the spleen but can also be expressed in multiple other organs including the liver (Bgee). It catalyzes the oxidative cleavage of heme resulting in biliverdin which is subsequently converted to bilirubin (UniProt, STRING). It is involved in "porphyrin and chlorophyll metabolism", "ferroptosis", "mineral absorption", and "hepatocellular carcinoma" (STRING). The human ortholog of Hmox1 is HMOX1, which encodes for a similar protein (MGI, UniProt).
GO terms enriched in cluster 2 that were represented by *Hmox1* included "regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis", "positive regulation of epithelial cell apoptotic process", "regulation of blood vessel endothelial cell migration", "regulation of angiogenesis", "regulation of vasculature development", and "negative regulation of immune system process".
Protein interaction analysis revealed that the proteins encoded by *Fech, Blvra, Cav1, Heph, Blvrb, Cp, Nfe2l2, Fxn, Cox10*, and *Nqo1* were predicted as most direct functional partners of heme oxygenase 1 (STRING). In the single-cell transcriptomics dataset, *Cav1* (Caveolin-1; plays a role in T-cell receptor mediated T-cell activation) was also significantly upregulated in d9p21 compared to both d9ctrl and healthy.

C-C motif chemokine 2

Ccl2 encodes for C-C motif chemokine 2. It is mainly expressed in endothelial cells of the lymphatic vessels and usually not in the liver (Bgee). It is a ligand for C-C chemokine receptor CCR2 (UniProt). Binding CCR2 induces a chemotactic response attracting monocytes, but not neutrophils or eosinophils (UniProt). C-C motif chemokine 2 is involved in "viral protein interaction with cytokine and cytokine receptor", the "IL-17-signaling pathway", the "TNF signaling pathway", "Apoptosis", and the "Toll-like receptor signaling pathway" (STRING). The human ortholog is CCL13 which has a similar function as *Ccl2* (MGI, UniProt).
GO terms enriched in cluster 2 and 6 that were represented by *Ccl2* were "macrophage migration", "positive regulation of macrophage chemotaxis", "regulation of granulocyte chemotaxis", "regulation of leukocyte chemotaxis", and "monocyte homeostasis".
Protein interaction analysis showed that *Ccr2, Ccr5, Ccl7, Ccr1, Jun, Cxcl10, Il6, Ccr3, Cxcl1,* and *Tnf* were predicted as most direct functional partners of C-C motif chemokine 2 (STRING). Interestingly, *Jun* (a transcription factor) and *Cxcl10* (a proinflammatory cytokine) were DEGs in cluster 1 and 9 compared to all other clusters of the scRNA-seq dataset, and *Tnf* (a cytokine mainly secreted by macrophages and a potent pyrogen) was

significantly upregulated in cluster 2 compared to all other clusters and in d9p21 compared to both d9ctrl and healthy (**Figure 11**).

Calponin-2

Cnn2 encodes for Calponin-2, which is a thin filament-associated protein that plays a role in regulating and modulating smooth muscle contraction and can bind to calmodulin, troponin C, and tropomyosin (UniProt). *Cnn2* is involved in "hypertrophic and dilated cardiomyopathy", "adherens junctions", and "vascular smooth muscle contraction" (STRING). Its' expression is therefore highest in smooth muscle and tissues containing a significant amount of smooth muscle (Bgee). *Cnn2* can also be expressed in the liver (Bgee). The human ortholog is CNN2 which exhibits a similar function as *Cnn2* (MGI, UniProt).

GO terms enriched in cluster 2 and 6 that are represented by *Cnn2* were "regulation of macrophage migration", "negative regulation of cell migration and cell motility", and "negative regulation of immune system process".

Protein interaction analysis revealed that proteins encoded by *Smtn, Utrn, Myh11, Cald1, Tpm1, Flna, Acta2, Actg1, Lqgap1,* and *Tpm4* were predicted as most direct functional partners of Calponin-2. Of these genes, *Acta2* (Actin) was also a DEG in OG3 compared to all other cells in the scRNA-seq dataset.

Thrombospondin 1

Thbs1 encodes for thrombospondin 1, which is an adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions (STRING, UniProt). *Thbs1* is multifunctional as it is involved in angiogenesis, inflammation, wound healing, reactive oxygen species (ROS) signaling, nitrous oxide (NO) signaling, apoptosis, stemness, and senescence (UniProt). *Thbs1* is expressed in multiple tissues including the ovaries, limbs, lungs, bladder (Bgee, NCBI). It is usually not expressed in the liver, but it is known to be highly expressed in fetal liver hematopoietic progenitor cells (Bgee). The human ortholog is THBS1 which encodes for a similar protein as *Thbs1* (MGI, UniProt).

GO terms enriched in cluster 2 that were associated with *Thbs1* were "regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis", "positive regulation of endothelial cell apoptotic process", "positive regulation of epithelial cell apoptotic process", "regulation of macrophage migration", "regulation of blood vessel endothelial cell migration", "negative regulation of angiogenesis", "negative regulation of cell migration and motility", and "negative regulation of immune system process".

Protein interaction analysis showed that *Cd47, Fn1, Calr, Sdc4, Cd36, Lrp8, Nlgn1, Itgb1, Mmp9*, and *Ltbp1* were predicted as most direct functional partners of thrombospondin 1.

The protein interaction analysis of the 6 selected genes described here showed some predicted functional partners that were also significantly upregulated in some clusters compared to all other clusters. This suggests that the different clusters of BECs might be interacting with each other.

Based on the above information, 3 genes were selected for *in vitro* validation. The 3 most interesting genes based on the Log2FC of gene expression between conditions and clusters (**Figure 14-15**) were *Serpina1c, Gc,* and *Ccl2*.
Serpina1c and *Gc* are both hepatocyte markers and carry out important hepatocyte functions. It could be that these genes are upregulated in d9p21 compared to the other conditions as a result of BECs that are differentiating into hepatocytes after injury, which in turn would also suggest that the functions of both genes might be of high importance for the liver to be restored. However, these genes could also play a different role in the d9p21 HPCs that might be important to promote liver regeneration by BECs. For example, the high expression of both genes in cluster 2 and 6 (**Figure 14-15**) together with their associated GO terms (**Figure 13B-C**) and the known immunomodulatory functions of both genes described above might suggest a role for *Serpina1c* and *Gc* in interacting with the vasculature and immune system after liver injury. This could be contributing to such a potential role in liver regeneration.
Ccl2 is usually not expressed in the liver and plays a role in inflammation by attracting monocytes. This corresponds to the GO terms in cluster 2 and 6 associated with *Ccl2*. It is therefore interesting to find out the potential role of *Ccl2* in HPCs in the context of liver injury, and if upregulation of *Ccl2* can enhance liver regeneration.

To choose the most promising genes for *in vitro* validation, it was important to keep in mind the restrictions of *in vitro* experiments as there is no immune system, vasculature system, or other environmental systems of the body to respond to changes in gene expression in HPCs. The function of genes such as *Ccl2, Hmox1, Cnn2*, and *Thbs1* were therefore more limited to study *in vitro* than *Serpina1c* and *Gc*. However, by studying the role of these genes in HPCs *in vitro*, it is possible to find out if these genes play a role within HPCs and their surrounding neighbors that may be relevant to enhance their inherent regenerative capacity.
Considering both the data analysis results and the known functions of the genes, *Serpina1c, Gc*, and *Ccl2* were most of interest and were therefore selected for *in vitro* experiments to study their function in HPCs.

Chapter 4 - CRISPR-editing of *Serpina1c*, *Gc*, and *Ccl2*

For *in vitro* validation of the findings of the bioinformatics analysis, I developed a CRISPR-Cas9 knockout (KO) strategy, separately targeting *Serpina1c, Gc*, and *Ccl2.* The goal of this experiment was to study the role of these genes in hepatic progenitor cells (HPCs) in homeostasis, by indirectly eliminating their functional effects. We were particularly interested in the effects of these genes' ablation on the expression of hepatocyte and cholangiocyte phenotypes, proliferation, stemness, senescence, and inflammation.

1. Lentivirus transduction of HPCs

3 gRNA sequences were selected for each gene (*Serpina1c, Gc*, and *Ccl2*). Target sequence cloning was performed for each gRNA using a Lentiviral vector (LentiCRISPRv2). However, due to contamination and incorrectly ligated plasmids of some of the colonies, subsequent lentivirus production was only carried out for *Serpina1c* (guide 1-3), *Gc* (guide 3), and *Ccl2* (guide 3). The lentivirus carries a GFP-fluorescent reporter that allows us to identify the transduced cells of interest by GFP-positive (GFP+) expression. The LentiCRISPRv2 backbone was designed so that - after transduction - GFP, Cas9, and the gRNA are integrated in the host genome and expressed from the nucleus of the cell. This means that GFP+ cells theoretically represent successfully transduced cells.
Additionally, the lentivirus backbone also contained Bleomycin resistance allowing us to select the transduced cells with antibiotics.

Mouse HPCs were transduced with different dilutions of lentivirus (see **Table 3**) in order to find the optimal dilution for a successful transduction, as assessed by the presence of GFP+ labelled cells.
To determine if transduction was successful, the cells were therefore observed in a microscope to find green GFP+ cells. Additionally, recovery and expansion of the cells was analyzed over the course of passages to select the most viable cells for sorting and subsequent *in vitro* experiments.

At first, HPCs were transduced with *Serpina1c* (guide 1-3), *Gc* (guide 3), and scramble virus diluted at 1/100 and 1/500 (**Figure 16, 17,** and **18**; all 1/100). Polybrene was used as adjuvant to improve transduction efficiency. In addition, polybrene was assessed independently to determine potential toxicity.
The transductions with virus diluted at 1/500 yielded no GFP+ HPCs (data not shown). At day 7, HPCs transduced with *Gc* (guide 3) diluted in 1/100 showed GPF+ cells, but not many cells survived the transduction.

HPCs transduced with *Serpina1c* (guide 1) and *Serpina1c* (guide 2) diluted in 1/100 both survived the transduction, but no GFP+ was observed.
Surprisingly, HPCs that received only polybrene did not expand. Potential explanations are technical errors (wells containing less cells before adding polybrene compared to the other wells; inadvertently adding more polybrene) or that the selected image is not fully representative of the complete well. However, as part of the cells still look viable and most of the transduced cells recovered well from transduction, it is unlikely that polybrene-induced toxicity is responsible for this phenomenon.

Transduction with *Serpina1c* (guide 3) and scramble diluted in 1/100 resulted in GFP+, viable and growing cells. These HPCs were therefore expanded and FACS-sorted for GFP+, using DRAQ7 as a marker of cell death (**Figure 19**). FACS-sorting of *Serpina1c* (guide 3) transgenic and scramble HPCs resulted in 56,2% and 62,6% GFP+, DRAQ7- cells respectively. For both resulting cell lines, a higher percentage of GFP+ cells was observed in comparison with the unsorted HPCs population (**Figure 17A-B and 18A**). After culture, on day 20, clusters of *Serpina1c* (guide 3) KO cells started to detach and part of these cells did not survive. However, the HPCs reattached to the plate as shown on day 26. The same pattern was observed again after 1-2 weeks (day 35 after transduction). To date, due to the time limitations of this project, it is unknown if the cells keep repeating the same pattern over the course of time or if this represents a unique event.
However, HPCs transduced with scramble virus showed the same pattern of detaching at day 32 after passage 21 (P21). At a similar timepoint, the control HPCs were also detaching (also at P21, data not shown). This repeated pattern in all three lines may indicate an intrinsic phenotype of the HPCs in culture over the course of time. Further experiments are required to understand if this phenomenon is transduction dependent.

Transduction of HPCs with *Serpina1c* (guide 1-2) and *Gc* (guide 3) was repeated to improve survival and the percentage of GFP+ cells. HPCs were transduced with *Serpina1c* (guide 1-2) diluted at 1/50 and *Gc* (guide 3) diluted at 1/250 (**Figure 20, 21, and 22**). These transductions were performed twice in HPCs after passage 19 and 20 (P19, P20) respectively. *Serpina1c* (guide 1) transduction resulted in a few GFP+ cells after both transductions, with minimal survival and failed cell expansion.
Interestingly, P19 HPCs transduced with *Serpina1c* (guide 2) and *Gc* (guide 3) showed a similar detachment pattern as above (**Figure 21A and 22A**). Both *Serpina1c* (guide 2) and *Gc* (guide 3) transgenic HPCs showed a few GFP+ cells before day 26, potentially indicating that the transduction had improved in comparison with the 1/100 dilution.
Transduction of *Serpina1c* (guide 1-2) and *Gc* (guide 3) in P20 HPCs resulted in a few GFP+ cells, but these HPCs did not recover from transduction (**Figure 20B, 21B, and 22B**).
Again, due to time limitations and a low expansion rate to reach confluency, I did not manage to get a viable culture within the time frame of my project.

***Ccl2* (guide 3) transductions** were performed at the same time as the *Serpina1c* (guide 1-2) and *Gc* (guide 3) P20 transductions and 4 different virus dilutions were added to

HPCs (**Figure 23**). Transduction with 1/50 dilution resulted in a very low survival rate and HPCs were therefore discarded between day 7-13. Transduction with 1/100 and 1/250 dilutions resulted in GFP+ HPCs and a good survival rate compared to 1/50. However, the observed phenotype of the HPCs at all dilutions (dark, spheric) was completely different from control HPCs and similar to the P20 HPCs transduced with *Serpina1c* (guide 1-2) and *Gc* (guide 3).
Around day 13 after *Ccl2* (guide 3) transduction, I observed the pattern of detachment and death in the transgenic, scramble control, and non-transduced control HPCs as described above. It is therefore unclear if the lower recovery rate and change in phenotype of the *Serpina1c* (guide1-2), *Gc* (guide 3), and *Ccl2* (guide 3) transgenic cells is caused by the KO of the genes, by a decreased viability of the HPCs at later transductions, or by something else. The *Ccl2* (guide 3) transduced HPCs did not recover and expand well enough to FACS-sort the GFP+ cells within the time frame of this project but I would suggest that the 1/100 and 1/250 are the most promising dilutions for a successful transduction protocol.

Altogether, the HPCs transduced with *Serpina1c* (guide 3) and scramble diluted in 1/100 were GFP+ and expanded adequately, indicating that this virus dilution results in viable transgenic *Serpina1c* (guide 3) and scramble control HPCs that can be used for *in vitro* experiments.
FACS-sorted cell lines, some of the other transduced P19 HPCs (*Serpina1c* (guide 1-2) diluted in 1/50 and *Gc* (guide 3) diluted in 1/250), and non-transduced control HPCs suddenly detached (and died) at various time points after transduction, which limited the yield of GFP+ viable cells. As this pattern was observed in almost all cell lines - including non-transduced control HPCs - and the latest transduced cells show a lower recovery and expansion rate together with an altered phenotype, it is likely that HPCs show reduced viability over the course of culture. For future directions, I would therefore suggest to repeat all transductions in fresh viable HPCs (at low passages) using the most promising virus dilutions (**Table 4**)
The resulting *Serpina1c* (guide 3) and scramble HPCs' lines were kept in culture and were subsequently used for RT-qPCR.

Table 4: Most promising virus dilutions for HPC transduction

gRNA	dilution
***Serpina1c* guide1**	1/50
***Serpina1c* guide 2**	1/50
***Serpina1c* guide 3**	1/100
***Gc* guide 3**	1/250
***Ccl2* guide 3**	1/100 and 1/250
Scramble	1/00

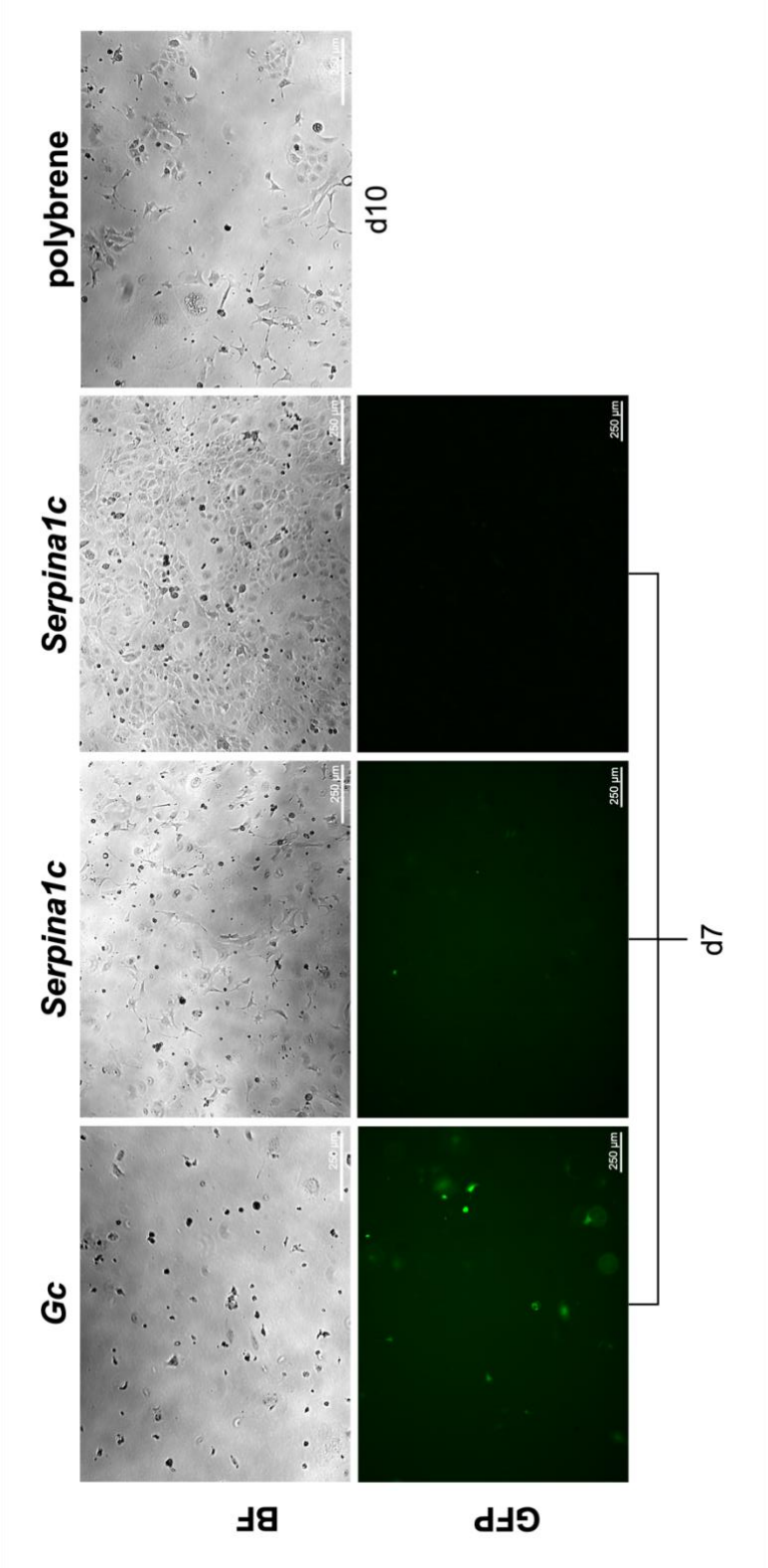
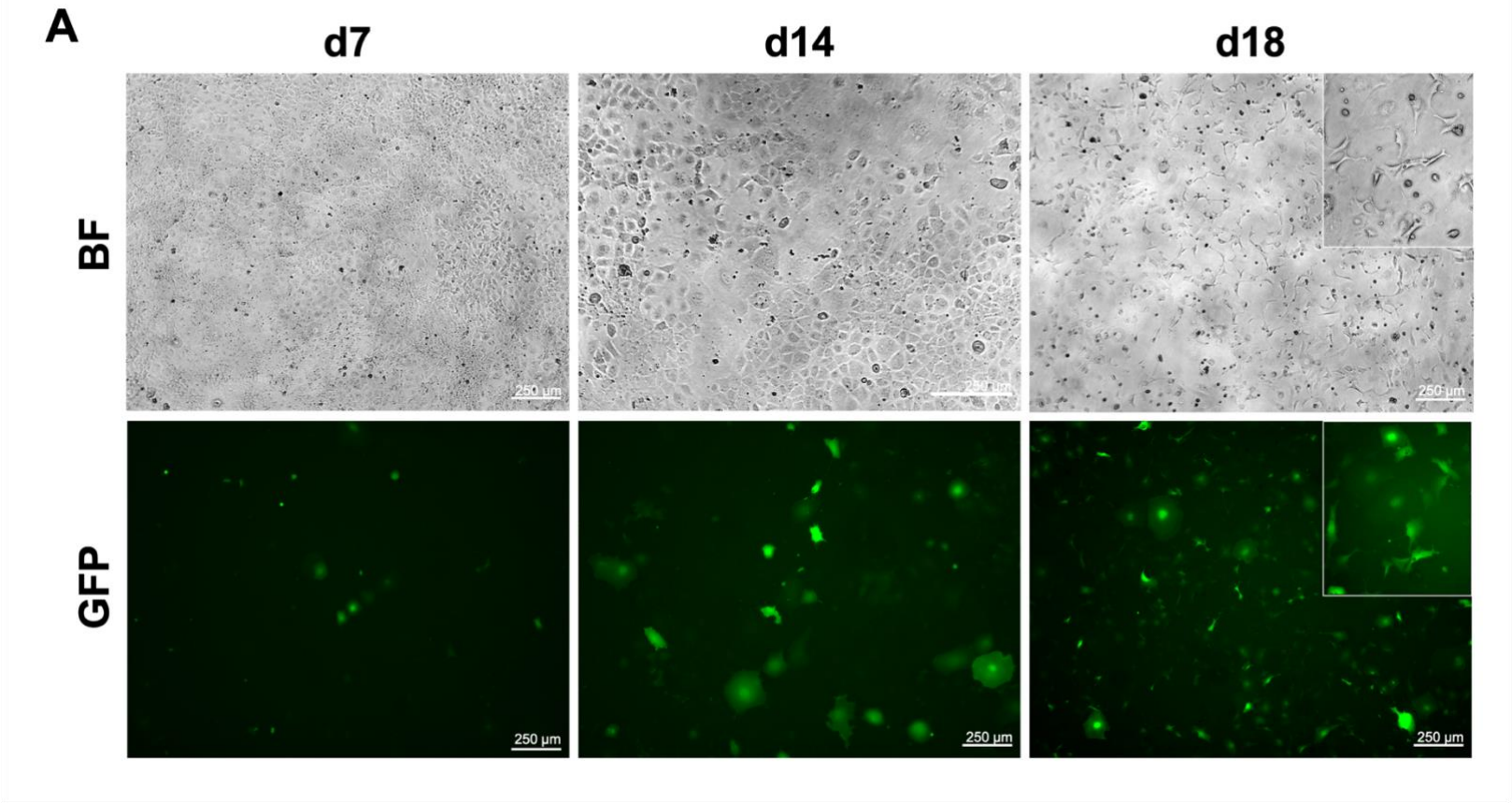


Figure 16: Transduction of HPCs with *Gc* (guide 3) and *Serpina1c* (guide 1-2) (1/100), and polybrene control.

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with *Gc* (guide 3) and *Serpina1c (*guide 1, left and guide 2, right) at day 7. Polybrene control at day 10. Scale bars, 250 µm.

Figure 17: Transduction of HPCs with scramble (1/00)



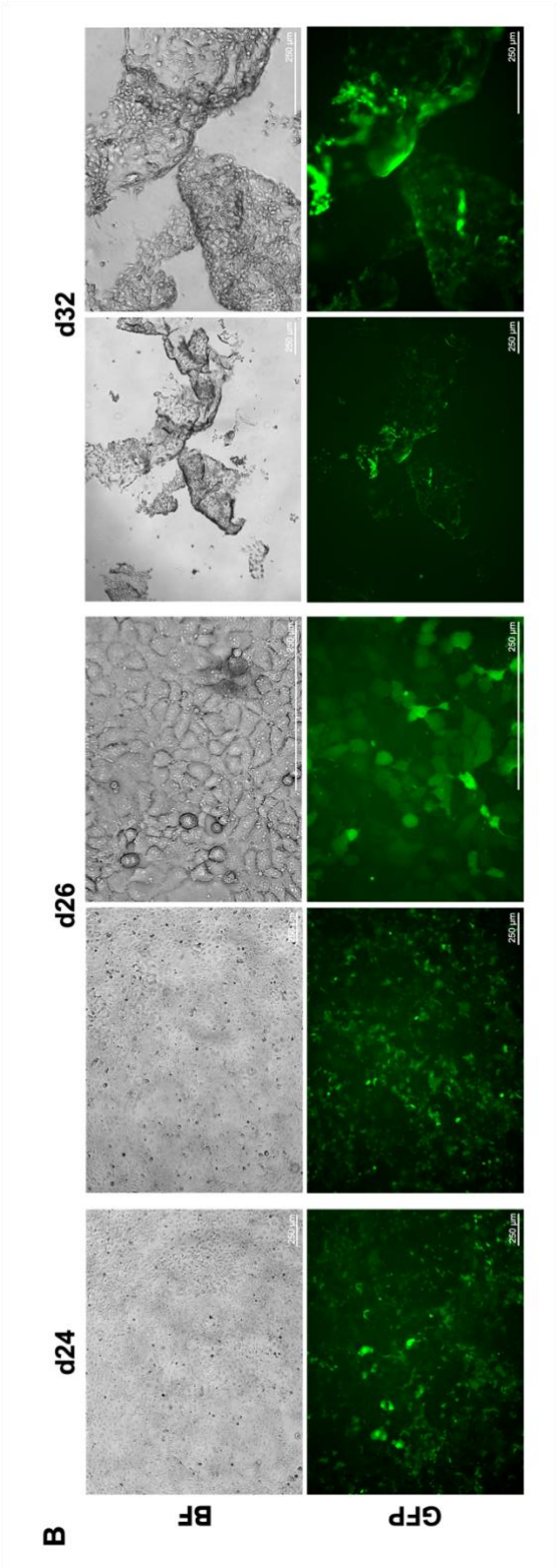
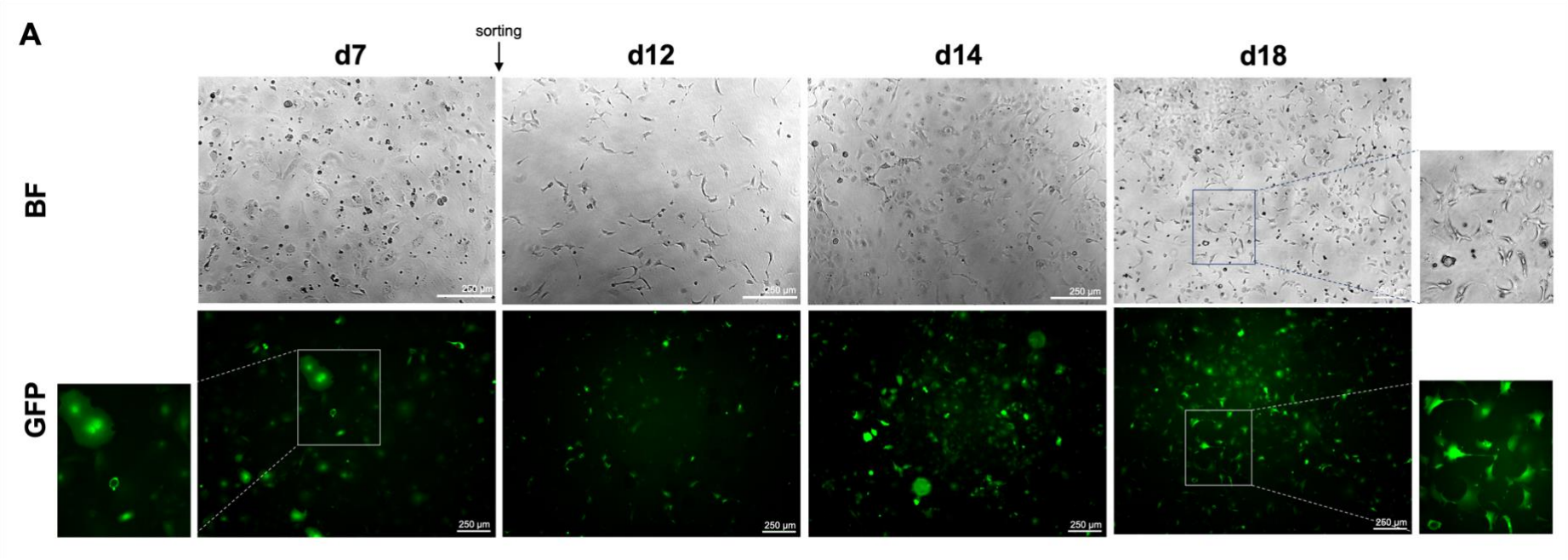


Figure 17: Transduction of HPCs with scramble (1/00).

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with scramble at different time points after transduction. (**A**) Transduced HPCs before FACS-sorting GFP+ cells. Scale bars, 250 µm. Top right: digital magnification. (**B**) Transduced HPCs after FACS-sorting GFP+ cells. Scale bars, 250 µm.

Figure 18: Transduction of HPCs with *Serpina1c* (guide 3) (1/100)



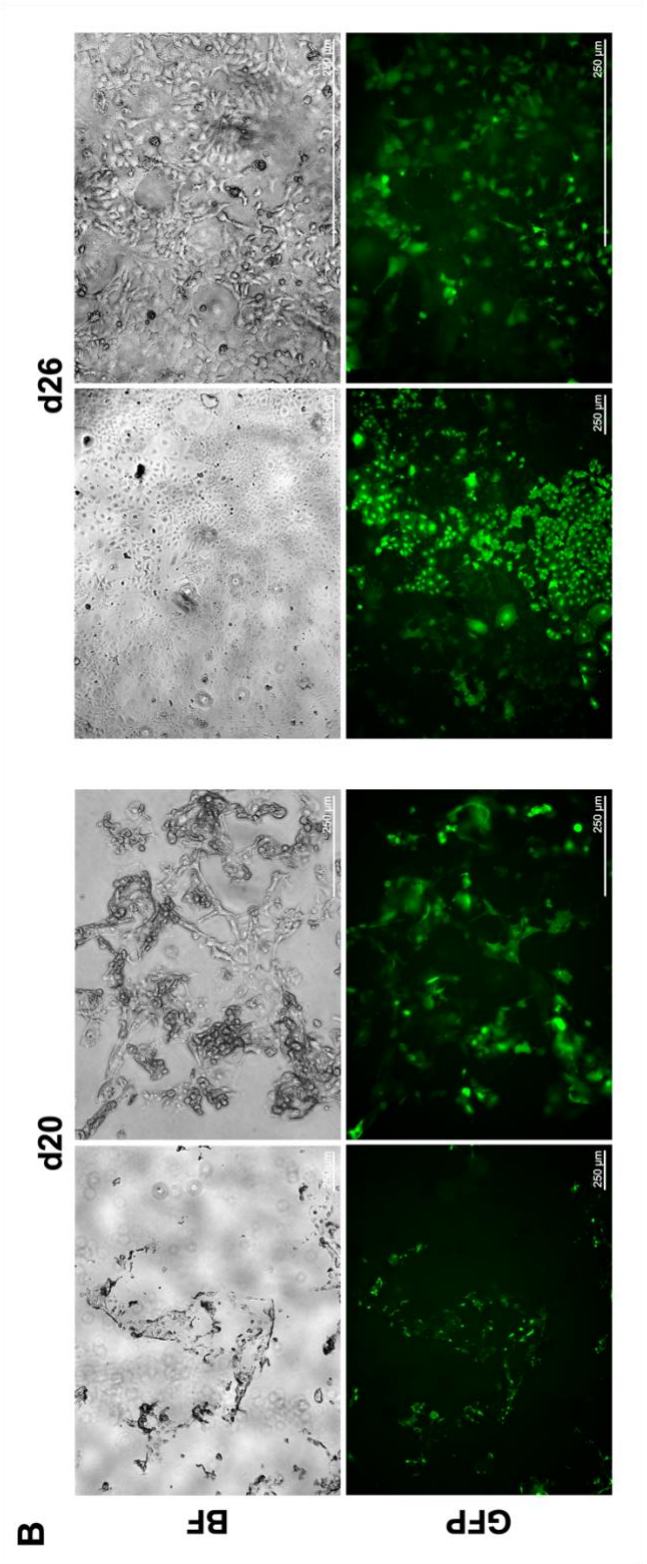


Figure 18: Transduction of HPCs with *Serpina1c* (guide 3) (1/100).

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with *Serpina1c* (guide 3) at different time points after transduction. (**A**) day 7 (d7) – day 18. Far left and right: digital magnification of area of interest. Scale bars, 250 µm. (**B**) Day 20 (d20) and day 26 (d26). Right: digital magnification. Scale bars, 250 µm.

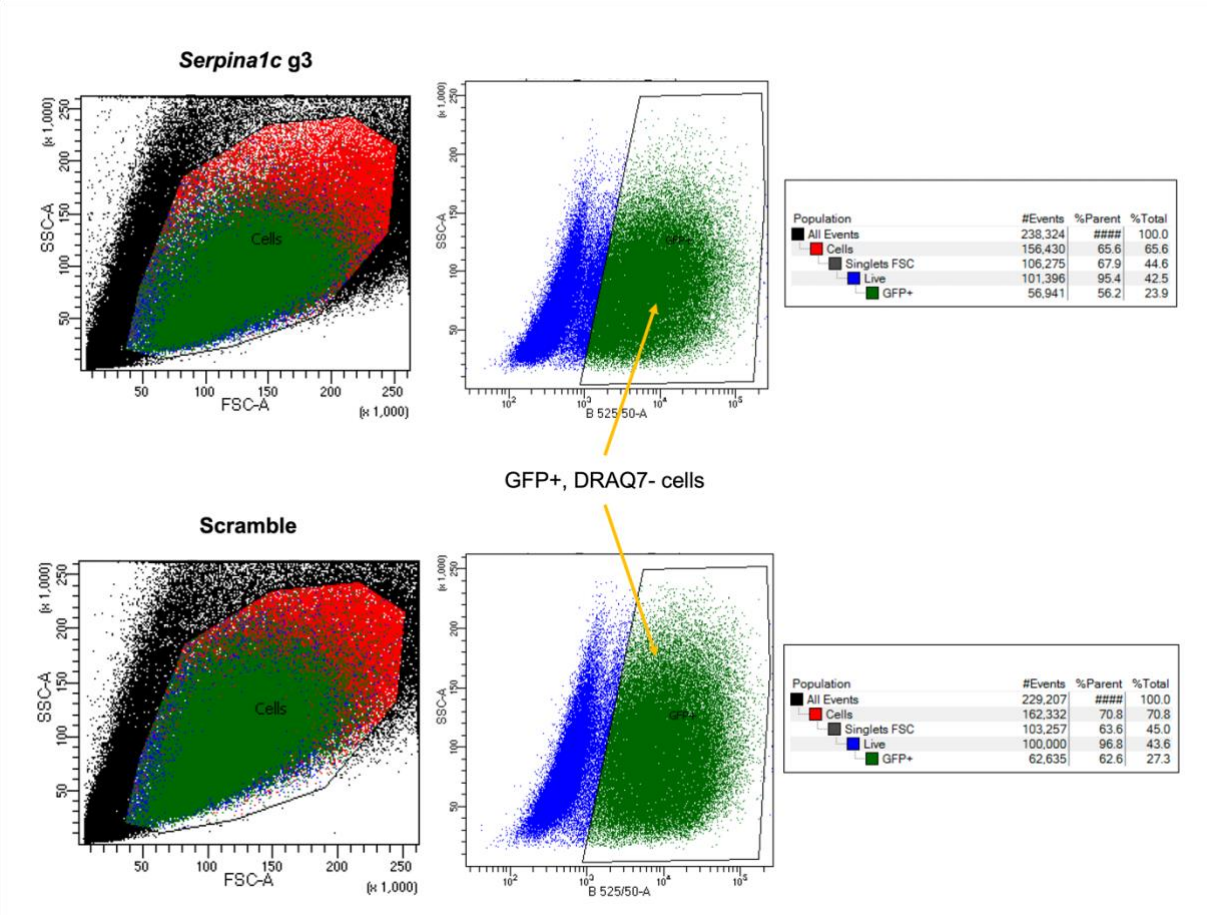


Figure 19: FACS sorting of *Serpina1c* KO and scramble cells.

FACS-sorting for GFP+ (green), DRAQ7- (blue) HPCs resulted in 56.2% and 62.6% *Serpina1c* (guide 3) transgenic and scramble HPCs (yellow arrows) respectively.

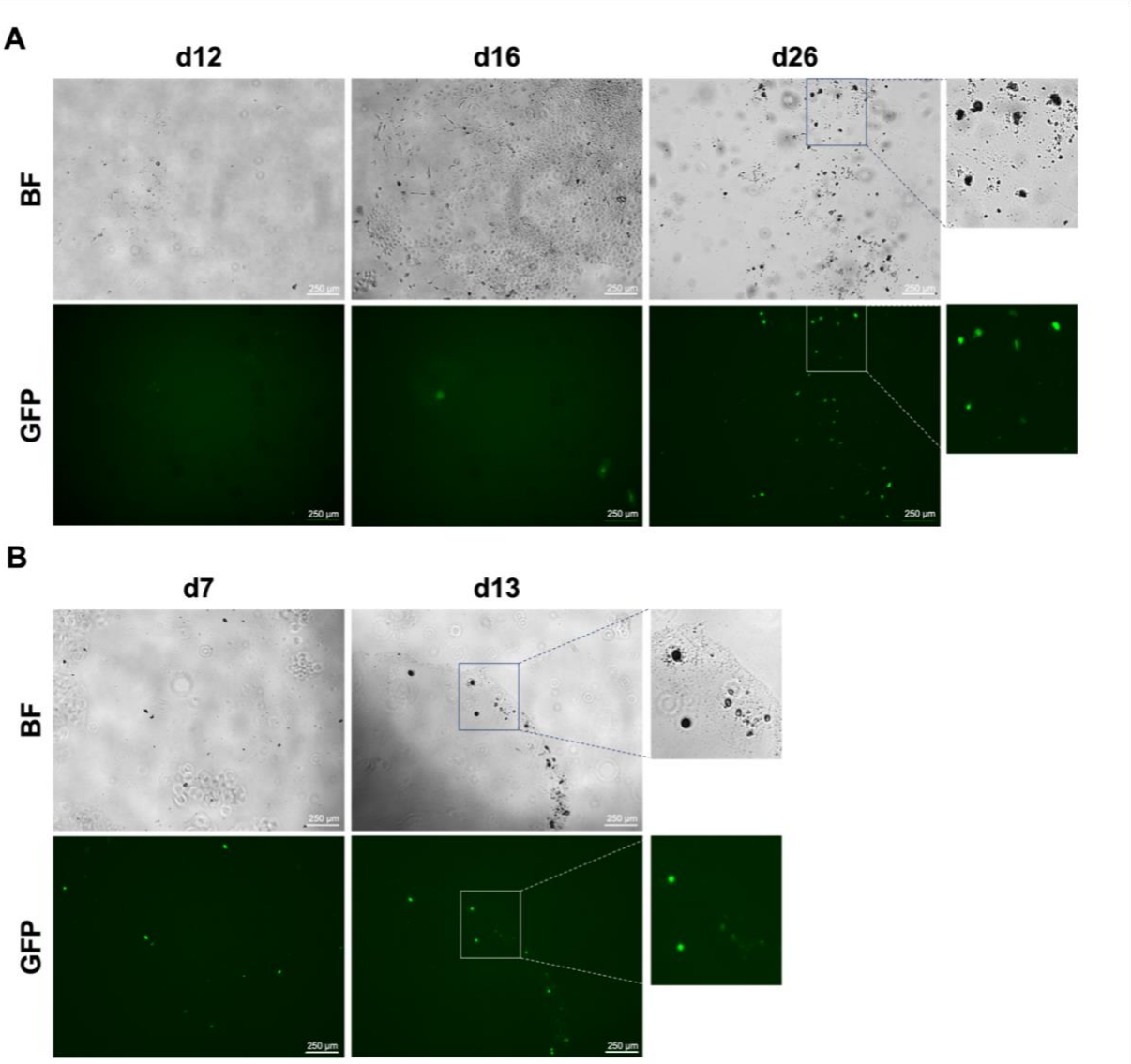


Figure 20: Transduction of HPCs with *Serpina1c* (guide 1) (1/50).

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with *Serpina1c* (guide 1) at different time points after transduction. Transduction was carried out twice after passage 19 (**A**) and passage 20 (**B**). Right: digital magnification. Scale bars, 250 µm.

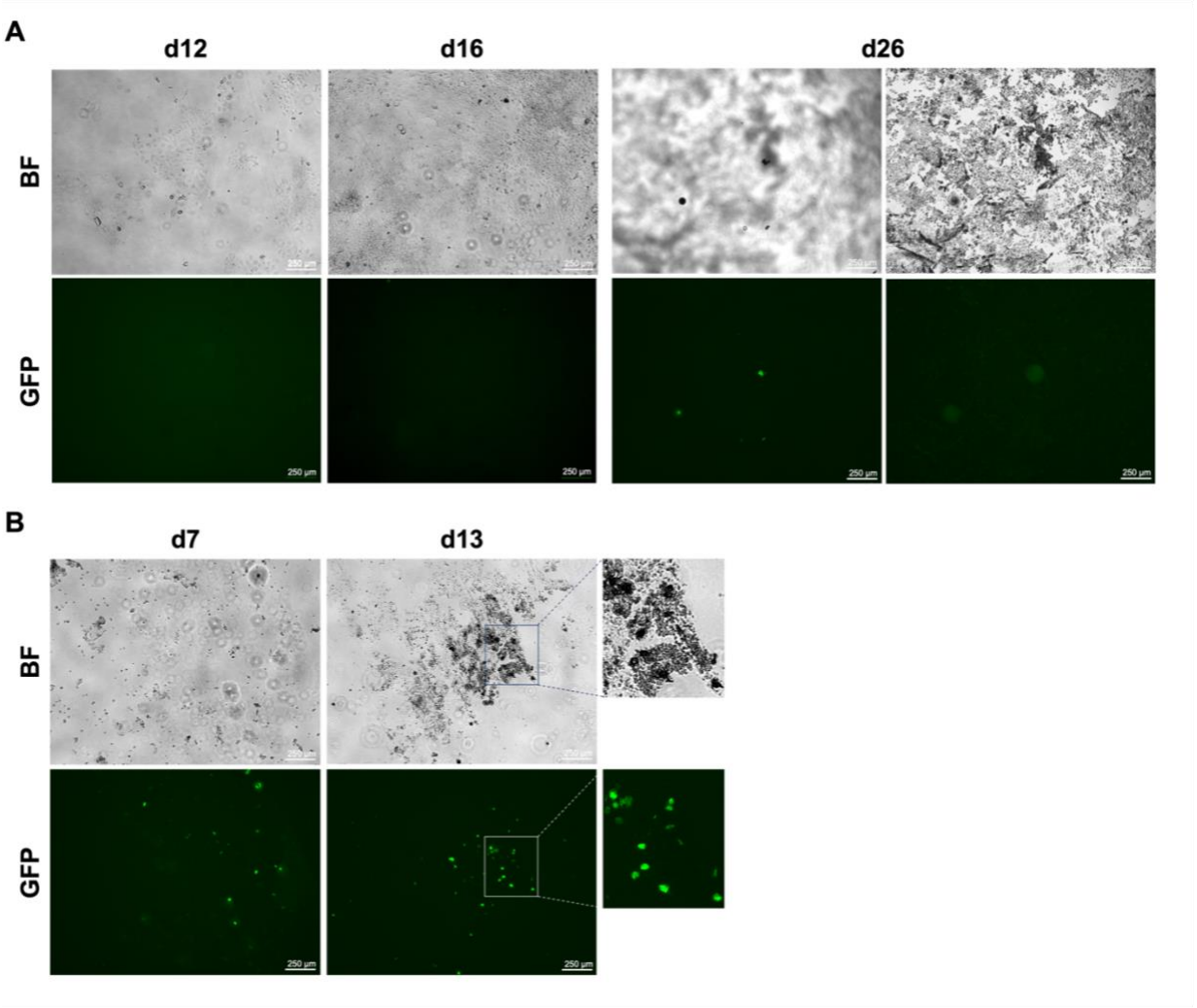


Figure 21: Transduction of HPCs with *Serpina1c* (guide 2) (1/50).

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with *Serpina1c* (guide 2) at different time points after transduction. Transduction was carried out twice after passage 19 (**A**) and passage 20 (**B**). Bottom right: digital magnification. Scale bars, 250 µm.

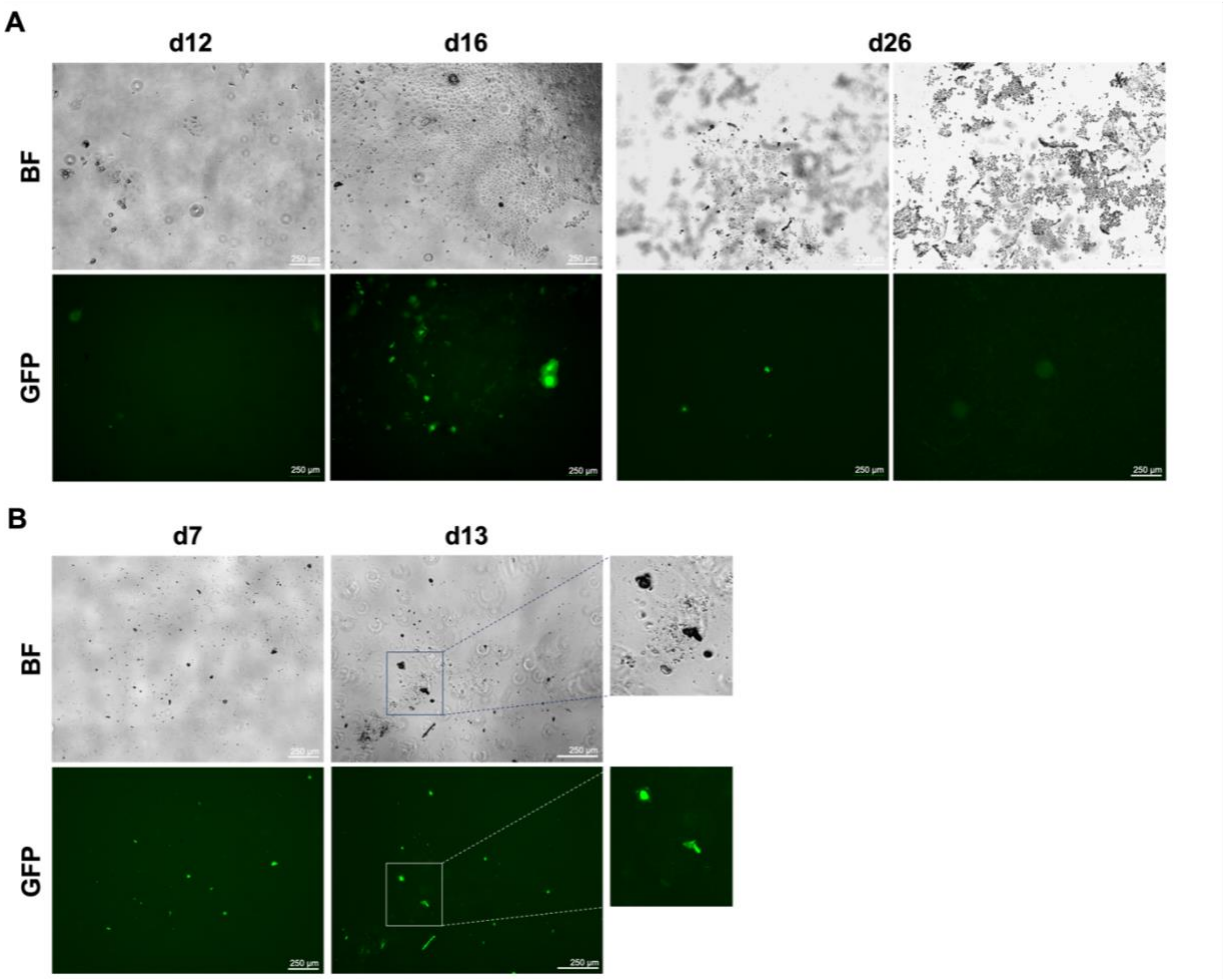


Figure 22: Transduction of HPCs with *Gc* (guide 3) (1/250)

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with *Gc* (guide 3) at different time points after transduction. Transduction was carried out twice after passage 19 (**A**) and passage 20 (**B**). Bottom right: digital magnification. Scale bars, 250 µm.

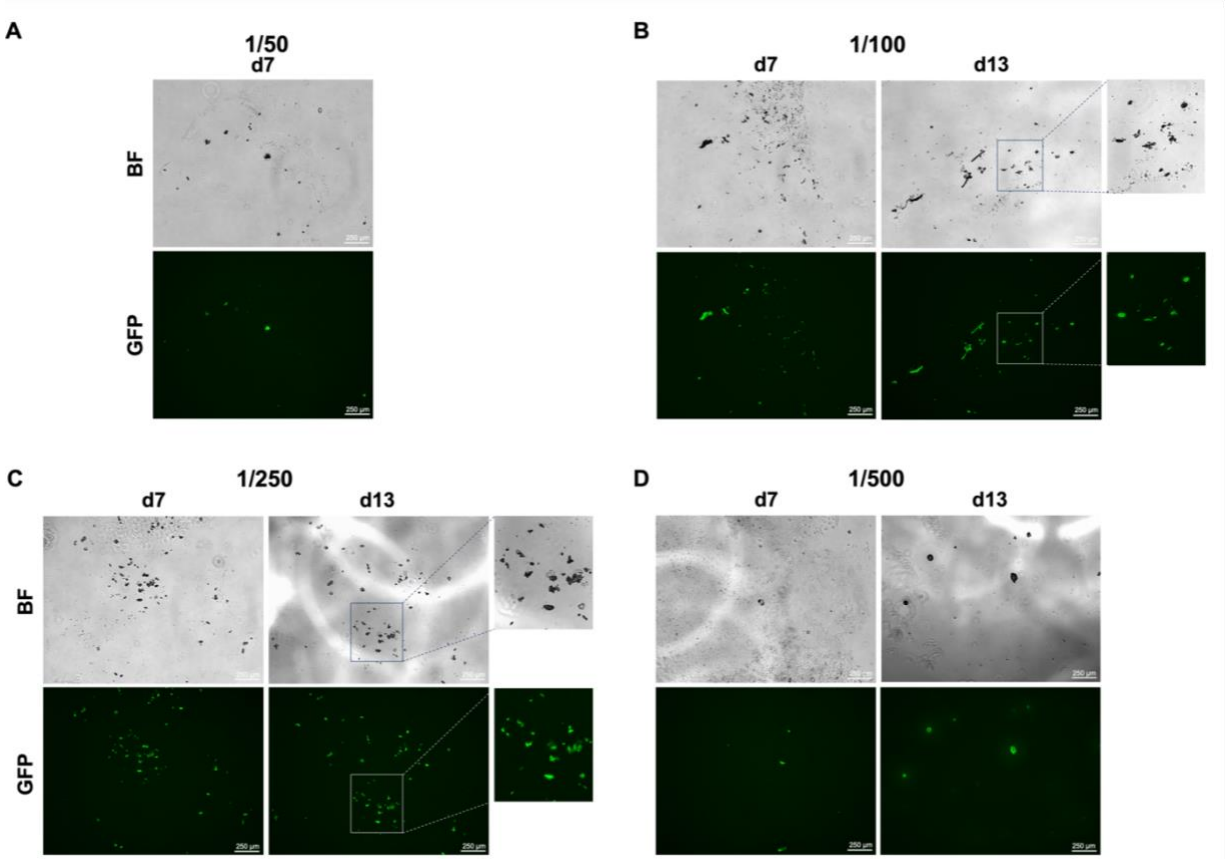


Figure 23: Transduction of HPCs with *Ccl2* (guide 3)

Brightfield (BF) and Green fluorescent protein (GFP+) images showing HPCs transduced with 4 different *Ccl2* (guide 3) dilutions at different time points after transduction. (**A**) Transduction performed with 1/50 dilution. Scale bars, 250 µm. (**B**) Transduction performed with 1/100 dilution. Right: digital magnification. Scale bars, 250 µm (**C**) Transduction performed with 1/250 dilution. Right: digital magnification. Scale bars, 250 µm. (**D**) Transduction performed with 1/500 dilution. Scale bars, 250 µm.

1.1 Zeocin selection

As the LentiCRISPRv2GFP backbone contains Bleomycin resistance, Zeocin selection was used as an alternative method to select successfully transduced HPCs. First, we determined that 600 µg/mL Zeocin was effective to kill approximately 80% of the HPCs (data not shown). Then, all 1/500 transgenic HPCs that showed a low percentage of GFP+ HPCs were used to study Zeocin selection. 600µg/mL Zeocin was added to the HPCs on day 10 after transduction. Fresh media without Zeocin was added to one well per gRNA as negative controls. 4 days later, the confluency and percentage of GFP+ HPCs were analyzed using a microscope.

Zeocin-treated HPCs showed an increase in the number of dead HPCs that impacted confluency in comparison with control wells (observe *Gc* (guide 3) transgenic HPCs in **Figure 24**). However, no increase of GFP+ HPCs was observed (data not shown).

A possible explanation for this could be the protocol adjustments (Zeocin™ Selection Reagent User Guide (2012)) as Zeocin was added only once instead of every 3-4 days and at a different time point (10 days instead of 72 hours) than instructed.
Future experiments that alter the timing for Zeocin selection (e.g. one addition at 72 hours) could potentially yield more GFP+ HPCs.

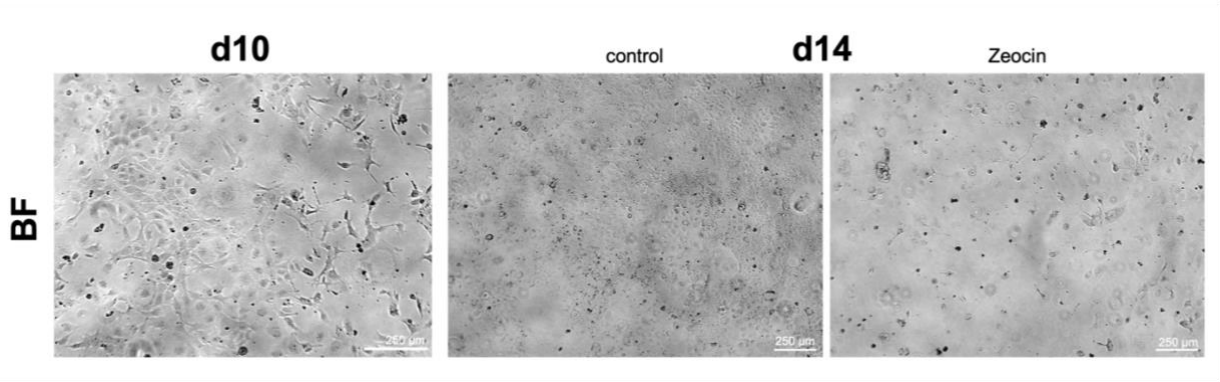


Figure 24: Zeocin selection of *Gc* (guide 3) transgenic HPCs
Brightfield (BF) images showing HPCs transduced with *Gc* (guide 3) before (d10) and after (d14) treatment with Zeocin. Scale bars, 250 µm.

2. Reverse Transcription and Real Time PCR

After FACS-sorting of scramble and *Serpina1c* (guide 3) KO GFP+ HPCs, the transcriptional differences between these cell lines and non-transduced control HPCs were assessed. Non-transduced control HPCs (from now on named 'control HPCs') were included to study the effect of the transduction itself on the expression of the assessed genes in HPCs. We were most interested in transcriptional differences of the 3 genes of interest, hepatocyte, cholangiocyte, HPC, proliferation, stemness, senescence, and inflammatory marker genes. However, due to a limited amount of time we only assessed differences in *Serpina1c*, *Gc*, and *Ccl2*; the hepatocyte markers *Alb, Hnf4a*, and *Cyp2e1*; the cholangiocyte markers *Epcam*, *Sox9, Krt19*, and *Hnf1β*; the proliferation markers *Mki67* and *Pcna*; and the stemness markers *Sox9, Onecut1, Prom1*, and *Lgr5*. Of these, *Sox9* and *Epcam* are also of interest as they are known as potential HPC markers (Font-Burgada et al., 2015; Furuyama et al., 2011; Han et al., 2019; Yoon et al., 2011; Yovchev et al., 2008).
Difference in expression of *Serpina1c* was assessed to validate the KO and expression of *Gc* and *Ccl2 was assessed* to determine if the ablation of *Serpina1c* affected the expression of the other 2 genes of interest. Expression of the other genes mentioned above was assessed to study the effect of the *Serpina1c* KO on the differentiation status of HPCs.

To perform RT-qPCR on the above genes, I purchased and tested the *Serpina1c, Gc*, and *Ccl2* primers. Primers for *Alb, Hnf4a*, *Cyp2e1, Epcam*, *Sox9, Krt19*, *Hnf1β, Mki67*, *Pcna, Sox9, Onecut1, Prom1*, and *Lgr5* were already available and had been tested by members of the Forbes lab (data not shown).

The available primers for *Serpina1c* detect both *Serpina1c* and *Serpina1a (*serine (or cysteine) peptidase inhibitor, clade A, member 1A). Both genes encode for α-1-antitrypsin which is represented by 6 individual *Serpina1* genes: *Serpina1a-f* (UniProt, Heit et al., 2013). However, human α-1-antitrypsin is only represented by one gene: SERPINA1 (Heit et al., 2013).
As there was no better primer option and *Serpina1c* was the only α-1-antitrypsin gene that was highly expressed in HPCs within the scRNA-seq dataset, I decided to use the best available primer (Qiagen, #QT01046038) despite the overlapping detection of *Serpina1a*. The *Serpina1c* primer was tested on hepatocytes (both healthy and injured with ethanol, provided by Dr Daniel Rodrigo Torres), the *Gc* primer on injured bulk liver tissue (obtained from Dr Marilena Candela), and the *Ccl2* primer on injured bulk liver tissue and HPCs (both healthy and injured with paracetamol, obtained from Dr Marilena Candela and Mr. MacMillan). Tests were marked successful when melt curve analysis displayed a single peak and when Ct values were between 15-20. All tests were successful except for the *Ccl2* primer in injured bulk liver tissue. As we only needed to use this primer in HPCs, this was not an issue for this project.

4 biological replicates of the scramble and *Serpina1c* (guide 3) KO cell lines and 3 biological replicates of control HPCs were used for RT-qPCR. I hypothesized that control and scramble would yield no significant differences as they both represent wild type HPCs. However, significant differences were found between these two groups (Observe **Figure 25, panel A, C,** and **K**) suggesting further alterations of the scramble cell line.

Serpina1c was significantly downregulated in both the transgenic *Serpina1c* KO and scramble HPCs compared to control HPCs and there was no difference in expression between *Serpina1c* KO and scramble HPCs (**Figure 25A**). As described above, the *Serpina1c* primer detects both *Serpina1a* and *-1c* which limits the interpretation of these results. In addition, the used primer for *Serpina1c* amplifies exons 1-2 whereas the *Serpina1c* (guide 3) sequence works downstream of exon 2 (**Figure 26**). It is therefore likely that exons 1-2 were detected but were later degraded or made a protein that is later degraded due to the mutation. To find out if *Serpina1c* is knocked out in the *Serpina1c* KO cell line and not in the scramble cell line, it is needed to develop a *Serpina1c* primer that targets the area of the genome containing the guide sequence and repeat RT-qPCR, sequence both genomes, and study the abundance and functionality of α-1-antitrypsin 1-3 in both cell lines.
In addition, *Ccl2* was significantly upregulated in the scramble HPCs compared to both control and the *Serpina1c* KO HPCs (**Figure 25B**). This also suggests an unexpected difference between the scramble and control HPCs. Assuming that the *Serpina1c* KO was successful, this difference could also be explained by an inflammatory response of the HPCs to the viral transduction, resulting in upregulation of *Ccl2* in the scramble HPCs that was impaired in the *Serpina1c* KO HPCs. However, this experiment should be repeated, and KO of *Serpina1c* requires further experimentation to confirm this hypothesis.

The following suggestions are made not taking into consideration the scramble results, and comparing the *Serpina1c* KO with the control and scramble HPCs:

1) There was no significant difference in the expression of the hepatocyte markers *Cyp2e1, Gc* (**Figure 25C**), and *Hnf4α. Cyp2e1* and *Hnf4α* were both non detectable in any of the groups (data not shown). *Albumin* displays a trend for increased expression in *Serpina1c* KO HPCs compared to control and scramble HPCs (**Figure 25D**).
2) The cholangiocyte markers *Hnf1β and Sox9,* and *Onecut1* (a regulator of biliary development (Clotman et al., 2002)) were all significantly upregulated in *Serpina1c* KO HPCs compared to scramble HPCs (**Figure 25G-I**). This might suggest that downregulation of *Serpina1c* results in differentiation of the HPCs towards a more cholangiocyte phenotype. *Epcam*, *Krt19,* and *Prom1* do not display significant differences in gene expression between these groups, but both *Epcam* and *Prom1* show a trend for increased expression in *Serpina1c* KO HPCs compared to control and scramble HPCs supporting the above findings (**Figure 25, panels E, F, and J**).
3) The stem cell marker *Lgr5* showed a significantly decreased expression in *Serpina1c* KO HPCs compared to scramble HPCs (**Figure 25K**). *Lgr5* (a Wnt pathway gene) has been identified as a HPC marker in biliary derived HPCs that were able to differentiate towards hepatocytes (Huch et al., 2013). This finding might suggest that HPCs lose stem cell characteristics due to *Serpina1c* ablation, which corresponds with the finding that *Serpina1c* KO HPCs express more cholangiocyte markers in comparison with scramble HPCs.
4) Finally, *Pcna* was significantly downregulated in *Serpina1c* KO HPCs in comparison with control and scramble HPCs (**Figure 25L**). This might suggest that *Serpina1c* plays a role in HPCs' proliferation. However, *Mki67* does not show a significant difference in expression (**Figure 25M)**. This could be explained by the functional differences between both marker genes. Expression of *Ki67* is increased during the late G1, S, G2, and M phase of the cell cycle (Bologna-Molina et al., 2013; Gerdes et al., 1983), whereas *Pcna* is only elevated during the G1/S-phase (Bologna-Molina et al., 2013; Kelman, 1997). It could be that the difference in proliferation between *Serpina1c* KO and scramble HPCs is mainly due to a decreased number of HPCs in G1/S-phase but not when combining G1, S, G2, and M phases together.

Altogether, these results might suggest that *Serpina1c* interacts with *Ccl2* expression*;* affects stemness and proliferative capacities; and plays a role in the differentiation status of HPCs in which decreased *Serpina1c* expression promotes differentiation towards a cholangiocyte phenotype.
It is however unclear whether the KO of *Serpina1c* has succeeded in the *Serpina1c* (guide 3) transgenic HPCs (and not in the scramble HPCs) based on these results alone.

To validate these results, RT-qPCR should be repeated using a *Serpina1c* primer that amplifies the region targeted by *Serpina1c* (guide 3), and genome sequencing should be performed for all conditions.

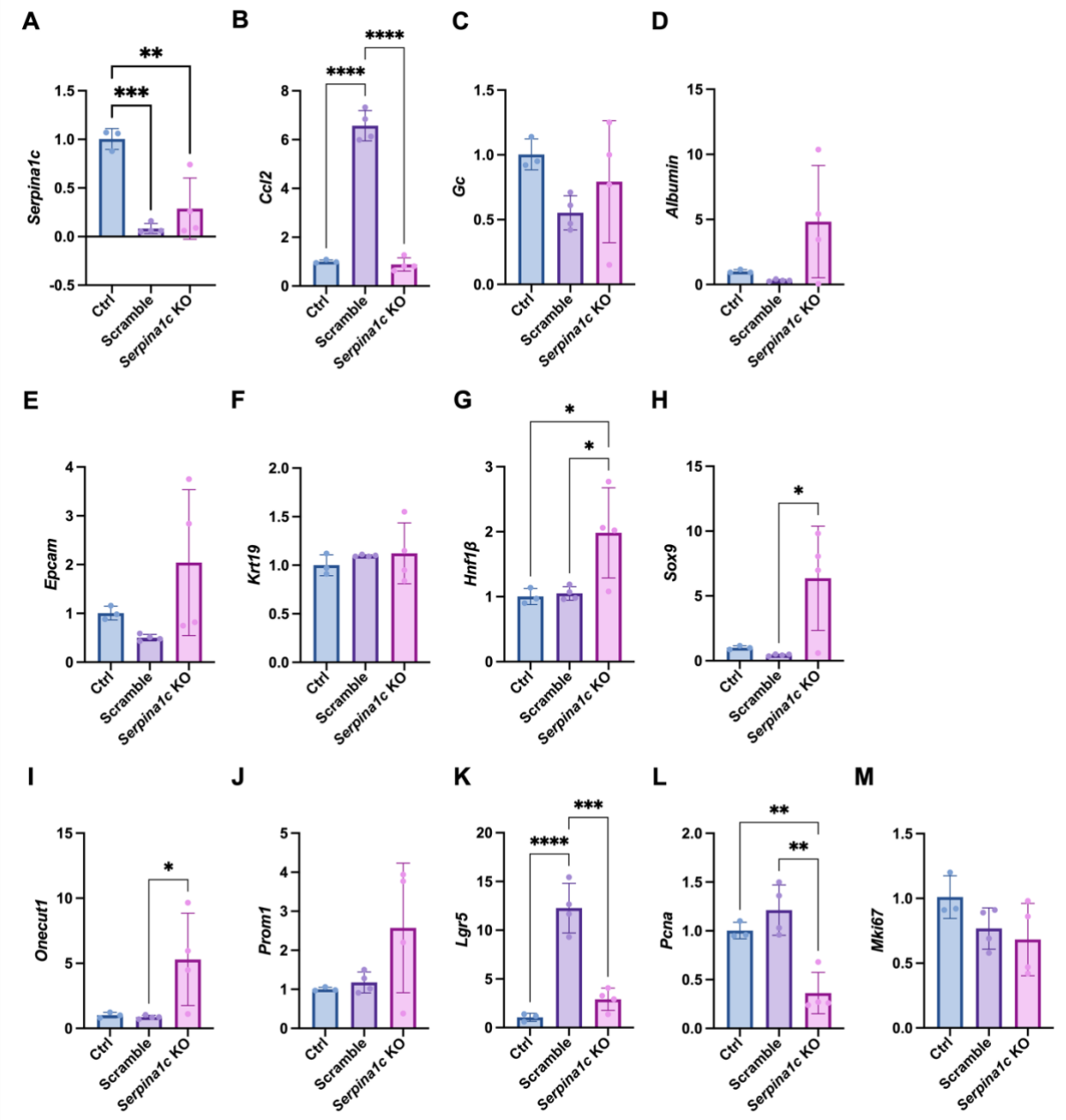


Figure 25: transcriptional differences between *Serpina1c* KO, scramble, and control HPCs.

Comparison of mRNA abundance of *Serpina1c* (**A**), *Gc* (**B**), *Ccl2* (**C**), *Albumin* (**D**), *Epcam* (**E**), *Krt19* (**F**), *Hnf1β* (**G**), *Sox9* (**H**), *Onecut1* (**I**), *Prom1* (**J**), *Lgr5* (**K**), *Pcna* (**L**), *and Mki67* (**M**) in control (blue), scramble (purple), and *Serpina1c* KO (pink) HPCs. Data is normalized to *Ppia* and *Gapdh*. * indicates p<0.05, **p<0.01, ***p<0.001, and ****p<0.0001 (Mean ± SEM). one-way ANOVA test. n = 4 biological replicates for scramble and *Serpina1c* KO HPCs and n = 3 biological replicates for control HPCs.

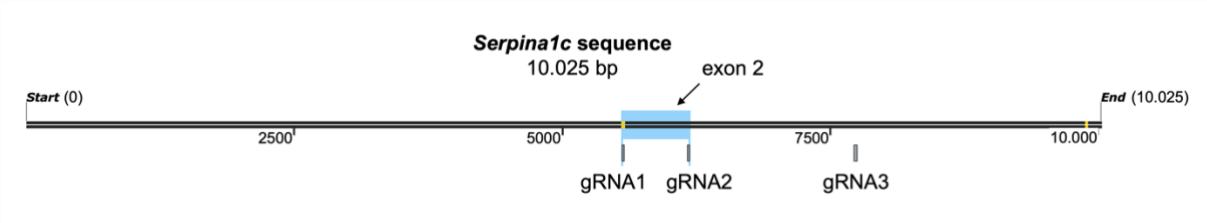


Figure 26: *Serpina1c* primer amplification.
Serpina1c gene sequence consisting of 10.025 base pairs (bps) showing the target sites of *Serpina1c* (guide 1-3): gRNA1-3. Blue region: exon 2. Adapted from SnapGene.

3. Immunohistochemistry

Immunohistochemistry is a suitable tool to study the effect of knocking out the genes of interest at a protein level. I first performed immunofluorescence stainings for the cholangiocyte markers K19 and EpCAM (**Figure 27A**), and the proliferation markers Ki67 and PCNA (**Figure 27B**) in murine three-dimensional bile duct-derived organoids. As expected, a widespread expression of K19 and EpCAM was visible, and part of the cells also expressed PCNA or Ki67 showing their proliferative capacity.

To stain for the genes of interest, antibodies for α-1-antitrpysin (*Serpina1c*), GC, and Monocyte Chemoattractant Protein 1 (MCP1, *Ccl2*) had to be tested before use. Antibodies for the other genes (*Alb, Hnf4a*, *Cyp2e1, Epcam*, *Sox9, Krt19*, *Hnf1β, Mki67*, *Pcna, Sox9, Onecut1, Prom1*, and *Lgr5,* described in Chapter 4.2) were already available and tested by members of the Forbes lab (data not shown). To test the antibodies, stainings were performed in tissues or cells known to express the markers of interest.

1) α-1-antitrypsin 1-3 tested positive in healthy liver tissue (provided by Dr Hannah Esser and Dr Niya Aleksieva) (**Figure 28**). Similar to the *Serpina1c* primer, the only available antibody for α-1-antitrpyin 1-3 (*Serpina1c*) have been shown to exhibit 70% cross-reactivity with recombinant mouse *Serpina1a* (*Mouse Serpin A1c/ Alpha 1-Antitrypsin Antibody AF2979: R&D Systems*, n.d.). Therefore, interpretation of staining results using this antibody is limited. To use this antibody for α-1-antitrpyin 1-3 stainings in HPCs, I would suggest performing a positive control staining in hepatocytes to develop the right staining conditions that can be used for HPCs as well.
2) GC immunostainings were also performed in healthy liver tissue (provided by Dr Hannah Esser and Dr Niya Aleksieva) but resulted in a low signal (data not shown). Due to limited amount of time, staining conditions could not be further improved. However, in the future, I would suggest performing a positive control staining in hepatocytes for the same reasons as for α-1-antitrypsin 1-3.
3) MCP1 immunostaining was performed using both immunofluorescence and immunohistochemistry. MCP1 immunohistochemistry was performed in liver and kidney tissue from thioacetamide-injured mice or AAV8.TBG.p21 mice treated

with MCD diet (provided by Dr Ben Dwyer and Dr Niya Aleksieva). Two antibodies were tested (Abcam, #ab8101 and #ab25124). Unfortunately, none of these stainings resulted in a positive signal. Additionally, the #ab8101 antibody was tested in control HPCs and bile duct organoids using immunofluorescence, but without any positive results. Potential explanations include expiration dates of the antibodies (that were sorted for more than five years). Using a fresh antibody in the future might improve these results.

Unfortunately, there was no time left within this project to improve the positive control stainings for GC and MCP1 and to perform stainings in the *Serpina1c* KO, scramble, and control HPCs.

In summary, the α-1-antitrpysin antibody was working in healthy liver tissue and can be tested in hepatocytes in the future. GC and MCP1 stainings were unsuccessful, requiring further improvements for future use in transgenic and control HPCs.

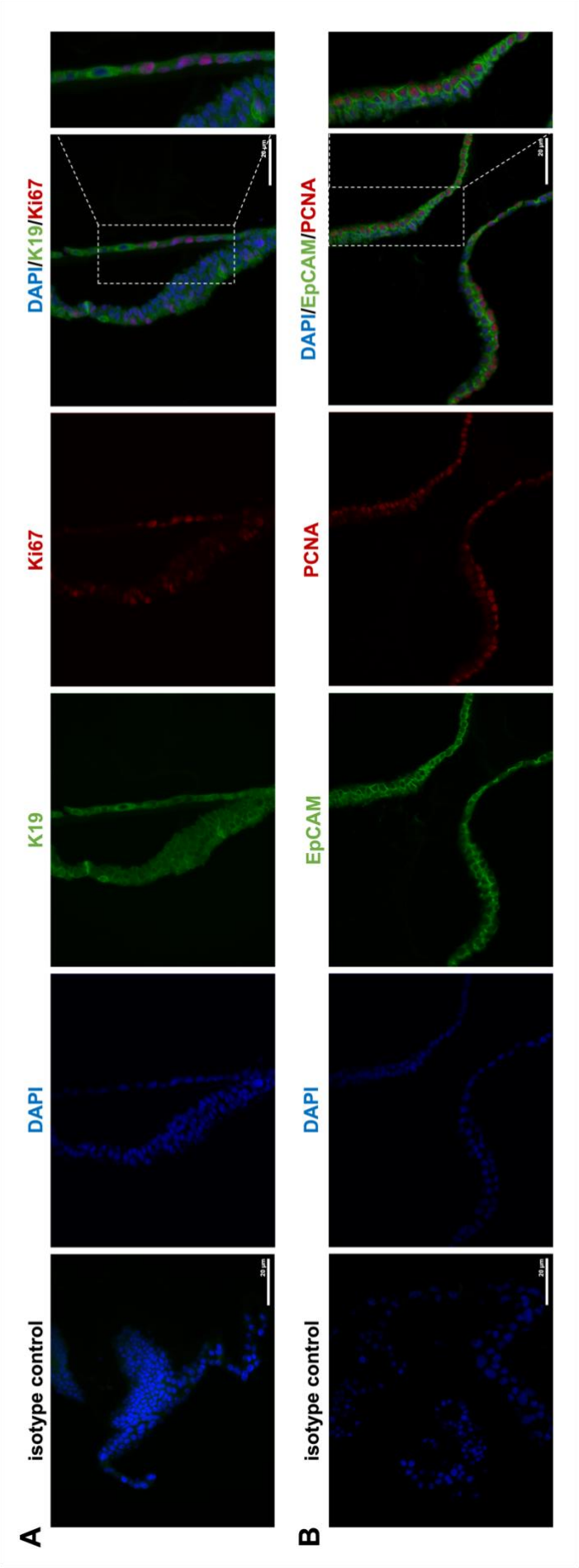


Figure 27: Characterization of bile duct organoids

Figure 27: Characterization of bile duct organoids

(**A**) DAPI (blue), K19 (green), and Ki67 (red) positive bile duct organoids. Left: isotype control. Right: digital magnification. Scale bars, 20 µm. (**B**) DAPI (blue), EpCAM (green), and PCNA (red) positive bile duct organoids. Left: isotype control. Right: digital magnification. Scale bars, 20 µm.

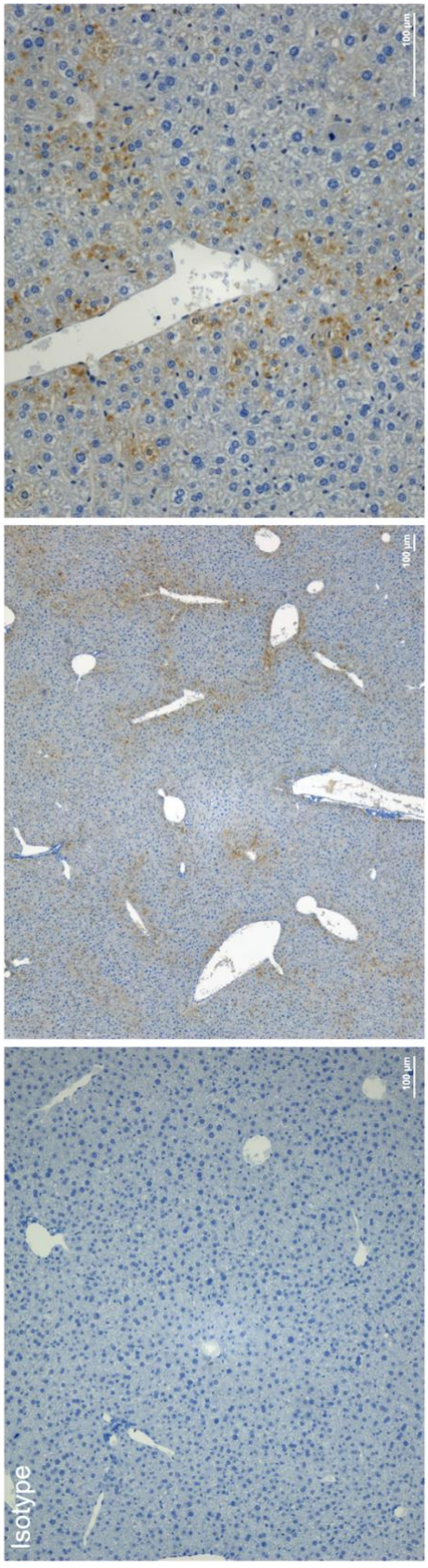


Figure 28: Positive control staining for α-1-antitrypsin 1-3

α-1-antitrypsin 1-3 (brown) positive healthy liver tissue. Left: isotype control. Scale bars, 100 µm.

4. Summary

To knock out the genes of interest, HPCs were transduced with lentivirus directed towards *Serpina1c* (guide 1-3), *Gc* (guide 3), and *Ccl2* (guide 3) or with a non-target scramble lentivirus using several viral dilutions.
Of these, transgenic *Serpina1c* (guide 3) and scramble HPCs showed visible GFP+ green cells and recovered and expanded well. These lines were FACS-sorted for GFP-positivity, with all cells displaying detaching at various time points after transduction for an unknown reason.

The FACS-sorted HPCs were subsequently used for RT-qPCR. Unfortunately, this did not result in a positive validation of *Serpina1c* KO, which will require further experimental settings (such as genome sequencing and studying the abundance and functionality of α-1-antitrypsin 1-3).

With the discussed aforementioned limitation of these results, qPCR data suggest that downregulation of *Serpina1c* affects the level of *Ccl2* expression, results in differentiation of HPCs towards a more cholangiocyte phenotype, and plays a role in HPCs' stemness and proliferative capacities. However, this experiment should be repeated to confirm these results.
Positive control stainings for α-1-antitrypsin 1-3, GC, and MCP1 still have to be improved to be able to use them for comparing the presence and abundance of the proteins between the transgenic and control HPCs.

Chapter 5 - Discussion and future perspectives

Chronic liver disease (CLD) is an extremely common clinical condition involving progressive deterioration of liver functions leading to progressive fibrosis, cirrhosis and, if untreated, death. Due to very limited treatment options for CLD, it is necessary to develop alternative therapeutic options.
CLD is characterized by hepatocellular senescence that compromises hepatocyte-mediated liver regeneration (Lu et al., 2015; Marshall et al., 2005; Raven et al., 2017; Richardson et al., 2007). Still, the extraordinary regenerative properties of the liver present a secondary tier of defense as a last resource: the hepatic progenitor cell.
HPCs of biliary origin can act as facultative liver stem cells, being able to transform into hepatocytes or cholangiocytes, to repair the liver parenchyma (Aleksieva, 2021; Deng et al., 2018; Manco et al., 2019; Raven et al., 2017; Russell et al., 2019). Therefore, HPCs are a potential source for liver cell therapy as an alternative therapeutic strategy to treat CLD.

Dr Aleksieva studied the conversion of biliary epithelial cells (BECs) into hepatocytes using murine models of CLD, comparing transcriptional differences between BECs isolated from mice in which hepatocytes can regenerate (control), and mice in which the regenerative capacity of hepatocytes was compromised (CLD, p21-driven model) (Aleksieva, 2021). This study confirmed that BECs differentiate into hepatocytes after liver injury. Through bulk RNA-seq and scRNA-seq at different time points during liver injury and recovery, transcriptional differences between the groups were examined. The BECs in this study contained cell populations that only emerged in mice with compromised hepatocyte regeneration identifying heterogeneity within the biliary epithelium (Aleksieva, 2021).

In this project, I analyzed Dr Aleksieva's scRNA-seq dataset to identify the biological pathways and genes responsible for BEC-mediated liver regeneration and repair. By doing so, I was able to determine the transcriptional differences between the 3 animal groups (healthy, d9ctrl, and d9p21) at day 9 recovery after liver injury. This led to the identification of 17 genes of interest based on a pseudo-bulk analysis between the groups; a cluster analysis identifying differentially expressed genes (DEGs) for the clusters that only appeared in the d9p21 group; an enrichment analysis on the DEGs of each cluster; the level of expression of the DEGs in the clusters and groups; and the frequency of the DEGs being of interest when combining the different analyses. The three most promising genes that were selected for *in vitro* validation were *Serpina1c, Gc*, and *Ccl2*.
As this analysis was only performed for day 9 recovery, this does not inform about the transcriptional differences at any of the other recovery time points, nor about the long-term transcriptional differences between the groups. In the future, it could be interesting

to perform a similar analysis at the other recovery time points to study the pathways and genes that are important for BECs to differentiate towards a hepatocyte phenotype. As the d9p21 BECs already express hepatocyte marker genes, identifying DEGs for p21 cell clusters at earlier time points might inform about the genes responsible for inducing this differentiation process.

The performed analysis on the transcriptional differences between the groups and clusters was mainly focused on the upregulated genes (as I initially aimed at ablating the expression of those genes using CRISPR strategies). However, additional analysis could be performed by also assessing the downregulated genes. In addition, clusters of cells were characterized and selecting genes of interest was performed combining the above performed analyses. A limitation of this method is that heterogeneity within d9p21 BECs is disregarded for this particular study. As the enrichment analysis suggested, the DEGs in cluster 2 and 6 appear to indicate functional differences. It would therefore be interesting to analyze differences in gene expression between the d9p21 clusters (2, 6, 10, 11, and 12) and assess any potential associated functional differences. This would characterize the different functions of the biliary epithelium after liver injury in animals with compromised hepatocyte regeneration.

The selection of 3 genes (*Serpina1c*, *Gc, and Ccl2)* for *in vitro* validation was based on a few considerations:

(i) The top 3 genes of interest based on the performed bioinformatics analysis: *Serpina1c*, *Gc, and Ccl2.*

(ii) *Serpina1c* and *Gc* are both hepatocyte markers. Upregulation of these genes in d9p21 BECs might therefore represent a differentiation process towards a hepatocyte phenotype.

(iii) Upregulation of these hepatocyte markers suggests that their associated roles (serine protease inhibition and vitamin D transport and storage) are highly important hepatic functions that need to be restored during early liver repair. The known genetic disorder causing α-1-antitrypsin deficiency highlights the importance of a well-functioning α-1-antitrypsin in humans.

(iv) The high expression of *Serpina1c* and *Gc* in the clusters of interest together with their known immunomodulatory functions and associated GO terms might suggest interaction of these genes with the vasculature and immune system after liver injury. This could be contributing to a potential role in liver regeneration.

(v) Other genes of interest included *Hmox1, Ccl2, Cnn2*, and *Thbs1.* These genes exhibit functions that require a tight interaction between the vasculature and the immune system, suggesting a role for both systems in hepatic regeneration that might be regulated by HPCs. Studying the role of these genes in HPC-mediated hepatic regeneration *in vitro* is limited, as this system does not allow to determine potential interactions of HPCs with the surrounding *in vivo*

environment during homeostasis and disease. Despite this limitation, we chose to ablate *Ccl2* in HPCs to study the role of *Ccl2* in HPCs in homeostasis.

In the future, the interaction of the genes described in (v) with the vasculature and immune system could be studied by adding immune factors to HPCs *in vitro* to examine effects on differentiation towards a hepatocyte phenotype. Additionally, it would be informative to study the differences and potential correlation between vascular changes, immune responses, and liver regeneration between the 3 animal groups (healthy, d9ctrl, and d9p21) used by Dr Aleksieva on recovery day 9 and other recovery days. Subsequently, KO (or even knock in (KI)) of *Hmox1, Ccl2, Cnn2*, and *Thbs1* – individually or combined – in BECs *in vivo* would allow us to study in detail the role of these markers in the context of liver regeneration and repair. This would also enable us to study up to what extent the vascular and immune system are involved in this process.

Nonetheless, I decided to study the function of *Serpina1c, Gc*, and *Ccl2* in HPCs *in vitro* by performing a CRISPR-KO experiment. As an alternative, it would have been suitable to perform a KI experiment to study a positive correlation between upregulation of these genes and conversion of HPCs into hepatocytes, which is directly in line with my hypothesis.
The above genes were individually knocked out which is a suitable tool to study the function of a gene in cells of interest. However, a function of interest (e.g. affecting the differentiation towards a BEC or hepatocyte phenotype) might possibly only be affected if several genes are knocked out together. In that case, ablating one of the responsible genes might not result in a functional effect.
Of all selected gRNAs (3 for each gene) for *Serpina1c, Gc*, and *Ccl2*, I was able to produce viable lentivirus for *Serpina1c* (guide 1-3), *Gc* (guide 3), *Ccl2* (guide 3), and a non-target scramble gRNA. If this project had continued, I would have completed virus production for all selected gRNAs, performed transductions with all lentiviruses and examine which gRNA is the most effective to knock out each gene.
Comparing transgenic HPCs with HPCs that only received polybrene showed that the cell confluency of only polybrene cells on day 10 after transduction was lower than for the transgenic *Serpina1c* (guide 1-2) cells on day 7. A potential explanation is that the wells that only received polybrene contained less cells before the start of the experiment. Alternatively, more polybrene could have been inadvertently added, or the confluence of the images may not be representative for the complete well. It is however unlikely that polybrene was toxic to the cells because part of the cells still looked viable and most of the transgenic cells recovered well from transduction.

Lentivirus transduction of HPCs resulted in visible green GFP-positive (GFP+) cells – indicative of a successful gene mutation – for all gRNAs and scramble. The optimal virus dilutions to obtain GFP+ and viable expanding cells varied between the gRNAs were: 1/50 for *Serpina1c* (guide 1-2), 1/100 for *Serpina1c* (guide 3) and scramble, 1/250 for *Gc* (guide 3), and 1/100 and 1/250 for *Ccl2* (guide 3).

At various time points after transduction, a pattern of detaching cells was visible in all transgenic cell lines, scrambled control, and non-transduced control HPCs. Considering this happened at day 20+ after transduction (and to all cell lines), it is unlikely that this is caused by the transduction process.
A possible explanation could be that cold media was added to cells on coated plates resulting in cold shock and subsequent detachment of cells. This would also explain why the cells detach in monolayers or islands of cells. Another possible explanation could be a progressive degradation of the collagen coating, leading it to cell detachment. As no other lab member had detaching cells in culture during the time that I performed the experiment and collagen coating was performed similarly by other members of the group, this is unlikely to be the cause.
Alternatively, the gRNA might have off-target effects in the genome that are tolerable in the short term but affects attachment or viability of the cells in the long term. The gRNAs have been computationally designed to minimize off-target effects throughout the whole genome (Joung et al., 2017), but this might not always predict what happens in reality. This possibility would explain why both scrambled control and transgenic cells are impacted, but not why control HPCs are also affected. In the future, the genome of the cell lines could be sequenced to find out if any off-target effects happened.

Transductions were performed at 3 different moments using the same HPC cell line after different passage numbers. A clear decrease in viability and recovery of the cells post transduction is seen after the last transduction (*Serpina1c* (guide 1-2, P20), *Gc* (guide 3, P20), and *Ccl2* (guide 3). The viability of the used HPCs was also decreased before this transduction (lower expansion rate and difference in expansion pattern) compared to the other transductions, which might explain this difference. Performing transductions at different time points limited the comparability of the different transductions. In the future, this experiment could therefore be improved by performing all transductions at the same time.

After transduction, GFP+ transgenic *Serpina1c* (guide 3) and scramble cells (diluted in 1/100) were successfully selected using FACS, resulting in 56.2% and 62.6% GFP+ HPCs respectively. Subsequently, RT-qPCR was performed on both cell lines and control HPCs to validate and study the functional effects of the ablation of *Serpina1c*. *Serpina1c* was downregulated in both the *Serpina1c* KO and scramble HPCs compared to control HPCs but there was no difference in expression between *Serpina1c* KO HPCs and scramble HPCs. It is unclear why expression of *Serpina1c* was downregulated in scramble HPCs, but this might be due to a technical problem (e.g. *Serpina1c* (guide 3) lentivirus were accidentally added to the scramble cells) or a problem with the primer. The used primer for *Serpina1c* amplifies both *Serpina1a* and *Serpina1c* which could have affected these PCR results. In addition, *Serpina1c* (guide 3) targets the gene downstream of the amplified exons of the used primer. It is therefore likely that the detected mRNA was later degraded or made a protein that is later degraded due to the mutation. Due to limited amount of time within

this project, I was unable to design new primers that would be able to detect the *Serpina1c* KO caused by *Serpina1c* (guide 3).
Assuming that *Serpina1c* was knocked out, *Ccl2* expression was downregulated in the *Serpina1c* KO cells compared to scramble cells which might suggest that ablating *Serpina1c* inhibits the expression of *Ccl2*. Interestingly, *Ccl2* expression is upregulated in scramble HPCs compared to control HPCs. This could indicate a problem with the scramble HPCs but might also indicate that scramble HPCs are upregulating *Ccl2* expression as a response to lentivirus transduction which is impaired in the *Serpina1c* KO HPCs. To find out if the level of *Serpina1c* and *Ccl2* expression are linked to each other, RT-qPCR for *Serpina1c* in a *Ccl2* KO cell line and stainings for both genes in both transgenic cell lines would provide more insights.
A similar effect on gene expression as for *Ccl2* was visible for *Lgr5*, a stem cell marker that is known to facilitate regeneration in the liver. Together with decreased *Pcna* expression, this might suggest a role for *Serpina1c* in cell survival. Furthermore, the biliary genes *Hnf1β, Sox9,* and *Onecut1* were upregulated in the *Serpina1c* KO HPCs compared to scramble HPCs, whereas the hepatocyte marker genes are not detectable in all cell lines or are not affected. This might suggest that *Serpina1c* plays a role in the differentiation status of HPCs by inhibiting differentiation towards a cholangiocyte phenotype or by promoting a HPC phenotype.
To validate these PCR findings, repeating transductions and RT-qPCR would be required. New primers should be designed for *Serpina1c* in the area that is affected by *Serpina1c* (guide 3) to validate the KO with PCR. Genomes should be sequenced, and gene expression should also be assessed on a protein level.
Another potential future approach could be to KI *Serpina1c* to find out if this promotes differentiation towards a HPC or hepatocyte phenotype, and results in increased expression of *Ccl2*, *Lgr5,* and *Pcna*.

Finally, to study the effect of the KOs on a protein level, I performed test stainings for α-1-antitrypsin 1-3, GC, and MCP1. The α-1-antitrypsin 1-3 staining was successful, but GC and MCP1 stainings require further optimization. However, the used α-1-antitrypsin 1-3 antibody was known to exhibit cross-reactivity with recombinant mouse *Serpina1a*. Therefore, interpretation of staining results using this antibody is limited. In the future, this project would require optimization of the staining protocols in HPCs for these 3 antibodies in order to validate the KO cell lines.

Future directions
In an ideal scenario this project would benefit from in vitro validation of all 17 genes of interest found in the bioinformatics analysis, in order to explore which of these genes might be important for the conversion of HPCs into hepatocytes after liver injury. For this purpose, both KO or KI strategies could be performed in HPCs *in vitro,* in a similar fashion to the work performed in this project. Subsequently, RT-qPCR and immunohistochemistry could be performed to examine the gene editing effects on the expression of hepatocyte and cholangiocyte, HPC phenotype, DR, proliferation, stemness,

senescence, and inflammatory markers. For genes that have a similar function (e.g. heat shock proteins or inflammatory genes) it would be interesting to perform combined gene editing or RNA interference experiments to study synergistic effects on these markers. Moreover, inducing *in vitro* models of liver disease using control and transgenic HPCs (e.g. senescence with etoposide) could be performed. The damaged HPCs could be compared with healthy HPCs by assessing the transcriptional differences for the 17 genes of interest. Additionally, damaging both transgenic and control HPCs could inform about the transcriptional differences between these groups. This would enable to study the role of the edited genes in HPCs after injury.
Furthermore, an artificial cholangiocyte to hepatocyte differentiation protocol could be performed on transgenic, scramble, and control HPCs to assess differences within this process between these groups. This could provide insights in the role of the edited genes in this process.
Finally, I would suggest performing the above *in vitro* experiments in human HPCs and a drug screening could be performed to test whether repurposed drugs could favorably affect gene expression of the genes of interest and promote the conversion of HPCs into hepatocytes. This could be a potential step towards the translation of the above findings to the clinic.

In summary, I found 17 genes that might play a role in BEC-mediated hepatic regeneration and repair. To validate these findings, I performed a CRISPR-KO experiment in HPCs targeting *Serpina1c, Gc,* and *Ccl2*. I performed RT-qPCR on *Serpina1c* (guide 3), scramble control, and non-transduced control HPCs, with results suggesting that *Serpina1c* might inhibit differentiation of HPCs towards a cholangiocyte phenotype, enhance stemness and proliferative capacities of HPCs, and promote *Ccl2* expression.
To validate these findings, genome sequencing should be performed, RT-qPCR should be repeated, and protein expression should be assessed. Future studies are needed to identify more genes of interest, and to determine their roles in the conversion of HPCs into hepatocytes. By doing so, this project could determine whether these genes can be therapeutically modulated to enhance HPC-mediated hepatic regeneration and repair.

References

Aleksieva, N. (2021). *Impaired hepatocyte regeneration confers lineage plasticity in the biliary epithelium-a study of the transcriptional heterogeneity within the biliary epithelium following injury*.

Aravinthan, A., Scarpini, C., Tachtatzis, P., Verma, S., Penrhyn-Lowe, S., Harvey, R., Davies, S. E., Allison, M., Coleman, N., & Alexander, G. (2013). Hepatocyte senescence predicts progression in non-alcohol-related fatty liver disease. *Journal of Hepatology*, *58*(3), 549–556. https://doi.org/10.1016/J.JHEP.2012.10.031

Asrani, S. K., Devarbhavi, H., Eaton, J., & Kamath, P. S. (2019). Burden of liver diseases in the world. *Journal of Hepatology*, *70*(1), 151–171. https://doi.org/10.1016/J.JHEP.2018.09.014

Bastian, F. B., Roux, J., Niknejad, A., Comte, A., Fonseca Costa, S. S., de Farias, T. M., Moretti, S., Parmentier, G., de Laval, V. R., Rosikiewicz, M., Wollbrett, J., Echchiki, A., Escoriza, A., Gharib, W. H., Gonzales-Porta, M., Jarosz, Y., Laurenczy, B., Moret, P., Person, E., … Robinson-Rechavi, M. (2021). The Bgee suite: integrated curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Research*, *49*(D1), D831–D847. https://doi.org/10.1093/NAR/GKAA793

Bologna-Molina, R., Mosqueda-Taylor, A., Molina-Frechero, N., Mori-Estevez, A. D., & Sánchez-Acuña, G. (2013). Comparison of the value of PCNA and Ki-67 as markers of cell proliferation in ameloblastic tumor. *Medicina Oral, Patología Oral y Cirugía Bucal*, *18*(2), e174. https://doi.org/10.4317/MEDORAL.18573

Bouillon, R., & Pauwels, S. (2018). The Vitamin D-Binding Protein. *Vitamin D: Fourth Edition*, *1*, 97–115. https://doi.org/10.1016/B978-0-12-809965-0.00007-0

Boulter, L., Govaere, O., Bird, T. G., Radulescu, S., Ramachandran, P., Pellicoro, A., Ridgway, R. A., Seo, S. S., Spee, B., Van Rooijen, N., Sansom, O. J., Iredale, J. P., Lowell, S., Roskams, T., & Forbes, S. J. (2012). Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. *Nature Medicine 2012 18:4*, *18*(4), 572–579. https://doi.org/10.1038/NM.2667

Boulter, L., Govaere, O., Bird, T. G., Radulescu, S., Ramachandran, P., Pellicoro, A., Ridgway, R. A., Seo, S. S., Spee, B., van Rooijen, N., Sansom, O. J., Iredale, J. P., Lowell, S., Roskams, T., & Forbes, S. J. (2012). Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. *Nature Medicine 2012 18:4*, *18*(4), 572–579. https://doi.org/10.1038/nm.2667

Boulter, L., Lu, W. Y., & Forbes, S. J. (2013). Differentiation of progenitors in the liver: a matter of local choice. *The Journal of Clinical Investigation*, *123*(5), 1867–1873. https://doi.org/10.1172/JCI66026

Chun, R. F. (2012). New perspectives on the vitamin D binding protein. *Cell Biochemistry and Function*, *30*(6), 445–456. https://doi.org/10.1002/CBF.2835

Clotman, F., Lannoy, V. J., Reber, M., Cereghini, S., Cassiman, D., Jacquemin, P., Roskams, T., Rousseau, G. G., & Lemaigre, F. P. (2002). The onecut transcription factor HNF6 is required for normal development of the biliary tract. *Development*, *129*(8), 1819–1828. https://doi.org/10.1242/DEV.129.8.1819

Consortium, T. U., Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/NAR/GKAC1052

Deng, X., Zhang, X., Li, W., Feng, R. X., Li, L., Yi, G. R., Zhang, X. N., Yin, C., Yu, H. Y., Zhang, J. P., Lu, B., Hui, L., & Xie, W. F. (2018). Chronic Liver Injury Induces Conversion of Biliary Epithelial Cells into Hepatocytes. *Cell Stem Cell*, *23*(1), 114-122.e3. https://doi.org/10.1016/J.STEM.2018.05.022

Duncan, A. W., Dorrell, C., & Grompe, M. (2009). Stem cells and liver regeneration. *Gastroenterology*, *137*(2), 466–481. https://doi.org/10.1053/J.GASTRO.2009.05.044

Falkowski, O., An, H. J., Ianus, I. A., Chiriboga, L., Yee, H., West, A. B., & Theise, N. D. (2003). Regeneration of hepatocyte 'buds' in cirrhosis from intrabiliary stem cells. *Journal of Hepatology*, *39*(3), 357–364. https://doi.org/10.1016/S0168-8278(03)00309-X

Ferreira-Gonzalez, S., Man, T. Y., Esser, H., Aird, R., Kilpatrick, A. M., Rodrigo-Torres, D., Younger, N., Campana, L., Gadd, V. L., Dwyer, B., Aleksieva, N., Boulter, L., Macmillan, M. T., Wang, Y., Mylonas, K. J., Ferenbach, D. A., Kendall, T. J., Lu, W. Y., Acosta, J. C., … Forbes, S. J. (2022). Senolytic treatment preserves biliary regenerative capacity lost through cellular senescence during cold storage. *Science Translational Medicine*, *14*(674). https://doi.org/10.1126/SCITRANSLMED.ABJ4375/SUPPL_FILE/SCITRANSLMED.ABJ4375_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Finding Significant Genes -Software -Single Cell Gene Expression -Official 10x Genomics Support. (n.d.). Retrieved 8 April 2023, from https://support.10xgenomics.com/single-cell-gene-expression/software/visualization/latest/tutorial-features

Font-Burgada, J., Shalapour, S., Ramaswamy, S., Hsueh, B., Rossell, D., Umemura, A., Taniguchi, K., Nakagawa, H., Valasek, M. A., Ye, L., Kopp, J. L., Sander, M., Carter, H., Deisseroth, K., Verma, I. M., & Karin, M. (2015). Hybrid Periportal Hepatocytes Regenerate the Injured Liver without Giving Rise to Cancer. *Cell*, *162*(4), 766–779. https://doi.org/10.1016/J.CELL.2015.07.026

Forbes, S. J., Gupta, S., & Dhawan, A. (2015). Cell therapy for liver disease: From liver transplantation to cell factory. *Journal of Hepatology*, *62*(1 Suppl), S157–S169. https://doi.org/10.1016/J.JHEP.2015.02.040

Furuyama, K., Kawaguchi, Y., Akiyama, H., Horiguchi, M., Kodama, S., Kuhara, T., Hosokawa, S., Elbahrawy, A., Soeda, T., Koizumi, M., Masui, T., Kawaguchi, M., Takaori, K., Doi, R., Nishi, E., Kakinoki, R., Deng, J. M., Behringer, R. R., Nakamura, T., & Uemoto, S. (2011). Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nature Genetics*, *43*(1), 34–41. https://doi.org/10.1038/NG.722

Gadd, V. L., Aleksieva, N., & Forbes, S. J. (2020). Epithelial Plasticity during Liver Injury and Regeneration. *Cell Stem Cell*, *27*(4), 557–573. https://doi.org/10.1016/J.STEM.2020.08.016

Gerdes, J., Schwab, U., Lemke, H., & Stein, H. (1983). Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *International Journal of Cancer*, *31*(1), 13–20. https://doi.org/10.1002/IJC.2910310104

Gouw, A. S. H., Clouston, A. D., & Theise, N. D. (2011). *Ductular Reactions in Human Liver: Diversity at the Interface*. https://doi.org/10.1002/hep.24613

Grant, G. A. (2018). D-3-Phosphoglycerate Dehydrogenase. *Frontiers in Molecular Biosciences*, *5*(DEC). https://doi.org/10.3389/FMOLB.2018.00110

Haines, D. D., & Tosaki, A. (2020). Heme Degradation in Pathophysiology of and Countermeasures to Inflammation-Associated Disease. *International Journal of Molecular Sciences*, *21*(24), 1–25. https://doi.org/10.3390/IJMS21249698

Han, X., Wang, Y., Pu, W., Huang, X., Qiu, L., Li, Y., Yu, W., Zhao, H., Liu, X., He, L., Zhang, L., Ji, Y., Lu, J., Lui, K. O., & Zhou, B. (2019). Lineage Tracing Reveals the Bipotency of SOX9+ Hepatocytes during Liver Regeneration. *Stem Cell Reports*, *12*(3), 624–638. https://doi.org/10.1016/J.STEMCR.2019.01.010

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

He, J., Lu, H., Zou, Q., & Luo, L. (2014). Regeneration of Liver After Extreme Hepatocyte Loss Occurs Mainly via Biliary Transdifferentiation in Zebrafish. *Gastroenterology*, *146*(3), 789-800.e8. https://doi.org/10.1053/J.GASTRO.2013.11.045

Heit, C., Jackson, B. C., McAndrews, M., Wright, M. W., Thompson, D. C., Silverman, G. A., Nebert, D. W., & Vasiliou, V. (2013). Update of the human and mouse SERPIN gene superfamily. *Human Genomics*, *7*(1), 1–14. https://doi.org/10.1186/1479-7364-7-22/TABLES/3

Henderson, C. M., Fink, S. L., Bassyouni, H., Argiropoulos, B., Brown, L., Laha, T. J., Jackson, K. J., Lewkonia, R., Ferreira, P., Hoofnagle, A. N., & Marcadier, J. L. (2019). Vitamin D–Binding Protein Deficiency and Homozygous Deletion of the GC Gene . *New England Journal of Medicine*, *380*(12), 1150–1157. https://doi.org/10.1056/NEJMOA1807841/SUPPL_FILE/NEJMOA1807841_DISCLOSURES.PDF

Home - Gene - NCBI. (n.d.). Retrieved 13 April 2023, from https://www.ncbi.nlm.nih.gov/gene/

Hsieh, W. C., Mackinnon, A. C., Lu, W. Y., Jung, J., Boulter, L., Henderson, N. C., Simpson, K. J., Schotanus, B., Wojtacha, D., Bird, T. G., Medine, C. N., Hay, D. C., Sethi, T., Iredale, J. P., & Forbes, S. J. (2015). Galectin-3 regulates hepatic progenitor cell expansion during liver injury. *Gut*, *64*(2), 312–321. https://doi.org/10.1136/GUTJNL-2013-306290

Hu, C., Yang, J., Qi, Z., Wu, H., Wang, B., Zou, F., Mei, H., Liu, J., Wang, W., & Liu, Q. (2022). Heat shock proteins: Biological functions, pathological roles, and therapeutic opportunities. *MedComm*, *3*(3). https://doi.org/10.1002/MCO2.161

Huch, M., Dorrell, C., Boj, S. F., Van Es, J. H., Li, V. S. W., Van De Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M. J., Haft, A., Vries, R. G., Grompe, M., & Clevers, H. (2013). In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature 2013 494:7436*, *494*(7436), 247–250. https://doi.org/10.1038/NATURE11826

Janciauskiene, S. M., Bals, R., Koczulla, R., Vogelmeier, C., Köhnlein, T., & Welte, T. (2011). The discovery of α1-antitrypsin and its role in health and disease. *Respiratory Medicine*, *105*(8), 1129–1139. https://doi.org/10.1016/J.RMED.2011.02.002

Janciauskiene, S., Wrenger, S., Immenschuh, S., Olejnicka, B., Greulich, T., Welte, T., & Chorostowska-Wynimko, J. (2018). The Multifaceted Effects of Alpha1-Antitrypsin on Neutrophil Functions. *Frontiers in Pharmacology*, *9*(APR). https://doi.org/10.3389/FPHAR.2018.00341

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols 2017 12:4*, *12*(4), 828–863. https://doi.org/10.1038/NPROT.2017.016

Kelman, Z. (1997). PCNA: structure, functions and interactions. *Oncogene 1997 14:6*, *14*(6), 629–640. https://doi.org/10.1038/sj.onc.1200886

Kim, H. Y., Baek, G. H., Lee, W., Lee, Y. J., Shim, W. S., Choi, Y. J., Lee, B., Kim, S. K., & Kang, K. W. (2022). CD44 is involved in liver regeneration through enhanced uptake of extracellular cystine. *Clinical and Translational Medicine*, *12*(5). https://doi.org/10.1002/CTM2.873

Kim, P. T. W., & Testa, G. (2016). Living donor liver transplantation in the USA. *Hepatobiliary Surgery and Nutrition*, *5*(2), 133–140. https://doi.org/10.3978/J.ISSN.2304-3881.2015.06.01

Kon, J., Ooe, H., Oshima, H., Kikkawa, Y., & Mitaka, T. (2006). Expression of CD44 in rat hepatic progenitor cells. *Journal of Hepatology*, *45*(1), 90–98. https://doi.org/10.1016/j.jhep.2006.01.029

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., & Forbes, S. J. (2015a). Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nature Cell Biology 2014 17:8*, *17*(8), 971–983. https://doi.org/10.1038/ncb3203

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., & Forbes, S. J. (2015b). Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nature Cell Biology*, *17*(8), 971. https://doi.org/10.1038/NCB3203

Manco, R., Clerbaux, L. A., Verhulst, S., Bou Nader, M., Sempoux, C., Ambroise, J., Bearzatto, B., Gala, J. L., Horsmans, Y., van Grunsven, L., Desdouets, C., & Leclercq, I. (2019). Reactive cholangiocytes differentiate into proliferative hepatocytes with efficient DNA repair in mice with chronic liver injury. *Journal of Hepatology*, *70*(6), 1180–1191. https://doi.org/10.1016/J.JHEP.2019.02.003

Marshall, A., Rushbrook, S., Davies, S. E., Morris, L. S., Scott, I. S., Vowler, S. L., Coleman, N., & Alexander, G. (2005a). Relation between hepatocyte G1 arrest, impaired hepatic regeneration, and fibrosis in chronic hepatitis C virus infection. *Gastroenterology*, *128*(1), 33–42. https://doi.org/10.1053/j.gastro.2004.09.076

Marshall, A., Rushbrook, S., Davies, S. E., Morris, L. S., Scott, I. S., Vowler, S. L., Coleman, N., & Alexander, G. (2005b). Relation between hepatocyte G1 arrest, impaired hepatic regeneration, and fibrosis in chronic hepatitis C virus infection. *Gastroenterology*, *128*(1), 33–42. https://doi.org/10.1053/J.GASTRO.2004.09.076

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols 2019 14:3*, *14*(3), 703–721. https://doi.org/10.1038/S41596-019-0128-8

Michalopoulos, G. K., & DeFrances, M. C. (1997). Liver regeneration. *Science (New York, N.Y.)*, *276*(5309), 60–65. https://doi.org/10.1126/SCIENCE.276.5309.60

Miyajima, A., Tanaka, M., & Itoh, T. (2014). Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming. *Cell Stem Cell*, *14*(5), 561–574. https://doi.org/10.1016/J.STEM.2014.04.010

Mouse Serpin A1c/ alpha 1-Antitrypsin Antibody AF2979: R&D Systems. (n.d.). Retrieved 24 April 2023, from https://www.rndsystems.com/products/mouse-serpin-a1c-alpha1-antitrypsin-antibody_af2979

Okabe, H., Yang, J., Sylakowski, K., Yovchev, M., Miyagawa, Y., Nagarajan, S., Chikina, M., Thompson, M., Oertel, M., Baba, H., Monga, S. P., & Nejak-Bowen, K. N. (2016). Wnt signaling regulates hepatobiliary repair following cholestatic liver injury in mice. *Hepatology (Baltimore, Md.)*, *64*(5), 1652. https://doi.org/10.1002/HEP.28774

Petersen, B. E., Zajac, V. F., & Michalopoulos, G. K. (1998). Hepatic oval cell activation in response to injury following chemically induced periportal or pericentral damage in rats. *Hepatology*, *27*(4), 1030–1038. https://doi.org/10.1002/HEP.510270419

Pimpin, L., Cortez-Pinto, H., Negro, F., Corbould, E., Lazarus, J. v., Webber, L., & Sheron, N. (2018). Burden of liver disease in Europe: Epidemiology and analysis of risk factors to identify prevention policies. *Journal of Hepatology*, *69*(3), 718–735. https://doi.org/10.1016/J.JHEP.2018.05.011

Raven, A., Lu, W.-Y., Yung Man, T., Ferreira-Gonzalez, S., O'duibhir, E., Dwyer, B. J., Thomson, J. P., Meehan, R. R., Bogorad, R., Koteliansky, V., Kotelevtsev, Y., Ffrench-Constant, C., Boulter, L., & Forbes, S. J. (2017). *Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration*. https://doi.org/10.1038/nature23015

Richardson, M. M., Jonsson, J. R., Powell, E. E., Brunt, E. M., Neuschwander-Tetri, B. A., Bhathal, P. S., Dixon, J. B., Weltman, M. D., Tilg, H., Moschen, A. R., Purdie, D. M., Demetris, A. J., & Clouston, A. D. (2007). Progressive Fibrosis in Nonalcoholic Steatohepatitis: Association With Altered Regeneration and a Ductular Reaction. *Gastroenterology*, *133*(1), 80–90. https://doi.org/10.1053/J.GASTRO.2007.05.012

Robinson, M. D., & Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics (Oxford, England)*, *23*(21), 2881–2887. https://doi.org/10.1093/BIOINFORMATICS/BTM453

Rodrigo-Torres, D., Affò, S., Coll, M., Morales-Ibanez, O., Millán, C., Blaya, D., Alvarez-Guaita, A., Rentero, C., Lozano, J. J., Maestro, M. A., Solar, M., Arroyo, V., Caballería, J., van Grunsven, L. A., Enrich, C., Ginès, P., Bataller, R., & Sancho-Bru, P. (2014). The biliary epithelium gives rise to liver progenitor cells. *Hepatology (Baltimore, Md.)*, *60*(4), 1367–1377. https://doi.org/10.1002/HEP.27078

Roskams, T. A., Theise, N. D., Balabaud, C., Bhagat, G., Bhathal, P. S., Bioulac-Sage, P., Brunt, E. M., Crawford, J. M., Crosby, H. A., Desmet, V., Finegold, M. J., Geller, S. A., Gouw, A. S. H., Hytiroglou, P., Knisely, A. S., Kojiro, M., Lefkowitch, J. H., Nakanuma, Y., Olynyk, J. K., … West, A. B. (2004). Nomenclature of the finer branches of the biliary tree: canals, ductules, and ductular reactions in human livers. *Hepatology (Baltimore, Md.)*, *39*(6), 1739–1745. https://doi.org/10.1002/HEP.20130

Russell, J. O., Lu, W. Y., Okabe, H., Abrams, M., Oertel, M., Poddar, M., Singh, S., Forbes, S. J., & Monga, S. P. (2019). Hepatocyte-Specific β-Catenin Deletion During Severe Liver Injury Provokes Cholangiocytes to Differentiate Into Hepatocytes. *Hepatology (Baltimore, Md.)*, *69*(2), 742–759. https://doi.org/10.1002/HEP.30270

Santoni-Rugiu, E., Jelnes, P., Thorgeirsson, S. S., & Bisgaard, H. C. (2005). Progenitor cells in liver regeneration: molecular responses controlling their activation and expansion. *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica*, *113*(11–12), 876–902. https://doi.org/10.1111/J.1600-0463.2005.APM_386.X

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *50*(D1), D20. https://doi.org/10.1093/NAR/GKAB1112

Schaum, N., Karkanias, J., Neff, N. F., May, A. P., Quake, S. R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M. B., Chen, S., Green, F., Jones, R. C., Maynard, A., Penland, L., Pisco, A. O., Sit, R. v., Stanley, G. M., Webber, J. T., … Weissman, I. L. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature 2018 562:7727*, *562*(7727), 367–372. https://doi.org/10.1038/S41586-018-0590-4

Schmitt, M., Metzger, M., Gradl, D., Davidson, G., & Orian-Rousseau, V. (2015). CD44 functions in Wnt signaling by regulating LRP6 localization and activation. *Cell Death and Differentiation*, *22*(4), 677–689. https://doi.org/10.1038/CDD.2014.156

Shin, S., & Kaestner, K. H. (2014). The Origin, Biology, and Therapeutic Potential of Facultative Adult Hepatic Progenitor Cells. *Current Topics in Developmental Biology*, *107*, 269. https://doi.org/10.1016/B978-0-12-416022-4.00010-X

So, J., Kim, A., Lee, S.-H., & Shin, D. (2020). Liver progenitor cell-driven liver regeneration. *Experimental & Molecular Medicine*, *52*, 1230–1238. https://doi.org/10.1038/s12276-020-0483-0

Spee, B., Carpino, G., Schotanus, B. A., Katoonizadeh, A., Vander Borght, S., Gaudio, E., & Roskams, T. (2010). Characterisation of the liver progenitor cell niche in liver diseases: potential involvement of Wnt and Notch signalling. *Gut*, *59*(2), 247–257. https://doi.org/10.1136/GUT.2009.188367

Strnad, P., McElvaney, N. G., & Lomas, D. A. (2020). Alpha1-Antitrypsin Deficiency. *The New England Journal of Medicine*, *382*(15), 1443–1455. https://doi.org/10.1056/NEJMRA1910234

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605. https://doi.org/10.1093/NAR/GKAA1074

Tchorz, J. S., Kinter, J., Müuller, M., Tornillo, L., Heim, M. H., & Bettler, B. (2009). Notch2 signaling promotes biliary epithelial cell fate specification and tubulogenesis during bile duct development in mice. *Hepatology (Baltimore, Md.)*, *50*(3), 871–879. https://doi.org/10.1002/HEP.23048

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L. P., & Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science*, *31*(1), 8–22. https://doi.org/10.1002/PRO.4218

Understanding UMAP. (n.d.). Retrieved 8 April 2023, from https://pair-code.github.io/understanding-umap/

Xiao, J. C., Ruck, P., Adam, A., Wang, T. X., & Kaiserling, E. (2003). Small epithelial cells in human liver cirrhosis exhibit features of hepatic stem-like cells: immunohistochemical, electron microscopic and immunoelectron microscopic findings. *Histopathology*, *42*(2), 141–149. https://doi.org/10.1046/J.1365-2559.2003.01544.X

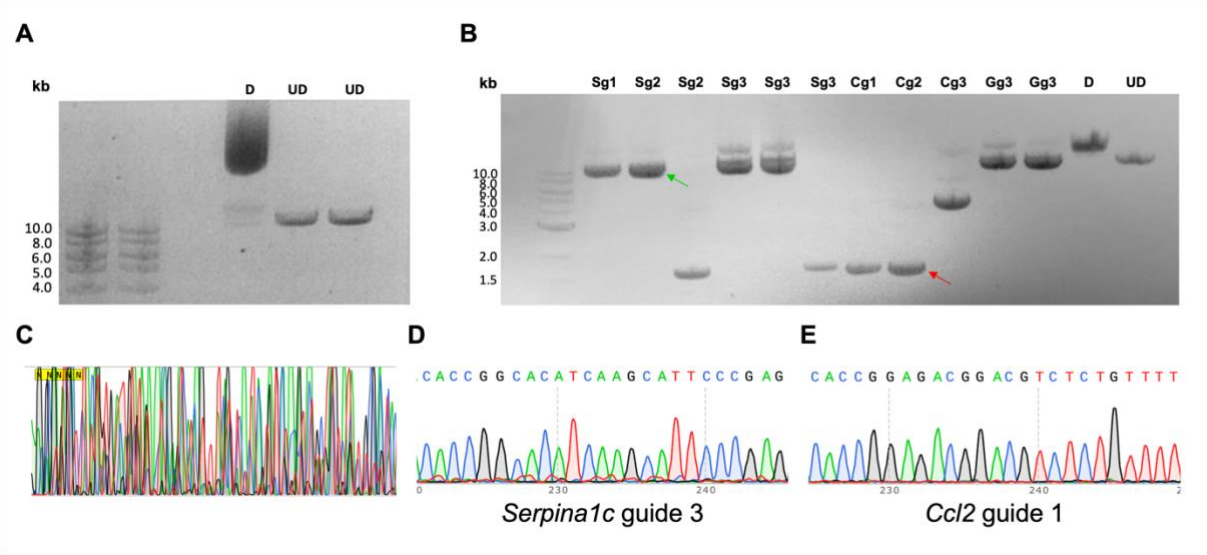
Yoon, S. M., Gerasimidou, D., Kuwahara, R., Hytiroglou, P., Yoo, J. E., Park, Y. N., & Theise, N. D. (2011). Epithelial cell adhesion molecule (EpCAM) marks hepatocytes newly derived from stem/progenitor cells in humans. *Hepatology (Baltimore, Md.)*, *53*(3), 964–973. https://doi.org/10.1002/HEP.24122

Yovchev, M. I., Grozdanov, P. N., Zhou, H., Racherla, H., Guha, C., & Dabeva, M. D. (2008). Identification of adult hepatic progenitor cells capable of repopulating injured rat liver. *Hepatology (Baltimore, Md.)*, *47*(2), 636–647. https://doi.org/10.1002/HEP.22047

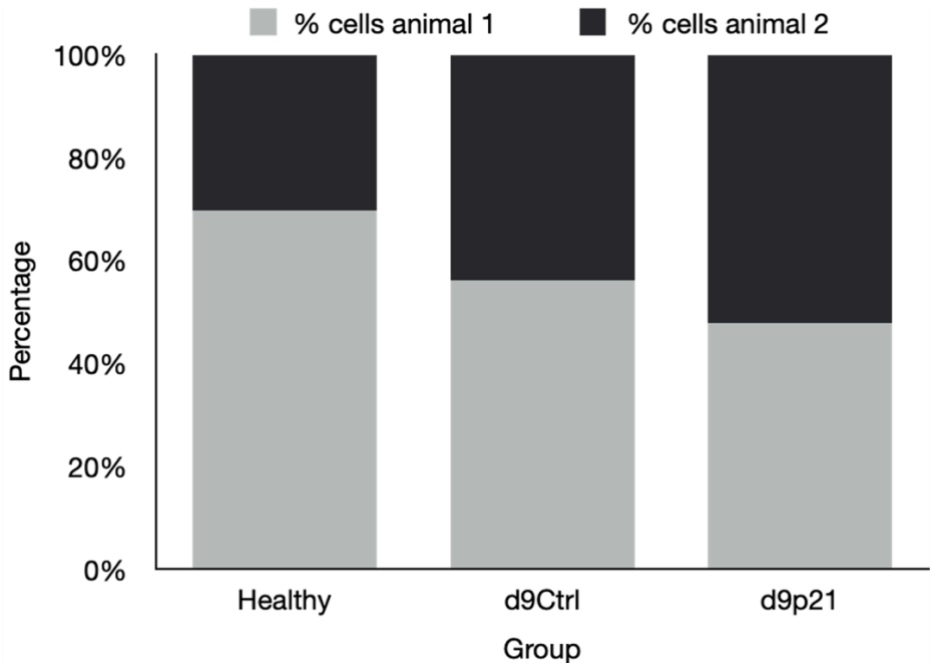
Yu, D., Huber, W., & Vitek, O. (2013). Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. *Bioinformatics (Oxford, England)*, *29*(10), 1275–1282. https://doi.org/10.1093/BIOINFORMATICS/BTT143

Zong, Y., Panikkar, A., Xu, J., Antoniou, A., Raynaud, P., Lemaigre, F., & Stanger, B. Z. (2009). Notch signaling controls liver development by regulating biliary differentiation. *Development (Cambridge, England)*, *136*(10), 1727. https://doi.org/10.1242/DEV.029140

Supplementary figures



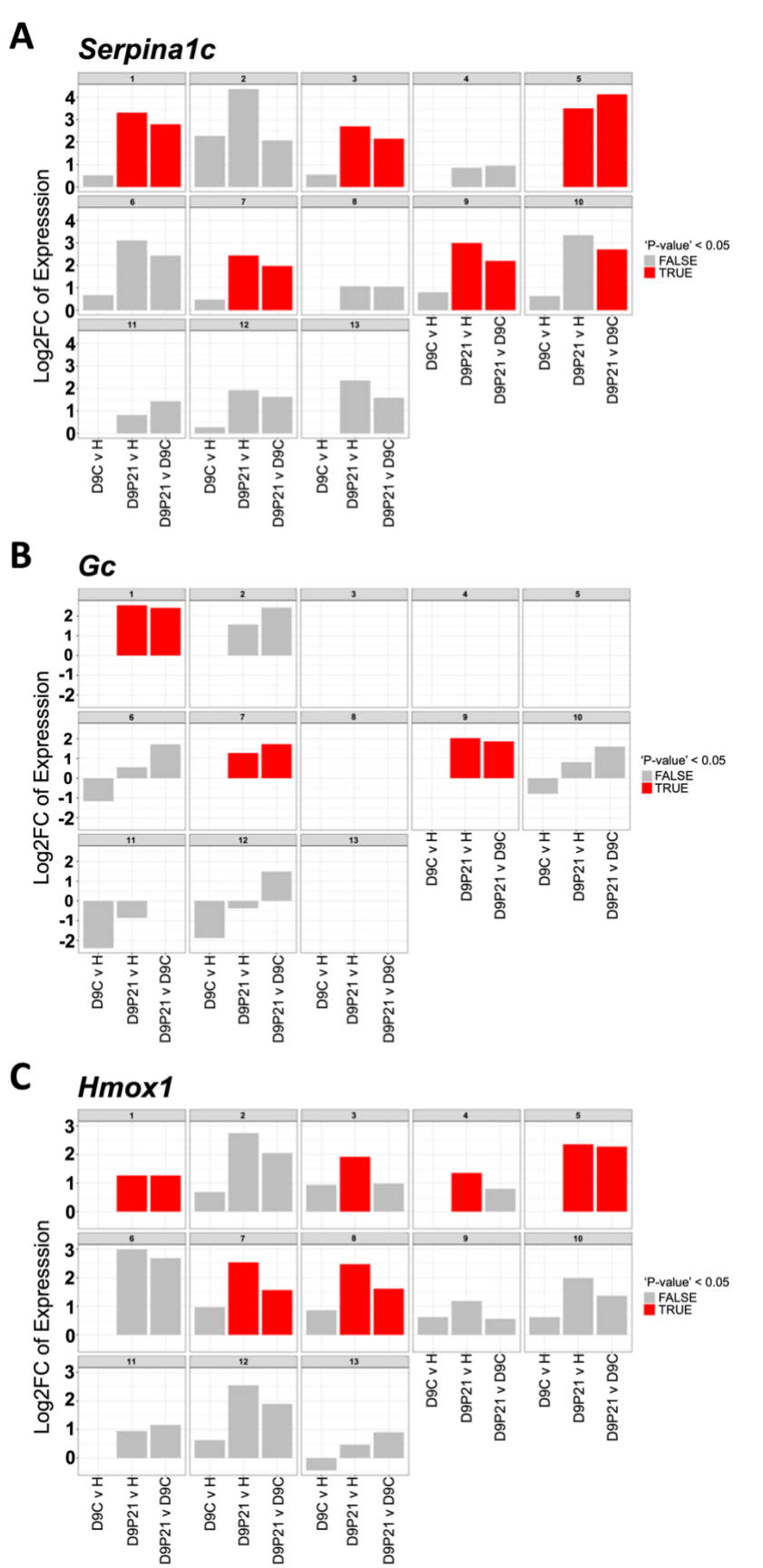
Supplementary figure 1: CRISPR-editing of *Serpina1c*, *Gc*, and *Ccl2.*
(**A**) Gel electrophoresis showing the molecular weight of digested (D) and undigested (UD) LentiCRISPRv2 plasmid. DNA ladder (New England Biolabs, #N3232). (**B**) Gel electrophoresis showing plasmids containing *Serpina1c* (guide 1-3) (Sg1-3), *Ccl2* (guide 1-3) (Cg1-3), *Gc* (guide 3) (Gg3), a digested control LentiCRISPRv2 plasmid (D), and an undigested control LentiCRISPRv2 plasmid (UD). Green arrow points towards the correct molecular weight of the LentiCRISPRv2 plasmid. Red arrow points towards a plasmid of incorrect molecular weight indicative for contamination. DNA ladder (New England Biolabs, #N3232). (**C**) Sanger sequencing result of plasmid contamination showing that sequencing was not able to determine any bases. (**D**) Correct Sanger sequencing result of the LentiCRISPRv2 plasmid containing *Serpina1c* (guide 3). (**E**) Incorrect Sanger sequencing result of the LentiCRISPRv2 plasmid for *Ccl2* (guide 3).



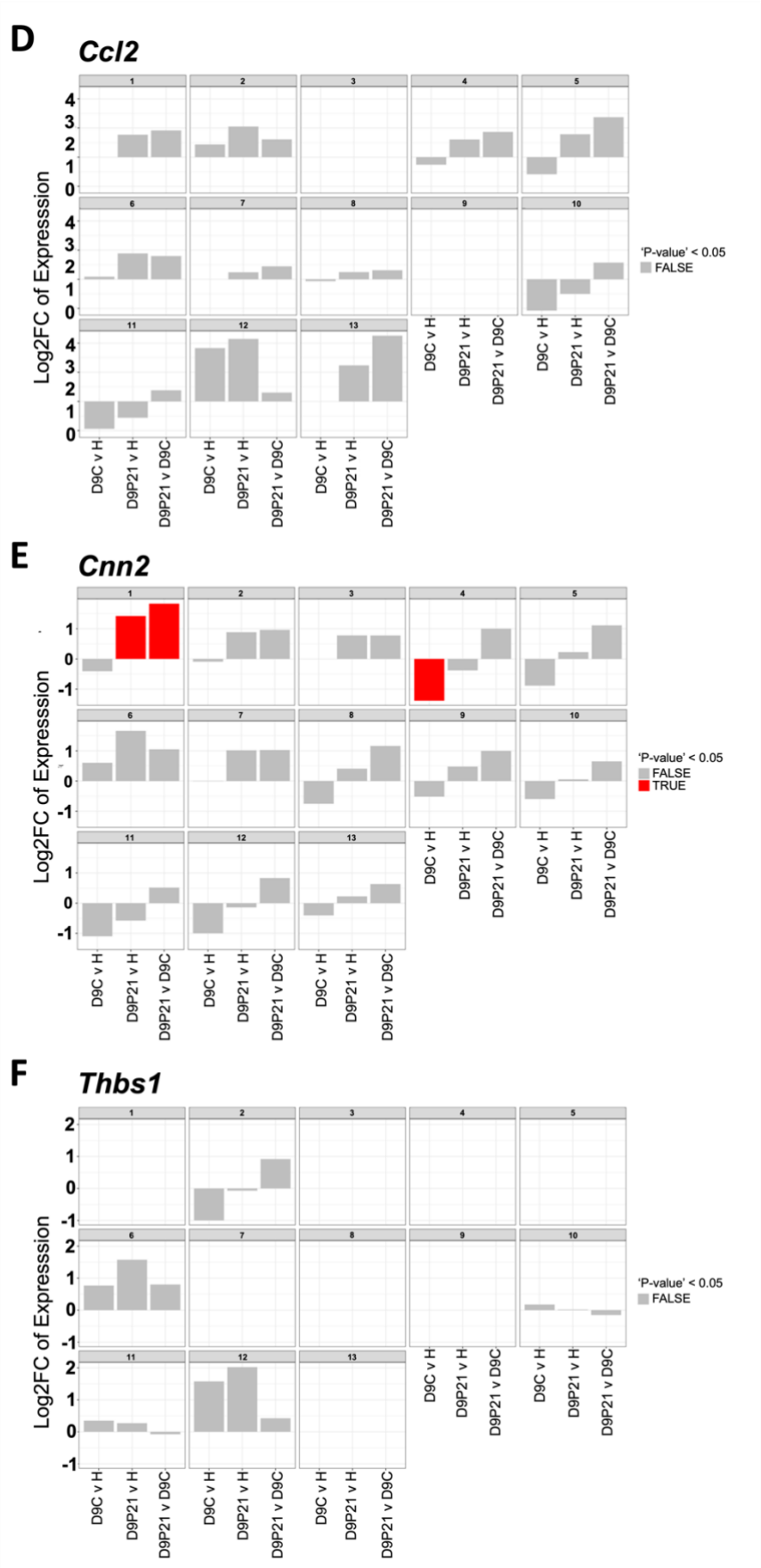
Supplementary figure 2: Distribution of biological replicas per condition in the scRNA-seq dataset.

Distribution of BECs in each condition of the scRNA-seq dataset presented as percentage of cells derived from each biological replicate (n=2 per condition).

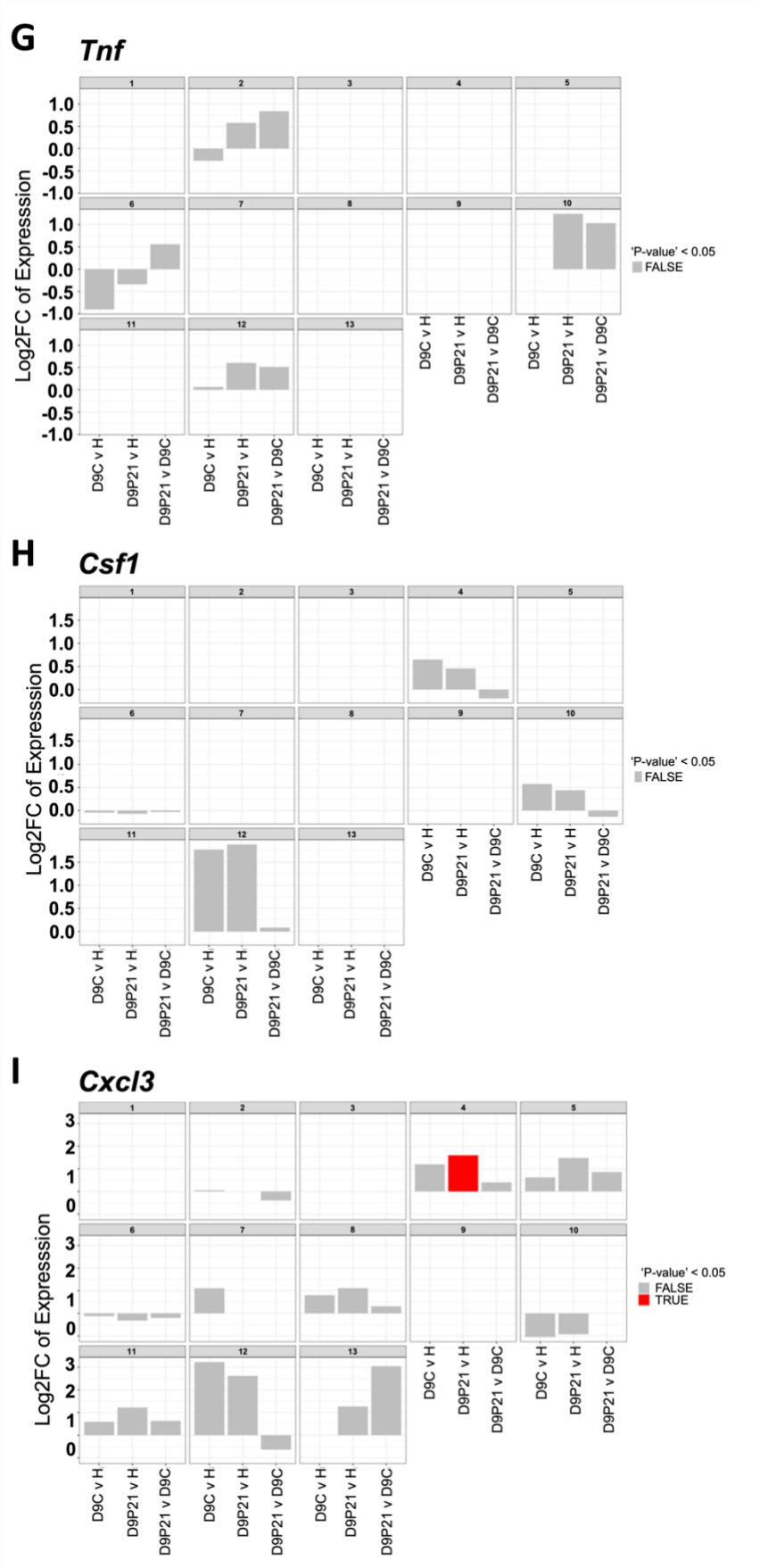
Supplementary figure 3: Expression of genes of interest in the experimental groups within clusters



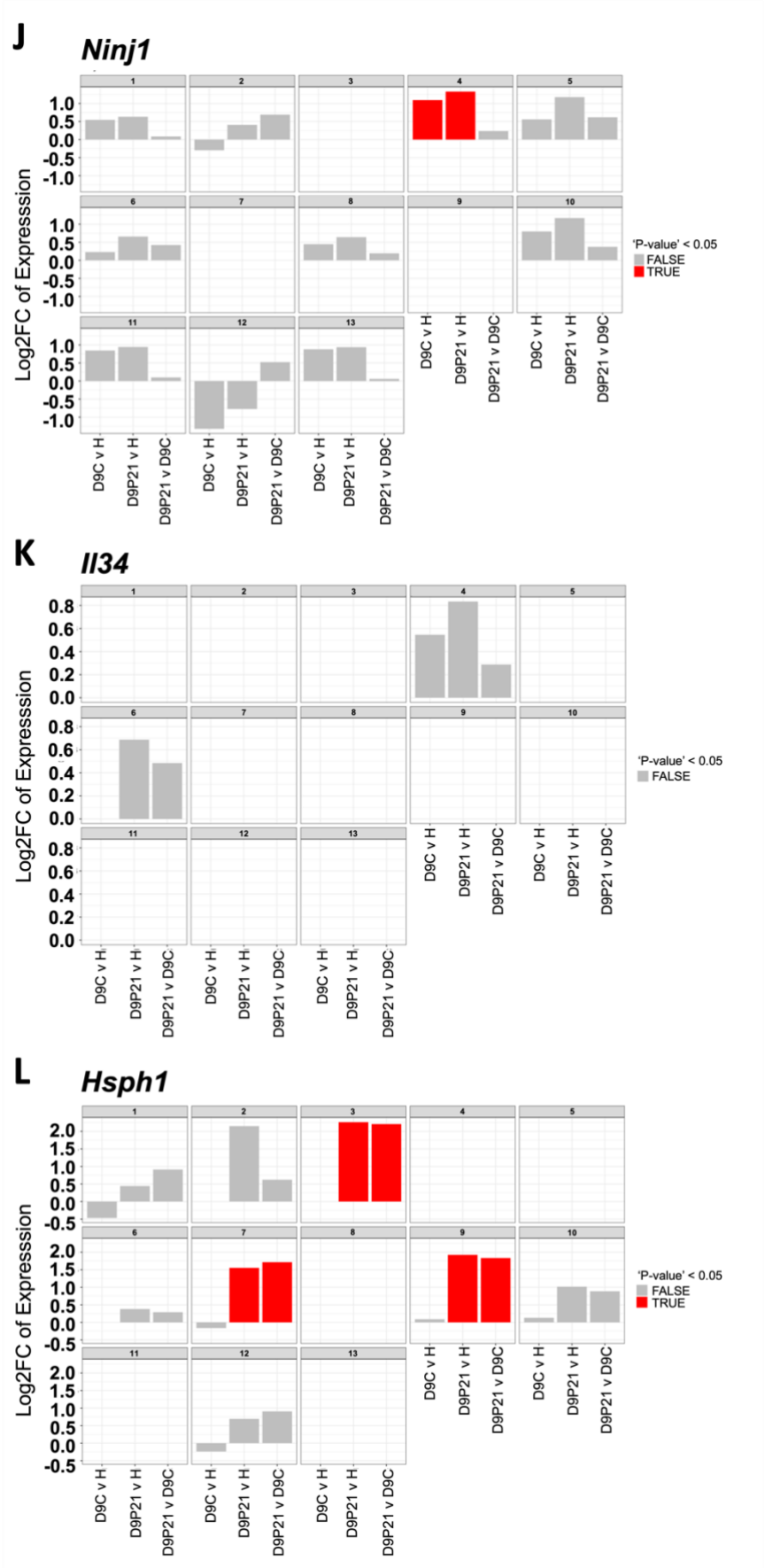
Supplementary figure 3: Expression of genes of interest in the experimental groups within clusters



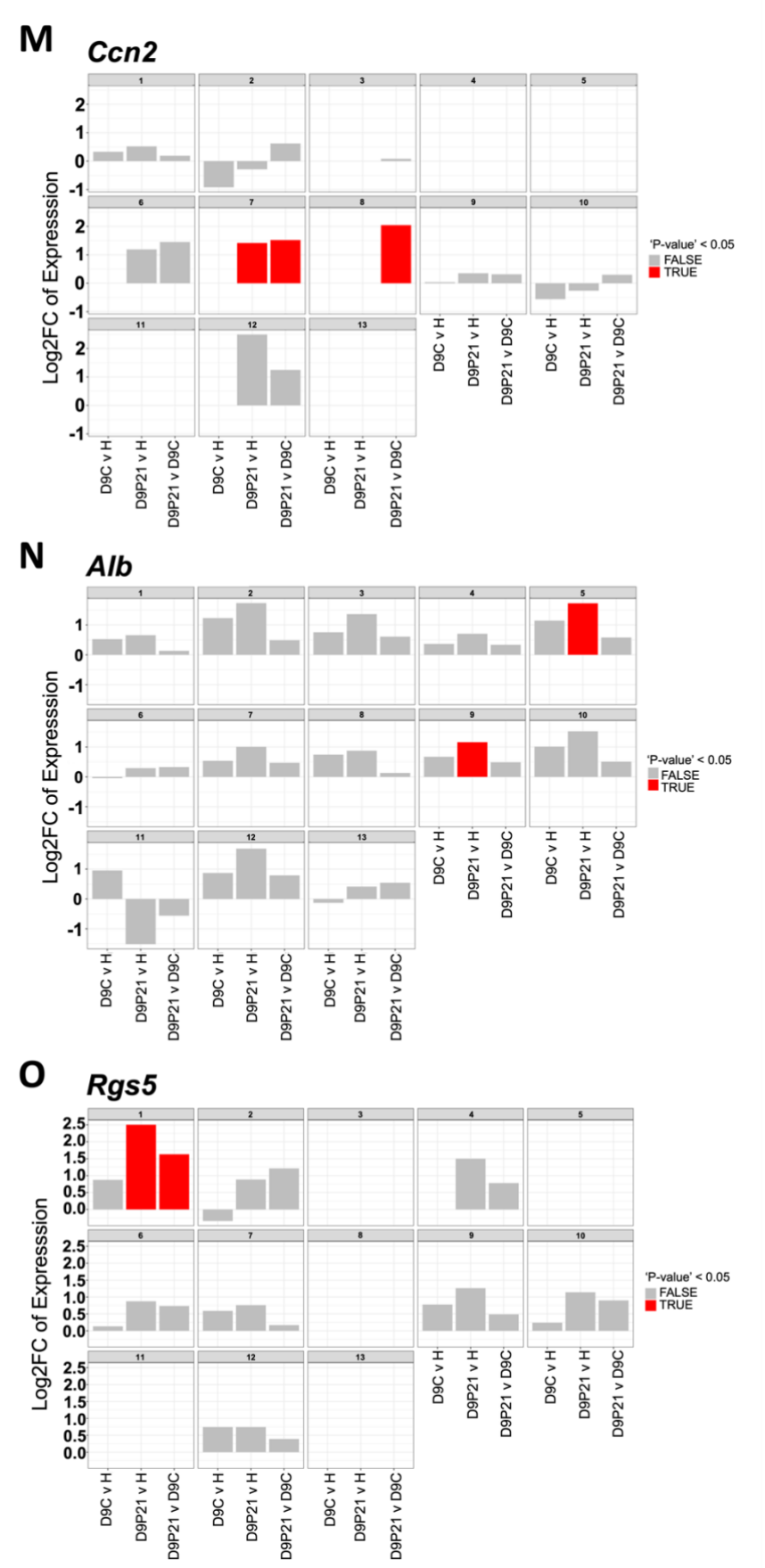
Supplementary figure 3: Expression of genes of interest in the experimental groups within clusters

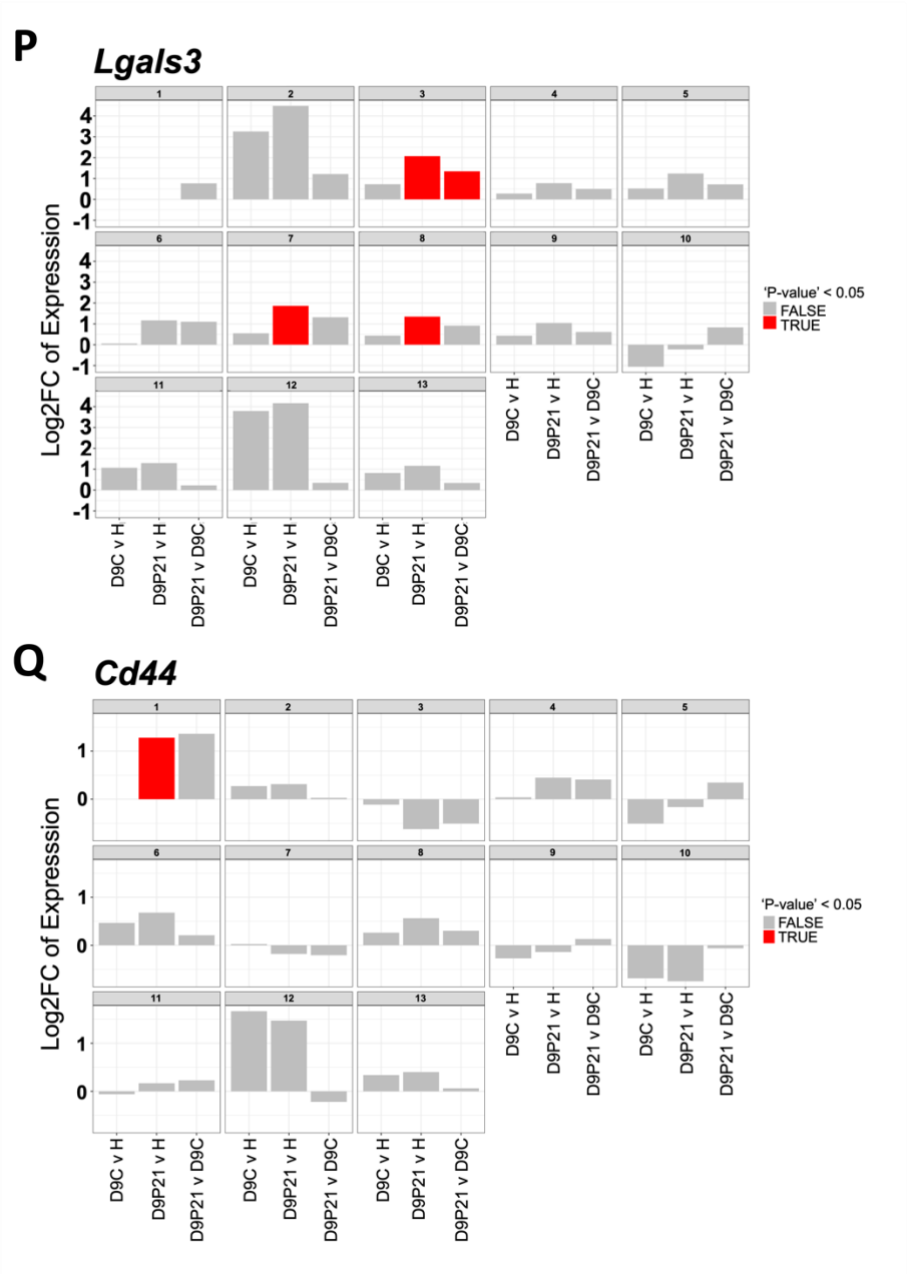


Supplementary figure 3: Expression of genes of interest in the experimental groups within clusters



Supplementary figure 3: Expression of genes of interest in the experimental groups within clusters





Supplementary figure 3: Expression of genes of interest in the experimental groups within clusters.

Comparing expression of *Serpina1c* (**A**), *Gc* (**B**), *Hmox1* (**C**), *Ccl2* (**D**), *Cnn2* (**E**), *Thbs1* (**F**), *Tnf* (**G**), *Csf1* (**H**), *Cxcl3* (**I**), *Ninj1* (**J**), *Il34* (**K**), *Hsph1* (**L**), *Ccn2* (**M**), *Alb* (**N**), *Rgs5* (**O**), *Lgals3* (**P**), and *Cd44* (**Q**) between healthy (H), d9ctrl (D9C), and d9p21 (D9P21) conditions within each cluster. Relative gene expression is presented as Log2 Fold Change (Log2FC) of expression.