Human Molecular Genetics, 2020, Vol. 29, No. 18 3145–3153

doi: 10.1093/hmg/ddaa188
Advance Access Publication Date: 21 August 2020
Bioinformatics Article

BIOINFORMATICS ARTICLE

Integrative analysis of scRNA-seq and GWAS data pinpoints periportal hepatocytes as the relevant liver cell types for blood lipids

Xingjie Hao[1,*], Kai Wang[1], Chengguqiu Dai[1], Zeyang Ding[2], Wei Yang[3], Chaolong Wang[1,4] and Shanshan Cheng[1,*]

[1]Department of Epidemiology and Biostatistics, Key Laboratory for Environment and Health, School of Public Health, [2]Hepatic Surgery Center, Tongji Hospital, [3]Department of Nutrition and Food Hygiene, School of Public Health and [4]Department of Orthopedic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China

*To whom correspondence should be addressed at: Tongji Medical College, Huazhong University of Science and Technology, 13 Hongkong Road, Wuhan 430030, Hubei, China. Tel: 8602783692757; Email: sscheng@hust.edu.cn (S. Cheng) or xingjie@hust.edu.cn (X. Hao)

Abstract

Liver, a heterogeneous tissue consisting of various cell types, is known to be relevant for blood lipid traits. By integrating summary statistics from genome-wide association studies (GWAS) of lipid traits and single-cell transcriptome data of the liver, we sought to identify specific cell types in the liver that were most relevant for blood lipid levels. We conducted differential expression analyses for 40 cell types from human and mouse livers in order to construct the cell-type specifically expressed gene sets, which we refer to as construction of the liver cell-type specifically expressed gene sets (CT-SEGS). Under the assumption that CT-SEGS represented specific functions of each cell type, we applied stratified linkage disequilibrium score regression to determine cell types that were most relevant for complex traits and diseases. We first confirmed the validity of this method (of delineating functionally relevant cell types) by identifying the immune cell types as relevant for autoimmune diseases. We further showed that lipid GWAS signals were enriched in the human and mouse periportal hepatocytes. Our results provide important information to facilitate future cellular studies of the metabolic mechanism affecting blood lipid levels.

Introduction

The liver is a heterogeneous tissue critical for metabolic and immune functions. The cellular organization of the liver is based on the building block of the hepatic acinus with different cell types inside. Among these cell types, hepatocytes make up the largest proportion of cell populations, and play an important role in metabolic, secretory and endocrine functions (1). Kupffer cells are the liver resident macrophages and have been described as the immunological sentinels of the liver (2–4). Hepatic stellate cells in the space of Disse are the main storage site for vitamin A, and are the major contributor to liver fibrosis (5). Some liver-infiltrating lymphocytes, including B cells, T cells and natural killer cells, are distributed in specific patterns, while many details remain unknown in terms of cellular locations and functions of these lymphocytes (6,7).

As an important tissue for metabolic and immune processes, the liver is constantly exposed to gut-derived dietary and micro-

Received: March 18, 2020. **Revised:** August 10, 2020. **Accepted:** August 18, 2020

bial antigens, and is thus relevant for many complex traits and diseases (8). The bile acids produced by the liver are necessary for the breakdown of fat and emulsification of lipids. In addition, the liver plays important roles in many other metabolic processes, including regulation of glycogen storage, decomposition of red blood cells, and production of hormones. Among the solid organs in the body, the liver has the largest population of tissue-resident macrophages, which can affect the progression of liver diseases (3,6).

Genome-wide association studies (GWAS) of blood lipid levels, including high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC) and triglycerides (TG), have identified numerous association signals enriched in the liver-specific expressed genes and the liver-specific epigenetic modification regions (9–15). Although it is clear that the liver is a relevant tissue for blood lipid levels, the resolution is insufficient to guide subsequent experiments to connect the GWAS results to cellular functions, because of the heterogeneity of cell type composition in the liver.

Recently, single-cell RNA sequencing (scRNA-seq) has emerged as a useful approach to quantify the transcriptome of individual cells and to cluster cells into putative populations based on their gene expressions (16,17). Many cell populations with different functions have been identified in human and mouse livers by scRNA-seq (6,7,18–21). Integrative analyses of scRNA-seq data and GWAS results have helped pinpoint the most relevant brain cell types for several neurological disorders (15,22–24) and insomnia (25). Given that the liver is a heterogeneous tissue consisting of various cell populations with distinct functions (6,7,18–21), identification of specific liver cell types relevant for blood lipid levels will provide essential information for future functional and cellular studies to understand the metabolism of lipids. Thus, in this study, we performed integrative analysis of GWAS summary statistics of blood lipid levels and scRNA-seq data from the human liver and the mouse liver, with the goal to pinpoint specific liver cell types relevant to blood lipid levels.

Results

Construction of the liver cell-type specifically expressed gene sets (CT-SEGS)

We selected 186 and 274 upregulated genes for each cell type from human and mouse livers, respectively, to construct the CT-SEGS (Fig. 1 and Methods). We provided a global picture of the sharing of upregulated genes across cell types of human and mouse liver by using the generalized Jaccard similarity index (i.e. the size of the intersection divided by the size of the union of the sample sets), a measure of overlap between upregulated genes in two cell types (Fig. 2). The average cross-cell-type Jaccard similarity indices were 0.009 in human liver CT-SEGS and 0.008 in mouse liver CT-SEGS. In addition, the average Jaccard similarity index was 0.003 between human and mouse liver CT-SEGS (Fig. 2). Through hierarchical clustering, similar cell types from both human and mouse liver tended to cluster together, broadly forming four cell type groups (Supplementary Note): endothelial cells, immune cells, hepatocyte cells and epithelial cells. For example, the endothelial cell group contained four cell types (i.e. central venous liver sinusoidal endothelial cells (LSECs), periportal (PP) LSECs, portal endothelial cells and stellate cells from the human liver), and the epithelial cell group contained three cell types (i.e. cholangiocytes from human liver, epithelia cell_Spp1 high and epithelial cell from the mouse liver) (Figs 2 and 3). The average Jaccard similarity indices were 0.189, 0.062, 0.085 and 0.092 within each cell type group (endothelial cell, immune cell, hepatocyte cell and epithelial cell).

Application to autoimmune diseases to confirm the functional relevance of CT-SEGS

To check the biological specificity of our liver CT-SEGS, we applied the stratified linkage disequilibrium score regression (LDSC) enrichment analyses to five well studied autoimmune diseases, including inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, rheumatoid arthritis and type 1 diabetes (Supplementary Material, Table S1). As expected, we found that the GWAS signals of autoimmune diseases tended to be enriched in both human and mouse liver immune CT-SEGS (Table 1 and Fig. 3). In particular, we found no enrichment in the hepatocytes for primary biliary cholangitis (Fig. 3C), a chronic disease in which the bile ducts in the liver are slowly destroyed. Interestingly, in addition to immune cells, we also identified cholangiocytes, the epithelial cells in bile ducts, as a top relevant cell type for primary biliary cholangitis ($P = 0.0188$) (Table 1).

Lipid GWAS signals were enriched in PP hepatocyte functional regions

Finally, we analyzed four lipid traits (HDL, LDL, TC and TG) by integrating the scRNA-seq profiles of human and mouse livers. We identified human Hep 3 and Hep 5 as the relevant cell types for blood lipid levels after Bonferroni correction for multiple tests ($P < 0.05/n$, where $n = 4$ is the number of cell type groups) (Table 2 and Fig. 4). We adjusted for the number of cell type groups rather than the number of cell types because gene expression levels in cell types from the same group were highly correlated and shared some specifically expressed genes (Fig. 2). Specifically, GWAS signals for HDL were enriched in human Hep 3, while the GWAS signals for LDL, TC and TG were significantly enriched in human Hep 5. In addition, among the four hepatocyte cell types in mouse liver, GWAS signals for blood lipids were only nominally significantly ($P < 0.05$) enriched in mouse PP hepatocyte cells. Specially, we found that mouse PP hepatocyte cells were significantly ($P = 1.05E\text{-}4$) relevant for the lipid traits after meta-analysis of four lipid traits simultaneously (Supplementary Note). To test the robustness of our results, we also changed the selection criteria of upregulated genes in CT-SEGS by applying a P value threshold rather than fixing the number of genes (details were provided in Supplementary Note). We found that the GWAS signals for HDL were still enriched in human Hep 3, and the GWAS signals for LDL, TC and TG were nominally significantly ($P < 0.05$) enriched in human Hep 5 and mouse PP hepatocyte cells (Supplementary Note). Except for hepatocytes from human and mouse, we did not identify any other cell types (*e.g.* immune cell types) whose CT-SEGS were enriched for lipid GWAS signals (Fig. 4).

Discussion

Previous studies have identified the liver as a relevant tissue for blood lipid levels by integrating GWAS results with functional annotations, such as those derived from bulk RNA-seq and chromatin immunoprecipitation sequencing (9–14,26–28). These studies, however, had limited resolution to identify specific liver cell types, because the functional annotations were derived from bulk sequencing of thousands of heterogeneous cells in the

Table 1. Estimates of the enrichment coefficient of human and mouse liver CT-SEGS for human autoimmune diseases

Liver	Trait	Cell type	τ	SE(τ)	P value
Human	IBD	NK-like cells	8.63E-08	4.87E-08	0.0382
Mouse	IBD	B cell_Jchain high	7.21E-08	4.08E-08	0.0388
Mouse	IBD	T cell_Gzma high	1.05E-07	5.79E-08	0.0344
Mouse	IBD	T cell_Trbc2 high	1.55E-07	7.34E-08	0.0176
Human	**MS**	**NK-like cells**	**2.09E-07**	**6.21E-08**	**0.0004**
Human	MS	CD3+ $\alpha\beta$ T cells	6.46E-08	3.37E-08	0.0278
Human	**MS**	**$\gamma\delta$ T cells 1**	**1.74E-07**	**5.69E-08**	**0.0011**
Mouse	**MS**	**Granulocyte**	**7.05E-08**	**3.02E-08**	**0.0099**
Mouse	MS	Erythroblast_Hbb-bs high	1.01E-07	4.31E-08	0.0093
Human	MS	Mature B cells	7.33E-08	4.07E-08	0.0358
Mouse	MS	B cell_Fcmr high	5.62E-08	3.26E-08	0.0423
Mouse	**MS**	**Dendritic cell_Cst3 high**	**1.09E-07**	**4.73E-08**	**0.0104**
Mouse	PBC	Kupffer cell	1.17E-07	7.07E-08	0.0492
Mouse	PBC	B cell_Jchain high	1.56E-07	7.59E-08	0.0198
Mouse	PBC	T cell_Trbc2 high	1.74E-07	9.94E-08	0.0398
Human	PBC	Cholangiocytes	1.57E-07	7.53E-08	0.0188
Human	**RA**	**Mature B cells**	**6.38E-08**	**2.80E-08**	**0.0113**
Mouse	**RA**	**B cell_Jchain high**	**6.53E-08**	**2.57E-08**	**0.0056**
Mouse	RA	T cell_Trbc2 high	8.07E-08	3.75E-08	0.0156
Mouse	**RA**	**B cell_Fcmr high**	**6.77E-08**	**2.62E-08**	**0.0049**
Human	**T1D**	**NK-like cells**	**1.34E-07**	**4.96E-08**	**0.0035**
Human	**T1D**	**CD3+ $\alpha\beta$ T cells**	**1.05E-07**	**4.27E-08**	**0.0069**
Human	T1D	Mature B cells	7.43E-08	3.94E-08	0.0296
Mouse	T1D	T cell_Gzma high	6.97E-08	3.79E-08	0.0328
Mouse	T1D	T cell_Trbc2 high	1.08E-07	5.22E-08	0.0190
Mouse	**T1D**	**B cell_Fcmr high**	**1.17E-07**	**3.96E-08**	**0.0015**
Mouse	**T1D**	**Dendritic cell_Cst3 high**	**8.91E-08**	**3.03E-08**	**0.0016**
Mouse	T1D	Hepatocyte_mt-Nd4 high	6.32E-08	3.39E-08	0.0313
Mouse	T1D	Epithelial cell	6.42E-08	3.55E-08	0.0354

CT-SEGS passing the Bonferroni significance thresholds are highlighted in bold. Only cell types with $P < 0.05$ for enrichment tests are listed. CT-SEGS, cell-type specifically expressed gene sets; IBD, inflammatory bowel disease; MS, multiple sclerosis; PBC, primary biliary cirrhosis; RA, rheumatoid arthritis; SE, standard error; T1D, type 1 diabetes.

Table 2. Estimates of the enrichment coefficient of human and mouse liver CT-SEGS for blood lipid levels

Liver	Trait	Cell type	τ	SE(τ)	P value
Mouse	HDL	Dendritic cell_Siglech high	3.02E-08	1.79E-08	0.0456
Mouse	HDL	Periportal (PP) hepatocyte	3.08E-08	1.81E-08	0.0442
Human	**HDL**	**Hep 3**	**8.66E-08**	**3.72E-08**	**0.0099**
Human	HDL	Hep 5	5.01E-08	2.44E-08	0.0200
Human	HDL	Hep 2	2.87E-08	1.34E-08	0.0158
Mouse	LDL	PP hepatocyte	3.71E-08	2.05E-08	0.0355
Mouse	LDL	Hepatocyte_Fabp1 high	2.88E-08	1.64E-08	0.0393
Human	**LDL**	**Hep 5**	**1.18E-07**	**4.13E-08**	**0.0022**
Mouse	TC	Neutrophil_Ngp high	3.58E-08	1.90E-08	0.0300
Mouse	TC	PP hepatocyte	5.38E-08	2.51E-08	0.0159
Human	TC	Hep 3	7.70E-08	4.21E-08	0.0338
Human	TC	Hep 4	8.08E-08	4.20E-08	0.0272
Human	**TC**	**Hep 5**	**1.17E-07**	**4.21E-08**	**0.0027**
Human	TG	$\gamma\delta$ T cells 2	2.69E-08	1.39E-08	0.0267
Human	TG	Non-inflammatory macrophages	2.48E-08	1.32E-08	0.0305
Mouse	TG	Erythroblast_Hbb-bt high	2.85E-08	1.35E-08	0.0172
Mouse	TG	PP hepatocyte	4.31E-08	2.28E-08	0.0296
Human	TG	Hep 3	4.37E-08	2.58E-08	0.0452
Human	**TG**	**Hep 5**	**6.46E-08**	**2.59E-08**	**0.0062**
Human	TG	Hep 1	4.06E-08	1.90E-08	0.0160

CT-SEGS passing the Bonferroni significance thresholds are highlighted in bold. Only cell types with $P < 0.05$ for enrichment tests are listed. HDL, high-density lipoprotein; LDL, low-density lipoprotein; SE, standard error; TC, total cholesterol; TG, triglycerides.

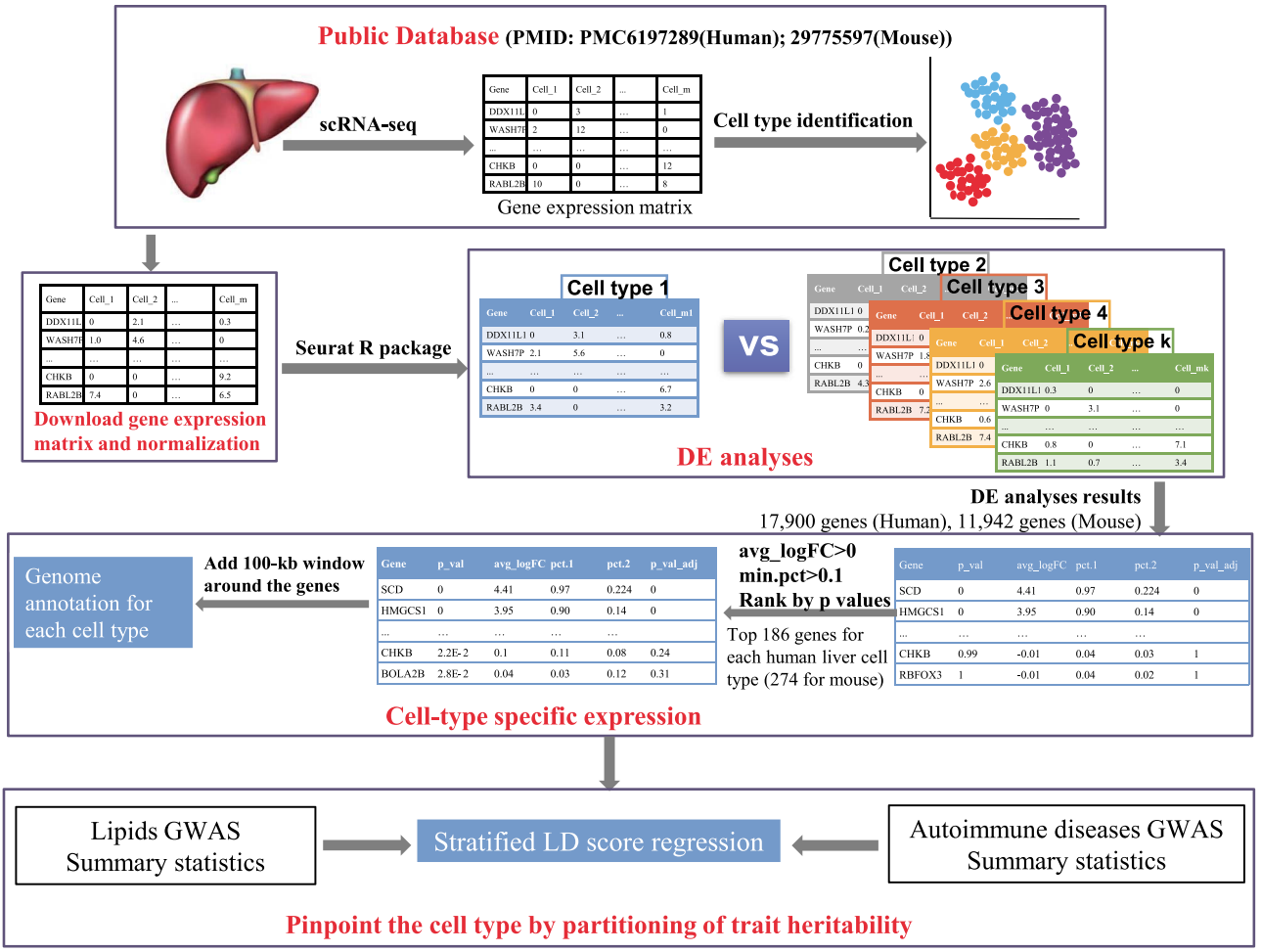

Figure 1. The flowchart of the study. For each cell type in the human and mouse liver scRNA-seq gene expression matrix, we used the likelihood-ratio test for differential expression for each gene. We then select the upregulated genes, rank the genes by the *P* values, take the top genes and add a 100-kb window to get a genome annotation. We used stratified LD score regression to test whether this annotation is significantly enriched for per-SNP heritability, conditional on the baseline model and the set of all genes.

liver tissue. In this study, we pinpointed specific liver cell types relevant for the lipid traits by leveraging two recent scRNA-seq datasets derived from human and mouse livers, respectively. As a proof of concept, we performed the same analysis and showed that in the liver only the immune cell types were relevant to autoimmune diseases, as expected.

Using upregulated genes as the specifically expressed gene sets has been shown to be effective in identifying relevant tissues and cell types for many complex diseases and traits (11,13,15,23). Because different tissues and cell types might share chromatin states in the epigenome map (29), it is common that gene expressions are correlated in similar cell types, as reflected by the larger Jaccard similarity indices within cell type groups than those across groups in our analysis (Fig. 2). Our results also confirmed that similar cell types from human and mouse analogous tissues were clustered together based on upregulated genes from scRNA-seq data (30). The small Jaccard similarity indices between most cell type pairs suggested the upregulated genes in the CT-SEGS could capture the specificity of each cell type, as confirmed by our analysis of the autoimmune diseases (Fig. 3). In addition, our stratified LDSC analysis tested each CT-SEGS annotation conditional on 53 general annotations and the set of all genes that were common to all cell types (details in Methods) (10,11). Finally, we have tested the robustness of our results with different selection criteria of upregulated genes in the CT-SEGS (Supplementary Note).

We found that GWAS signals for lipid traits were significantly enriched in the functional regions of Hep 3 and Hep 5 in the human liver (Table 2). As shown by the Jaccard similarity index in Fig. 2, human Hep 3 and Hep 5 cells share a larger proportion of specifically expressed genes with mouse PP hepatocyte cells than with mouse pericentral (PC) hepatocyte cells (0.147 vs 0.072 for Hep 3 and 0.217 vs 0.065 for Hep 5). Based on gene expression patterns between the human hepatocyte cell types and mouse zonated cell types, human Hep 3 and Hep 5 cells are transcriptionally similar to PP mouse layers (6,18). Consistently, we found that GWAS signals for lipid traits were nominally significantly ($P < 0.05$) enriched in mouse PP hepatocyte cells. We thus speculate that the lipid relevant liver cell types were likely located at the outer PP layers in the hepatic lobule. Interestingly, the spatial construction of mouse liver by scRNA-seq suggested that the mammalian liver has different zones to optimize the liver functions and to act in distinct processes (18). The outer PP layers in the hepatic lobule produced higher levels of enzymes for energy-demanding tasks such as gluconeogenesis and ureagenesis (18), consistent with their potential role in lipid metabolism, whereas the inner layers were specialized in glycolysis and xenobiotic metabolism.

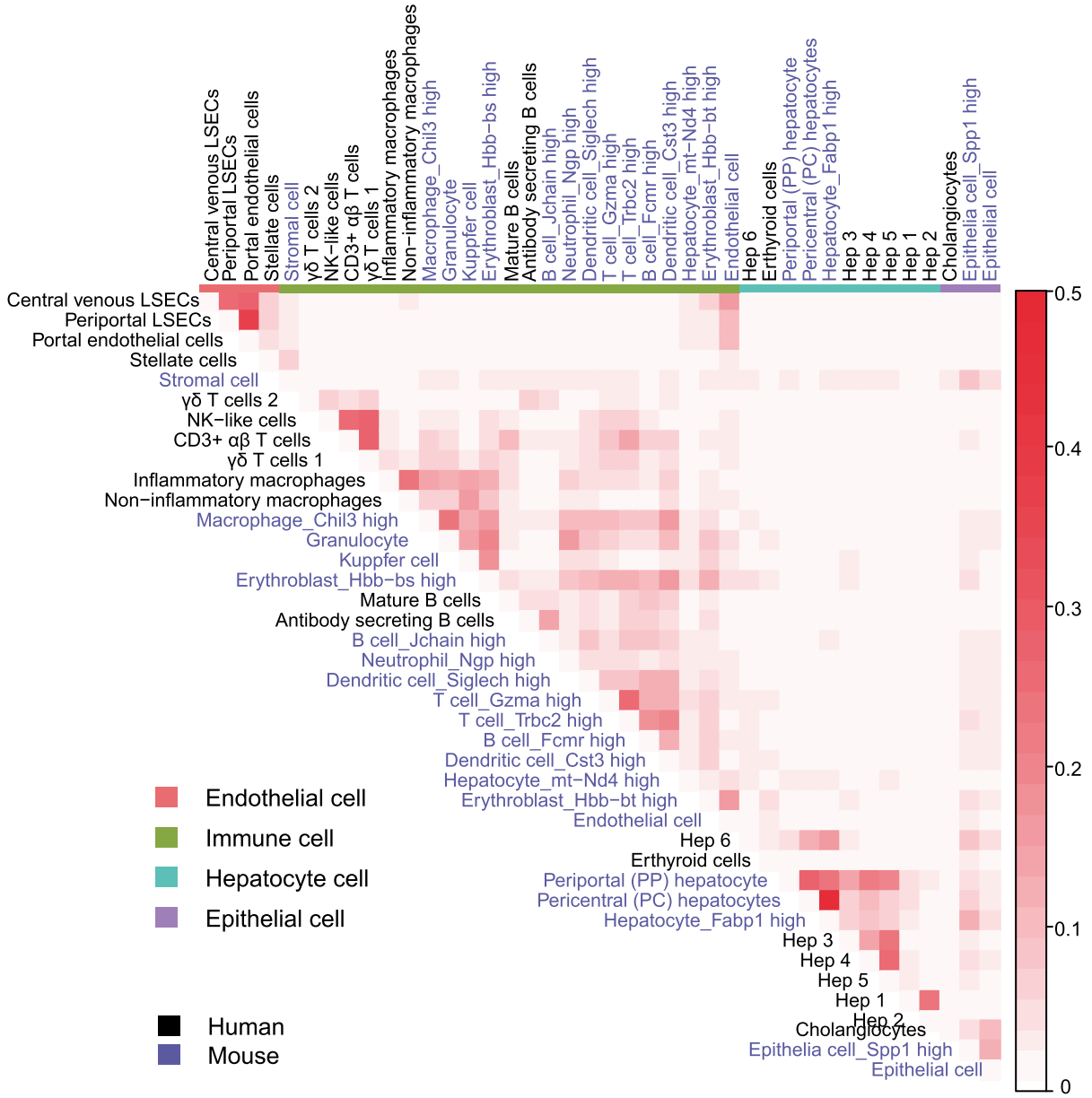

Figure 2. Jaccard Index of overlap among the upregulated genes in different cell types in human and mouse liver. Hierarchical clustering is used for clustering and ordering the different cell types. The colors of the text indicate the origin of the liver cell types, and the colors under the text indicate the group categories of the liver cell types.

In conclusion, we found that the PP hepatocyte cell types play an important role in blood lipid levels by integrative analyses of GWAS summary statistics and scRNA-seq of liver cell types. Our results provided important information for future cellular studies to investigate the metabolic processes underlying blood lipid levels. Nevertheless, the spatial distribution and functions of the liver cell populations have not been fully explored, such as human Hep 4 and Hep 5, which could limit the interpretations of our results and future experiments. In addition, the power to identify the relevant cell types may be limited by the correlation of gene expression patterns between similar cell types. New scRNA-seq technologies with improved sensitivity and precision to detect lowly expressed genes are required for the identification of more differentially expressed genes, and thus to help differentiate similar cell types in future studies.

Materials and Methods

GWAS summary statistics

The blood lipid level GWAS summary statistics were downloaded from http://csg.sph.umich.edu/willer/public/lipids2010 (31). Briefly, the lipid GWAS was based on >100 000 individuals of European ancestry, and identified 95 loci at the genome-wide significant level ($P < 5E\text{-}8$). Given that the liver is highly infiltrated with immune cells, we also downloaded and analyzed GWAS summary statistics of five autoimmune diseases for comparison, including inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, rheumatoid arthritis and type 1 diabetes. All the GWAS results were derived from European samples (see Supplementary Material, Table S1 for details).

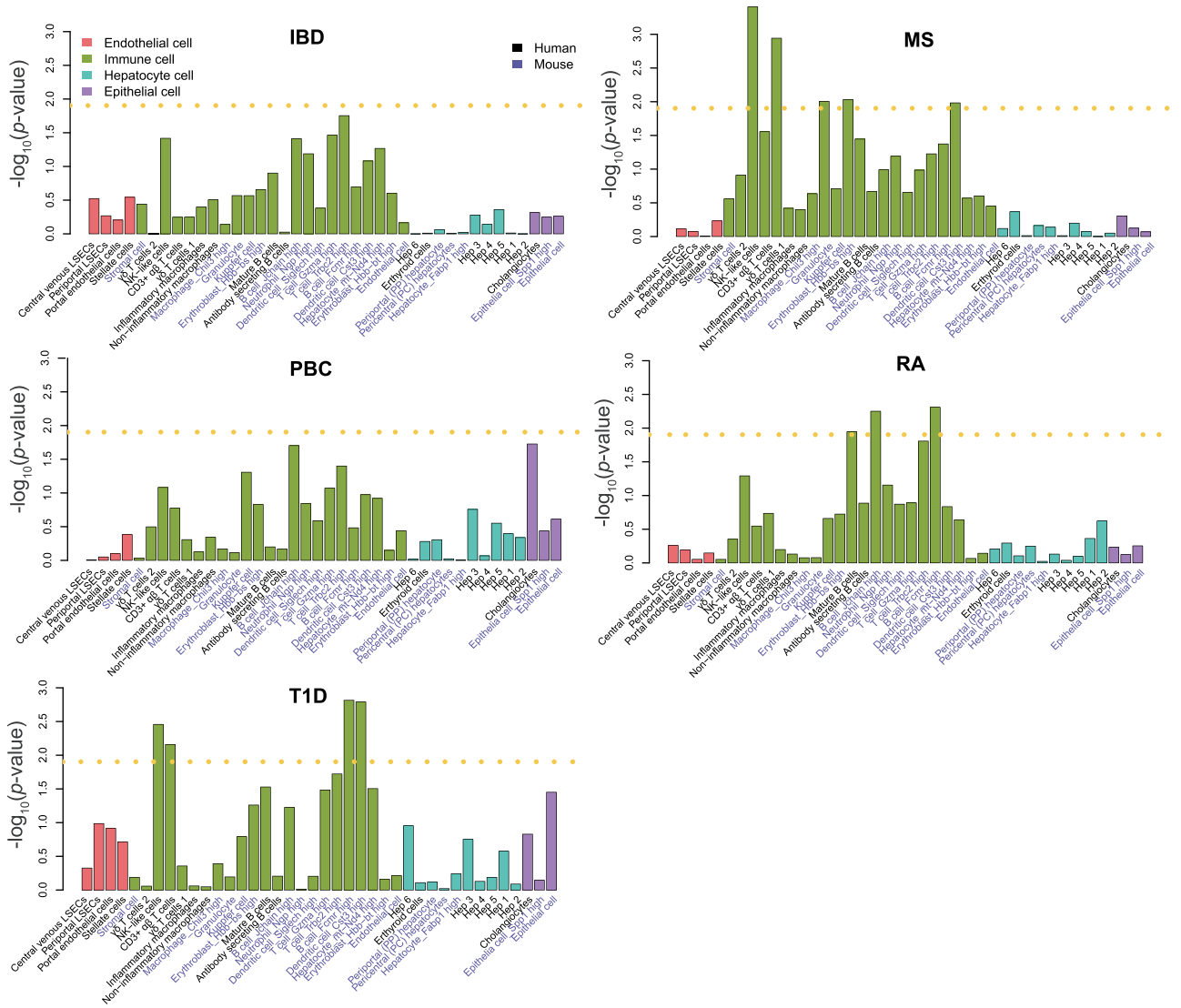

Figure 3. The enrichment of genetic signals of autoimmune diseases for human and mouse liver cell-type specifically expressed gene sets (CT-SEGS). The enrichment results of inflammatory bowel disease (**A**), multiple sclerosis (**B**), primary biliary cirrhosis (**C**), rheumatoid arthritis (**D**) and type 1 diabetes (C) for the human and mouse liver CT-SEGS. The gold dash lines are the Bonferroni significance thresholds ($P < 0.05/4$) after adjusting four cell type groups in the human and mouse liver. The order of the cell types is based on the hierarchical clustering of the Jaccard Index of overlap among the upregulated genes.

Cell-type specific expression

We constructed the specifically expressed gene set in certain human or mouse liver cell types using public datasets (6,7). Following previous studies (11,13,23), we chose the top upregulated genes in each focal cell type as the cell-type specifically expressed gene sets (CT-SEGS), which were not restricted to genes expressed only in the focal cell type (Fig. 1). Because the power to identify the differentially expressed genes depends on the sample size (i.e. the number of cells in each cell type) (32), the number of identified upregulated genes varied dramatically across different cell types. To avoid possible confounding caused by different cell type population sizes, we chose the same number of top upregulated genes for each cell type. To test the robustness of our results, we also constructed CT-SEGS using a fixed threshold of $P < 0.05$ in the differential expression analyses (described in this section), resulting in different number of upregulated genes in the CT-SEGS of different cell types (ranging from 50 to 1723; see Supplementary Note).

The human liver scRNA-seq was conducted on the 10X Genomics platform. Five cell type groups consisting of 20 cell types were identified among 8444 cells (6). We downloaded the human liver scRNA-seq counts (GSE115469), which were normalized using the default settings of the scran R package (33). To construct the specifically expressed gene sets, we performed 20 differential expression (DE) analyses for 17 900 genes using the likelihood-ratio test (34) implemented in Seurat 3.0 (35,36), each testing one cell type against the other 19 cell types. Similar to previous studies defining genes specifically expressed in a tissue (11,13,23), we selected the top 186 upregulated genes (fold change in log2 scale (logFC) > 0, expressed in > 10% of cells from the focal cell type) as the human CT-SEGS (Supplementary Material, Table S2) for each cell type. The

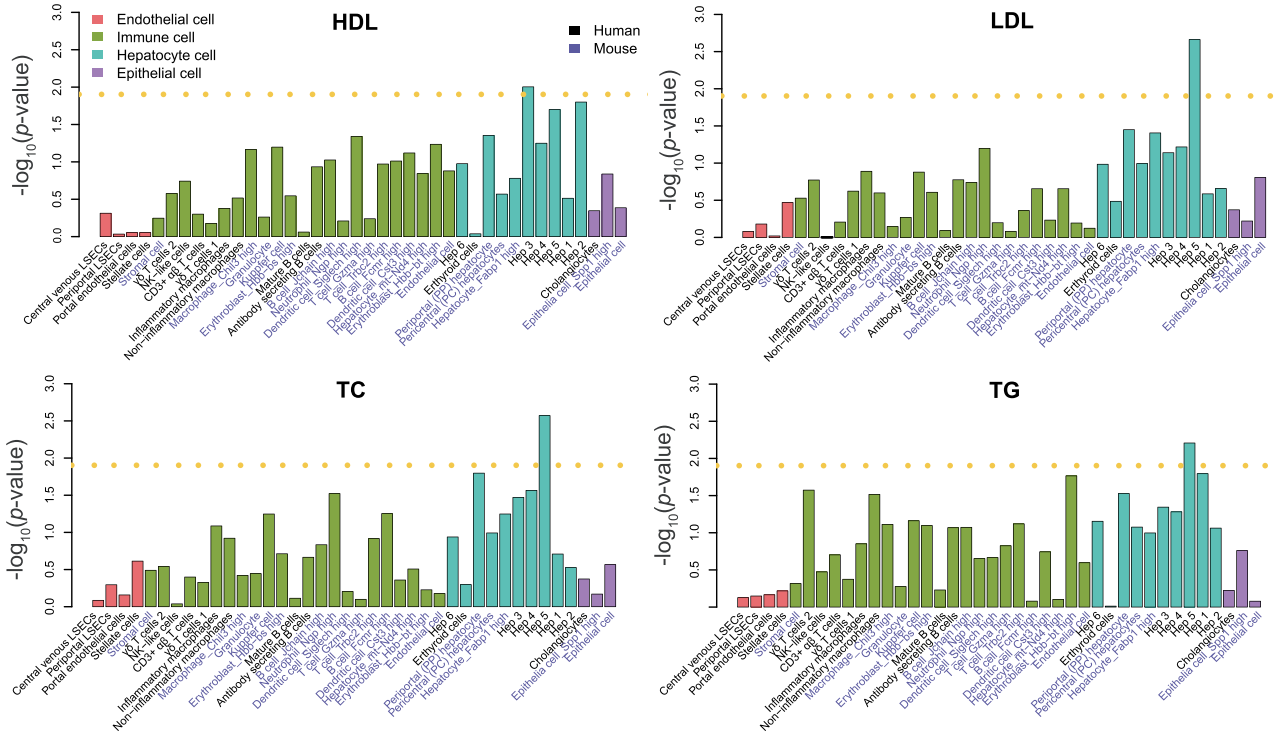

Figure 4. The enrichment of genetic signals of lipid traits for human and mouse liver cell-type specifically expressed gene sets (CT-SEGS). The enrichment results of HDL (**A**), LDL (**B**), TC (**C**) and TG (**D**) for the human and mouse liver CT-SEGS. The gold dash lines are the Bonferroni significance thresholds ($P < 0.05/4$) after adjusting four cell type groups in the human and mouse liver. The order of the cell types is based on the hierarchical clustering of the Jaccard Index of overlap among the upregulated genes.

number of 186 was determined based on the minimum number of upregulated genes across the 20 DE analyses of different human liver cell types. Among these genes, 88.0% had logFC > 0.1 and all of them had adjusted *P* value (false discovery rate [FDR]) < 0.05.

The mouse liver scRNA-seq was conducted on the Microwell-seq platform. Four cell type groups consisting of 20 cell types were identified among 4685 cells (7). We downloaded the raw sequencing counts data (https://figshare.com/articles/MCA_DGE_Data/5435866), and used an expert-curated human–mouse homolog list (http://www.informatics.jax.org/homology.shtml) to map mouse genes to their human homologs. We kept the genes with high mapping confidence (1:1 mapping). We normalized the counts data for the mappable genes using the default settings of the scran R package (33). We conducted DE analyses for the 11 942 genes as we did for the human liver scRNA-seq data. For each cell type, we selected the top 274 upregulated genes (the minimum number of upregulated genes across different mouse liver cell types) as the mouse liver CT-SEGS. Among these genes, 79.7% had logFC < 0.1 and 75.7% had FDR adjusted *P* value < 0.05 (Supplementary Material, Table S2).

Partitioning of trait heritability to CT-SEGS

To identify trait-relevant cell types, we used the stratified LDSC method, which tested if GWAS signals of a trait/disease were enriched in given gene sets (10,11) (Fig. 1). In our case, the gene sets, specifically the CT-SEGS, reflect the unique function of individual cell types. Additionally, we expanded each gene in the CT-SEGS by 100 kb upstream and downstream to represent the functional genomic regions of each liver cell type. We created a binary annotation for each cell type, indicating whether a single-nucleotide polymorphism (SNP) resides in the CT-SEGS of the cell type. LDSC estimates functional enrichment of the GWAS signals using the statistical the following model:

$$E\left(\chi_i^2\right) = 1 + N\alpha + N\sum_k \tau_k l\left(i,k\right),$$

where χ_i^2 is the GWAS summary statistic for SNP i, N is the GWAS sample size, α is a constant that reflects population structure and other sources of confounding, $l(i,k)$ is the linkage disequilibrium (LD) score of SNP i in annotation k, and τ_k is the regression coefficient of annotation k. We estimated LD scores using the European subset from the 1000 Genome Project (37), which includes 9 997 231 biallelic autosomal SNPs (minor allele counts, MAC > 5), with a default 1-centiMorgan window size. The regression coefficient τ_k quantifies the importance of annotation k accounting for all other annotations in the model. For cell-type-specific analysis, the *P*-value was computed to test whether τ_k was significantly greater than 0 (10,11). As previous studies (10,11), we focused on autosomal SNPs and excluded the major histocompatibility complex (MHC) region from all analyses. In each analysis, we jointly fit the following annotations: (i) the CT-SEGS annotations created for liver cell types; (ii) a common annotation indicating genic regions of all genes in the liver scRNA-seq dataset; and (iii) baseline annotations of 52 functional categories that were not specific to any cell type (e.g. coding, 3′ UTR, 5′ UTR, promoter and intron annotations) (10,11). The inclusion of annotations common to all cell types (ii and iii) can help remove potential confounding factors and enhance the cell type specific signals.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgments

We thank all the GWAS consortium studies for making the summary data publicly available and are grateful of all the investigators and participants contributed to those studies.

Conflict of Interest statement. The authors have declared no competing interests.

Funding

The National Natural Science Foundation of China (NSFC, 81973148, 82003561).

Author Contributions

XH conceived the study, performed the data collection and analysis. SC and CW supervised the study. XH drafted the manuscript with inputs from SC and CW. KW, DC, ZD and WY participated in data interpretation. All authors reviewed and approved the manuscript.

References

1. Wang, H., Liang, X., Gravot, G., Thorling, C.A., Crawford, D.H.G., Xu, Z.P., Liu, X. and Roberts, M.S. (2017) Visualizing liver anatomy, physiology and pharmacology using multiphoton microscopy. *J. Biophotonics*, **10**, 46–60.
2. Davies, L.C., Jenkins, S.J., Allen, J.E. and Taylor, P.R. (2013) Tissue-resident macrophages. *Nat. Immunol.*, **14**, 986–995.
3. Ginhoux, F. and Guilliams, M. (2016) Tissue-resident macrophage ontogeny and homeostasis. *Immunity*, **44**, 439–449.
4. Robinson, M.W., Harmon, C. and O'Farrelly, C. (2016) Liver immunology and its role in inflammation and homeostasis. *Cell. Mol. Immunol.*, **13**, 267–276.
5. Wu, J. and Zern, M.A. (2000) Hepatic stellate cells: a target for the treatment of liver fibrosis. *J. Gastroenterol.*, **35**, 665–672.
6. MacParland, S.A., Liu, J.C., Ma, X.-Z., Innes, B.T., Bartczak, A.M., Gage, B.K., Manuel, J., Khuu, N., Echeverri, J., Linares, I. *et al.* (2018) Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.*, **9**, 4383.
7. Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F. *et al.* (2018) Mapping the mouse cell atlas by microwell-Seq. *Cell*, **173**, 1307.
8. Carambia, A. and Herkel, J. (2018) Dietary and metabolic modulators of hepatic immunity. *Semin. Immunopathol.*, **40**, 175–188.
9. Backenroth, D., He, Z., Kiryluk, K., Boeva, V., Pethukova, L., Khurana, E., Christiano, A., Buxbaum, J.D. and Ionita-Laza, I. (2018) FUN-LDA: a latent Dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation: methods and applications. *Am. J. Hum. Genet.*, **102**, 920–942.
10. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C. and Farh, K. (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.
11. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N. *et al.* (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.*, **50**, 621–629.
12. Hao, X., Zeng, P., Zhang, S. and Zhou, X. (2018) Identifying and exploiting trait-relevant tissues with multiple functional annotations in genome-wide association studies. *PLoS Genet.*, **14**, e1007186.
13. Zhu, X. and Stephens, M. (2018) Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat. Commun.*, **9**, 4361.
14. Lu, Q., Powles, R.L., Abdallah, S., Ou, D., Wang, Q., Hu, Y., Lu, Y., Liu, W., Li, B., Mukherjee, S. *et al.* (2017) Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLoS Genet.*, **13**, e1006933.
15. Calderon, D., Bhaskar, A., Knowles, D.A., Golan, D., Raj, T., Fu, A.Q. and Pritchard, J.K. (2017) Inferring relevant cell types for complex traits by using single-cell gene expression. *Am. J. Hum. Genet.*, **101**, 686–699.
16. Andrews, T.S. and Hemberg, M. (2018) Identifying cell populations with scRNASeq. *Mol. Asp. Med.*, **59**, 114–122.
17. Stuart, T. and Satija, R. (2019) Integrative single-cell analysis. *Nat. Rev. Genet.*, **20**, 257–272.
18. Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E. *et al.* (2017) Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature*, **542**, 352–356.
19. 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. *et al.* (2018) Single-cell transcriptomics of 20 mouse organs creates a tabula Muris. *Nature*, **562**, 367–372.
20. Segal, J.M., Kent, D., Wesche, D.J., Ng, S.S., Serra, M., Oulès, B., Kar, G., Emerton, G., Blackford, S.J.I., Darmanis, S. *et al.* (2019) Single cell analysis of human foetal liver captures the transcriptional profile of hepatobiliary hybrid progenitors. *Nat. Commun.*, **10**, 3350.
21. Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J.S., Pessaux, P., Baumert, T.F. and Grün, D. (2019) A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*, **572**, 199–204.
22. Shang, L., Smith, J.A. and Zhou, X. (2020) Leveraging gene co-expression patterns to infer trait-relevant tissues in genome-wide association studies. *PLOS Genetics*, **16**, e1008734.
23. Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B. *et al.* (2018) Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.*, **50**, 825–833.
24. Skene, N.G. and Grant, S.G.N. (2016) Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.*, **10**. doi: 10.3389/fnins.2016.00016.
25. Lane, J.M., Jones, S.E., Dashti, H.S., Wood, A.R., Aragam, K.G., van Hees, V.T., Strand, L.B., Winsvold, B.S., Wang, H., Bowden, J. *et al.* (2019) Biological and clinical insights from genetics of insomnia symptoms. *Nat. Genet.*, **51**, 387–393.

26. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A. *et al*. (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet*., **51**, 414–430.
27. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L. *et al*. (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet*., **51**, 404–413.
28. Pickrell, J.K. (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet*., **94**, 559–573.
29. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. and Ziller, M.J. (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
30. Donovan, M.K.R., D'Antonio-Chronowska, A., D'Antonio, M. and Frazer, K.A. (2020) Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. *Nature Communications*, **11**, 955.
31. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I. and Willer, C.J. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.
32. Hart, S.N., Therneau, T.M., Zhang, Y., Poland, G.A. and Kocher, J.-P. (2013) Calculating sample size estimates for RNA sequencing data. *J. Comput. Biol*., **20**, 970–978.
33. Lun, A.T.L., McCarthy, D.J. and Marioni, J.C. (2016) A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res*, **5**, 2122–2122.
34. McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M. and Gottardo, R. (2013) Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics (Oxford, England)*, **29**, 461–467.
35. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol*., **36**, 411.
36. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019) Comprehensive integration of single-cell data. *Cell*, **177**, 1888–1902.e1821.
37. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.