Unveiling the pathogenesis of non-alcoholic fatty liver disease by decoding biomarkers through integrated single-cell and single-nucleus profiles

Wenfeng Ma [1,2,3,4], Xin Zhong [1,4], Benqiang Cai [1,4], Mumin Shao [6,7], Xuewen Yu [6,7], Minling Lv [1,4], Shaomin Xu [1,4], Bolin Zhan [1,4], Qun Li [1,4], Mengqing Ma [1,4], Mikkel Brejnholt Kjær [5,8], Jinrong Huang [2]*, Yonglun Luo [2,3]*, Henning Grønbæk [5]*, Lin Lin [2,3]*

1 Department of Liver Disease, Shenzhen Traditional Chinese Medicine Hospital, Shenzhen, Guangdong 518033, China.
2 Department of Biomedicine, Aarhus University, Aarhus, Denmark.
3 Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark.
4 Department of Liver Disease, The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine, Shenzhen, 518033, China.
5 Department of Hepatology and Gastroenterology, Aarhus University Hospital, Aarhus, Denmark.
6 Department of Pathology, Shenzhen Traditional Chinese Medicine Hospital, Shenzhen, Guangdong 518033, China.
7 Department of Pathology, The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine, Shenzhen, 518033, China.
8 Department of Clinical Medicine, Aarhus University, Aarhus, Denmark.
* = corresponding author

Highlights

- Integrated single-cell and single-nucleus profiles from 82 liver and PBMC specimens comprising NAFLD patients and healthy controls with increasing

severity were utilized to unveil the NAFLD pathogenesis through decoding candidate biomarkers of NAFLD.

- In cell-level observations, we decoded 16 up-regulated and 22 down-regulated secreting genes previously identified as associated with increasing NAFLD severity in the liver RNA-seq and plasma proteomics data.
- QSOX1, enriched in fibroblasts, and IL1RAP, enriched in hepatocytes, have been further validated and interpreted in integrated single-cell and single-nucleus profiles for their potential to predict NAFLD severity.
- The analysis of intercellular crosstalk, focusing on secreted signaling from our previously identified candidate biomarkers sourced from secreted genes, highlighted the IL1B-(IL1R1+IL1RAP) pathway between liver monocytes and hepatocytes/cholangiocytes. This suggests that this pathway might be a potential reason for the observed downregulation of IL1RAP in NAFLD liver.

Abstract

Background and Aims: Non-alcoholic fatty liver disease (NAFLD) is a heterogenous liver disease encompassing pathological changes ranging from simple steatosis, inflammation and fibrosis to cirrhosis. To further unravel NAFLD pathogenesis, we aimed to decode the candidate NAFLD biomarkers associated with NAFLD severity using publicly available single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA sequencing (snRNA-seq) data.

Methods: Seurat v5 and anchor-based reciprocal principal components analysis (RPCA) integration were performed to integrate and analyze the scRNA-seq and snRNA-seq data of 82 liver and Peripheral Blood Mononuclear Cell (PBMC) specimens from NAFLD patients and healthy controls to decode the candidate NAFLD biomarkers generated previously. Using the 'CellChat' R package, we analyzed ligand-receptor interactions of our candidate biomarkers from secreted genes to understand their signaling crosstalk and implications in NAFLD's biological processes.

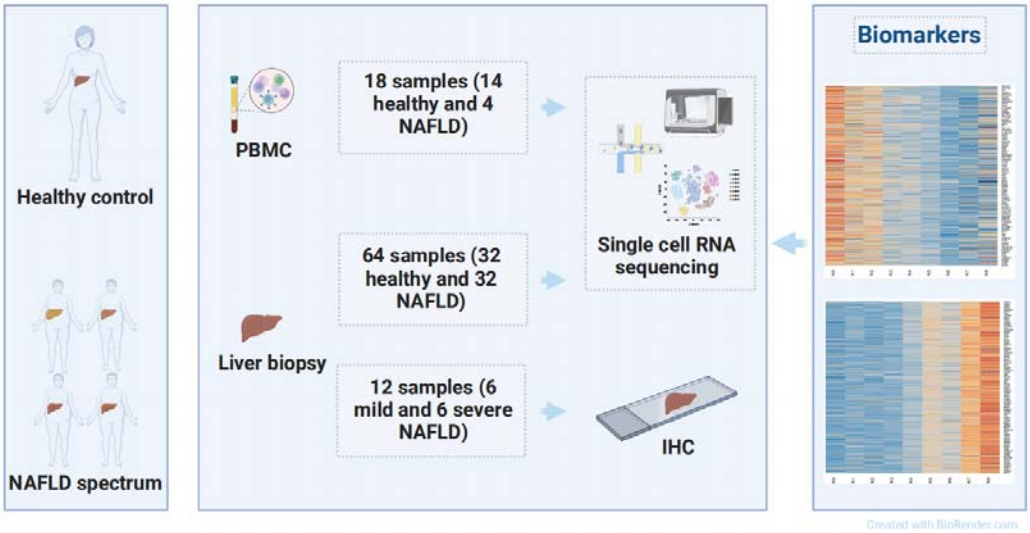
Results: We generated a database (https://dreamapp.biomed.au.dk/NAFLD-scRNA-seq/) to present the NAFLD pathogenesis by analyzing integrated scRNA-seq and snRNA-seq data. Through cell-level decoding, we discovered the expression distribution of the candidate biomarkers associated with NAFLD severity. The analysis of ligand-receptor pairs in NAFLD liver and PBMC data suggests that the IL1B-(IL1R1+IL1RAP) interaction between liver monocytes and hepatocytes/cholangiocytes may explain the correlation between NAFLD severity and IL1RAP down-regulation.

Conclusions: We confirmed a strong correlation between liver QSOX1/IL1RAP concentrations and NAFLD severity at the cellular level. Additionally, our analysis of comprehensive data unveiled new aspects of NAFLD pathogenesis and intercellular communication through the use of scRNA and snRNA sequencing data. (ChiCTR2300073940).

Lay Summary

We integrated single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA sequencing (snRNA-seq) data to unravel non-alcoholic fatty liver disease (NAFLD) pathogenesis. We decoded candidate biomarkers associated with NAFLD progression, which were previously screened from RNA sequencing (RNA-seq) data of 625 liver samples with a novel gene clustering method. A new version of the R package ''Seurat v5' and anchor-based reciprocal principal components analysis (RPCA) integration were performed to process and integrate scRNA-seq and snRNA-seq data of 82 liver and Peripheral Blood Mononuclear Cell (PBMC) specimens from NAFLD patients and healthy controls. The research delved deeper into the cellular expression patterns of the candidate biomarkers and examined the intercellular communication of their secreted signaling.

Graphical abstract



Keywords

Non-alcoholic fatty liver disease, pathogenesis, single-cell and single-nucleus RNA sequencing data integration, quiescin sulfhydryl oxidase 1, interleukin-1 receptor accessory protein

Introduction

Non-alcoholic fatty liver disease (NAFLD) is a multifaceted disease representing a spectrum from Non-alcoholic Fatty Liver (NAFL) of simple steatosis, non-alcoholic steatohepatitis (NASH) to NASH-induced fibrosis and cirrhosis (1, 2). As the global prevalence of NAFLD reached 38.0%, the upward trend in NAFLD and NASH-related liver cancer balanced the declining tendency of virally-caused hepatocellular carcinoma (3-5). Despite the noteworthy advancements in NAFLD research, the lack of FDA-approved treatments underscores the need for a systematic study to understand its pathogenesis (6, 7).

The advent of transcriptomics, specifically single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA sequencing (snRNA-seq) has triggered a paradigm shift in our comprehension of biological processes, cellular diversity, and the intricacies of physiological and disease states (8). These groundbreaking technologies allow us to examine gene expression at the cell level, significantly enhancing our understanding of various aspects of life sciences (9, 10). In 2019, Xiong et al. applied scRNA-seq to reveal the cellular heterogeneity of NAFLD, identifying key cell types and crosstalk during disease progression (11). Tamburini et al. focused on liver lymphatic endothelial cells (LECs) and observed the activation of the IL13 pathway in LECs of NASH patients (12). In 2020, Ramachandran et al. utilizing single-cell transcriptomics, resolved human liver tissue and blood samples of patients with cirrhosis and healthy controls, identifying new pathogenic subpopulations and providing the markers of cell subtypes (13). In 2021, Deczkowska and colleagues discovered an abundance and activation of type 1 conventional dendritic cells (cDC1) in both NAFLD and NASH conditions with scRNA-seq. They further highlighted XCR1+ cDC1 as a significant contributor to liver pathology in NASH (14). Recently, Wang et al. interpreted NASH-associated hepatic stellate cells (HSCs) with snRNA-seq and uncovered an autocrine HSC signaling circuit that emerged in advanced fibrosis (15). These studies underscore the immense potential of transcriptomics application in NAFLD research

120 for finding molecular signatures and pathways that could guide the development of
121 diagnosis and therapeutics for NAFLD.
122
123 Recently, we demonstrated that Quiescin sulfhydryl oxidase 1 (QSOX1) and
124 Interleukin-1 receptor accessory protein (IL1RAP) hold potential as indicators of liver
125 disease severity in NAFLD patients (16). QSOX1, a flavin-linked enzyme that facilitates
126 disulfide bond formation within the secretory structures of mammalian cells, is
127 expelled into the extracellular matrix (ECM) and blood as an independent protein
128 through the standard secretory pathway, encompassing the Golgi apparatus,
129 secretory vesicles, and the plasma membrane. Furthermore, it is notably amplified
130 and released from fibroblasts that are actively synthesizing ECM precursors (15-17).
131 On the other hand, IL1RAP, a member of the IL-1 receptor family, is ubiquitously
132 present but shows high levels in the liver, placenta, and white blood cells, according
133 to databases such as NCBI. This protein plays a critical role in inflammation and is
134 implicated in the initiation and progression of cancer. Studies suggested that both
135 QSOX1 and IL1RAP hold promise as potential biomarkers of NAFLD or obesity (17,
136 18).
137
138 In 2022, Hao's team unveiled Seurat v5, featuring a 'bridge integration' technique
139 designed to merge single-cell datasets from diverse molecular domains and various
140 measurement modalities. Employing concepts from 'dictionary learning' within the
141 field of representation learning, this method is highly scalable and ideal for handling
142 expansive datasets of millions of cells. It has wide applicability across different
143 technologies and molecular modalities, making it valuable for large research
144 consortia with existing single-cell RNA-seq databases. Additionally, the approach
145 shows potential in the identification and comprehensive study of rare cellular
146 populations. Overall, the introduction of bridge integration in Seurat v5 is poised to
147 make a significant impact on the advancement of single-cell research (19). Compared
148 to its predecessor (20), Seurat v5 enhanced its capacity for large-scale data
149 integration and facilitated a more comprehensive analysis of the cellular and nuclear

150 transcriptomic landscapes. With Seurat v5, we aimed to use publicly available
151 scRNA-seq and snRNA-seq data to decode the candidate NAFLD biomarkers we have
152 previously screened for unraveling NAFLD pathogenesis.
153
154 Understanding global cellular interactions necessitates precise representation and
155 analysis of cell-to-cell signaling pathways. "CellChat" serves as a tool, crafted to
156 extract and evaluate intercellular communication networks from sc- and snRNA-seq
157 data (21). It identifies primary signaling pathways for cells, decoding their
158 coordinated actions through network analyses and pattern detection methods. Using
159 manifold learning and quantitative assessments, CellChat classifies signaling routes
160 and differentiates between universal and context-specific pathways across diverse
161 datasets. Consequently, it aids in revealing novel intercellular communications and in
162 building communication blueprints for a range of tissues (22).
163

164 ## Materials and methods

165 **Single-cell and Single-nucleus RNA-seq Data Analysis**

166 1. Data Collection

167 The scRNA-seq and snRNA-seq data from human liver and Peripheral Blood
168 Mononuclear Cell (PBMC) of NAFLD and healthy controls were collected from the
169 NCBI GEO (until June 2023). Only datasets generated with 10x Genomics Single-Cell
170 Reagent Kit were included.
171

172 2. Data Alignment

173 The FASTQ-formatted data were downloaded and converted into feature, barcode
174 and matrix with the human GRCh38 genome using 'cellranger' (cellranger-6.1.2). For
175 snRNA-seq, intron was included for quantification (23). The doublets were removed
176 with the R package 'DoubletFinder' (DoubletFinder-2.0.3) for each sample.
177

178 3. Data Normalization and Clustering

With the R package 'Seurat V5' (Seurat 4.9.9.9042) (https://satijalab.org/seurat/articles/seurat5_integration.html), the datasets were integrated and processed using the anchor-based reciprocal principal components analysis (RPCA) method. After normalization and batch correction, the scRNA-seq and snRNA-seq data underwent unsupervised clustering.

For normalization, the 'LogNormalize' method was utilized with a default scale factor of 10,000, and subsequent feature identification and data scaling were done by default. The scaled data was then subjected to Principal Component Analysis (PCA) with a default setting of computing and storing 50 Principal Components (PCs). To perform integration in low-dimensional space, we selected the anchor-based RPCA integration method, which aimed to co-embed shared scRNA-seq and snRNA-seq library types across batches, yielding a dimensional reduction (integrated.rpca).

For cell clustering, a shared nearest neighbor (SNN) modularity optimization-based algorithm was employed, with the dimensions of reduction set to 1:15 and the resolution parameter set to 0.2 (liver) and 0.1 (PBMC).

4. Cell Type Annotation and Data Analysis

We annotated liver and PBMC clusters with canonical markers by manually curated the information to enhance the accuracy of gene annotation. To gain a more nuanced understanding of transcriptomic variations, we further segmented certain cell types in additional supplementary files. This enabled us to identify subtype-specific differences in both healthy and NAFLD conditions. We then compared the expressions of potential NAFLD biomarkers across multiple clustering methodologies and depicted the results on distinct UMAP plots. For subtype cell clustering, a shared nearest neighbor (SNN) modularity optimization-based algorithm was employed, with the dimensions of reduction set to 1:15 and the resolution parameter set to 0.05 (HSC/MFB).

209 5. Cell Communication Analysis

210 The sc- and snRNAseq data of NAFLD liver and PBMC were merged together and
211 RPCA integrated before normalization. A Seurat object with a new assay of 5,000
212 cells in each cell type in a tissue was subsetted. The R package 'CellChat' (CellChat
213 1.6.1) was utilized to analyze the data. We refined the CellChatDB database to
214 include only interactions involving secreted signaling, as the candidate NAFLD
215 biomarkers originate from genes that are secreted.

216

217 6. Database Construction

218 To visualize gene expression variation during the development of NAFLD severity at
219 the cell level, we generated a NAFLD sc- and snRNA-seq database (NAFLD-DB) using
220 the tool 'ShinyCell'. (https://github.com/SGDDNB/ShinyCell). This database was
221 deployed at https://dreamapp.biomed.au.dk/NAFLD-scRNA-seq/.

222

223 **Validation of QSOX1 and IL1RAP as Biomarkers in NAFLD at the Liver Cell Level**

224 1. Study Design

225 Our objective was to confirm the *QSOX1* and *IL1RAP* gene expression in cells of
226 different severity of NAFLD. To address this question, we performed
227 immunohistochemical staining (IHC) for liver QSOX1/IL1RAP on a group of NAFLD
228 patients from Shenzhen Traditional Chinese Medicine Hospital, China (SZTCMH).
229 Details of this cohort have been provided previously (16).

230

231 2. Human Samples and IHC

232 For IHC, we collected 12 fixed liver tissues from archived histological samples at
233 SZTCMH between 2014 and 2023, and details were in our earlier study (16).

234

235 3. Statistical Analysis

236 The significance for all statistical tests was two-sided, with $P < 0.05$. All data analysis
237 was presented in the plots using R-4.3.1. Statistical analysis in the Violin plots was
238 performed with Wilcoxon rank sum test to assess differences between unpaired

239 groups.

240

241 ## Results

242 **Alignment of ScRNA-seq and snRNA-seq Data**

243 We incorporated twelve datasets (GSE122083, GSE125188, GSE136103, GSE144430,
244 GSE159977, GSE169446, GSE175793, GSE179886, GSE182159, GSE189539,
245 GSE212837 and GSE217235) into our study. The scRNAseq and snRNA-seq data
246 encompass liver and PBMC samples of both healthy controls and patients with
247 NAFLD (**Supplementary Table 1**). The detailed information regarding the data
248 alignment was presented in the **Supplementary Table 2.**

249

250 By employing RPCA integration, we achieved precise comparative analyses across
251 single-cell and single-nucleus sequencing datasets. To address batch effects arising
252 from diverse sequencing technologies and studies involving various experimental
253 batches, donors, and conditions, we conducted integration and normalization, as
254 depicted in **Figure 1A**. After filtering out doublets, the dataset was refined, resulting
255 in a total of 427,787 liver cells (including 214,795 healthy cells and 212,992 NAFLD
256 cells) and 122,238 PBMC cells (including 92,319 healthy cells and 29,919 NAFLD cells)
257 available for subsequent analysis.

258

259 **The Effectiveness of Integration and Normalization on Liver Transcriptome Profiles**

260 PCA demonstrated the effectiveness of the RPCA integration and normalization
261 process in eliminating noticeable batch effects of liver transcriptome profiles
262 compared to the unintegrated UMAP, as depicted in **Figure 1B and 1C**. Moreover,
263 sample separation was not primarily driven by health/NAFLD group or library type, as
264 indicated in **Figure 1D and 1E**. Instead, cell type emerged as the primary factor
265 influencing transcriptome profiles, as illustrated in **Figure 1F**.

266

267 **Unsupervised Clustering and Annotation of Liver Cells and Nuclei Data**

268 Initially, we categorized all 427,787 liver cells into 12 distinct cell types using gene
269 markers. The identified cell types were annotated using signatures of known
270 canonical markers (**Figure 2A, B; Supplementary Table 3**).
271
272 **Tracing the Cellular Origins of Promising Biomarkers for Liver NAFLD Severity**
273 In our previous study (16), we pinpointed key genes that are significantly involved in
274 the advancement of NAFLD. Specifically, we discovered 16 genes, including A2M, C7,
275 COL6A3, COLEC11, ENPP2, FBLN1, FBN1, FCGBP, IGFBP6, LCN2, LUM, MMP2, PAPLN,
276 PTGDS, QSOX1, and VWF, that exhibited elevated secretory activity correlating with
277 the severity of NAFLD. On the other hand, we also identified another set of 22 genes,
278 namely AZGP1, C1RL, C4BPA, C6, C8B, CFHR3, CNDP1, F2, GC, HP, HPR, IL1RAP, ITIH1,
279 ITIH2, ITIH4, KLKB1, PON3, SERPINA10, SERPINC1, SERPING1, SMPDL3A, and TTR,
280 that showed reduced secretory activity, also in relation to the severity of the
281 condition. These particular genes were not only implicated through RNA-seq data but
282 were also corroborated by proteomics datasets. The correlation was such that the
283 expression patterns of these genes seemed to escalate or diminish in tandem with
284 increasing NAFLD severity. We investigated how these genes contribute to NAFLD
285 pathogenesis, and why their particular expression patterns are so closely associated
286 with the progression of NAFLD.
287
288 Delving into the intricate cellular milieu, the expression profiles of these potential
289 biomarkers were meticulously examined across diverse liver cell categories. It was
290 observed that the HSC/MFB cellular contingent distinctly exhibited pronounced
291 expression levels of C7, COL6A3, FBN1, and QSOX1 compared to other cellular
292 phenotypes. Intriguingly, these molecular markers were conspicuously prevalent in
293 NAFLD-afflicted cohorts over their healthy counterparts (**Figure 2C**). In light of the
294 overwhelming majority of hepatocytes in hepatic architecture, significant expression
295 of IL1RAP, ITIH2, TTR, and SERPINA10 was ascertained in these cells, spanning all
296 identified subtypes. This expression spectrum was notably skewed towards the

297 healthy clusters versus NAFLD groups, intimating a palpable downregulation at the
298 holistic hepatic expression matrix (**Figure 2D**).
299
300 Cutting-edge developments in medical science have clarified the primary source of
301 myofibroblasts, identifying hepatic stellate cells (HSCs) as their main origin (24, 25).
302 Functioning as lipid storage entities, HSCs share considerable similarity with
303 adipocytes, or fat cells, particularly when the liver is functioning optimally (26). This
304 newfound insight broadens our understanding of liver diseases, including liver
305 fibrosis and cirrhosis, where myofibroblasts play a significant role. Interestingly, the
306 elevated levels of QSOX1, which is enriched in the HSC/MFB, correlate with both the
307 NAS score and fibrosis score, offering a potential biomarker for NAFLD severity.
308
309 A more detailed exploration of the HSC/MFB clustering, and annotation using
310 canonical markers has revealed two distinct subsets: hepatic stellate cells and
311 myofibroblasts (**Figure 2E, F; Supplementary Table 3**). Within this context, QSOX1
312 manifested heightened expression within the NAFLD subset of HSC (**Figure 2G**). In
313 contrast, IL1RAP and CFHR3, bearing the torch for the loftiest expression within
314 hepatocytes amongst all subtypes, evinced diminished expression within the NAFLD
315 demographic compared to its healthier cohort (illustrated in **Figure 2H**).
316
317 In a supplementary validation, liver biopsies sourced from patients with varying
318 NAFLD severity were subjected to IHC in our antecedent study (16). Conclusive
319 findings reinforced that QSOX1 resonates with heightened expression specifically in
320 HSC/MFB cells of severe NAFLD patients (**Figure 2I**), while IL1RAP's expression was
321 attenuated in hepatocytes of these patients as seen in **Figure 2J** (16).
322
323 **The Effectiveness of Integration and Normalization on PBMC Transcriptome Profiles**
324 To explore the expression of biomarkers in PBMC between NAFLD patients and
325 healthy controls, we tracked their expression variations using RPCA integrated
326 scRNA-seq data. The results from PCA showed that normalization successfully

327 mitigated discernible batch effects compared to the unintegrated data, as shown in
328 **Figure 3A and B**. The primary factors driving sample differentiation were not the
329 health and NAFLD group, as depicted in **Figure 3C**. Instead, the cell type surfaced as
330 the primary determinant impacting transcriptome profiles, as illustrated in **Figure 3D**.
331

332 **Unsupervised Clustering Identifies Clusters of PBMC Cells**

333 We first clustered all the 122,238 PBMC cells into 6 cell types, and the identified cell
334 types were annotated using signatures of known canonical markers (**Figure 3E, F;**
335 **Supplementary Table 3**).
336

337 **Tracing the Cellular Origins of Promising Biomarkers for NAFLD severity in PBMC**

338 We identified 16 genes that are up-regulated (A2M, C7, COL6A3, COLEC11, ENPP2,
339 FBLN1, FBN1, FCGBP, IGFBP6, LCN2, LUM, MMP2, PAPLN, PTGDS, QSOX1, VWF) and
340 22 that are down-regulated (AZGP1, C1RL, C4BPA, C6, C8B, CFHR3, CNDP1, F2, GC, HP,
341 HPR, IL1RAP, ITIH1, ITIH2, ITIH4, KLKB1, PON3, SERPINA10, SERPINC1, SERPING1,
342 SMPDL3A, TTR) with increasing severity of NAFLD, as evidenced by both RNA-seq and
343 proteomics data. In PBMC, QSOX1 and IL1RAP have the highest expression levels in
344 Natural killer cells (NK cells) across all subtypes. Interestingly, healthy individuals
345 showed higher expression of these genes than those with NAFLD (**Figure 3G, H**).
346 Nonetheless, the baseline expression levels of these genes in PBMC were too faint to
347 support a significant comparison. The low expression within PBMC may not affect
348 their detectability and proportional concentrations in plasma. Even when a gene
349 exhibits varied expression in a distinct cell type or tissue, it's not a given that there
350 will be discernible differences in broader systems like plasma, especially when the
351 initial expression is inherently low (**Figure 3I, J**).
352

353 **The intercellular crosstalk in the liver and PBMC of NAFLD patients and healthy**
354 **controls revealed by integrated sc- and snRNAseq data analysis**

355 Since our candidate biomarkers were derived from secreted genes, it was essential to
356 delve deeper into the communication mechanisms of these genes. We utilized the

357 'CellChat' R package to analyze the ligand-receptor interactions, enabling us to better
358 understand the crosstalk of secreted signaling and their potential implications in the
359 biological processes of NAFLD.
360
361 The integrated RPCA analysis combined a total of 427,787 liver cells and 122,238
362 peripheral blood mononuclear cells (PBMCs). The data was then partitioned into
363 subsets comprising 5,000 cells for each cell type within the respective tissues.
364 Subsequent analysis was carried out using the 'CellChat' R package (**Figure 4A**). The
365 examination of ligand-receptor pairs of secreted signaling, which constitutes 61.8%
366 of all categories as depicted in **Figure 4B**, within the transcriptomic data from the
367 liver and PBMCs in the NAFLD group indicates a strong correlation. Specifically, the
368 interaction between IL1B-(IL1R1+IL1RAP) in liver monocytes and
369 hepatocyte/cholangiocyte may provide a convincing rationale for the correlation
370 between NAFLD severity and the down-regulated expression of IL1RAP in the liver, as
371 highlighted in **Figure 4C and F**. In our earlier research (16), we pinpointed potential
372 diagnostic markers from candidate genes by juxtaposing both up-regulated and
373 down-regulated gene clusters against the secretome database from the Human
374 Protein Atlas. Through this cross analysis, we discovered 349 genes that encode
375 secreted proteins: 249 of these genes were up-regulated, while 100 were
376 down-regulated. Of the totally 349 secreted genes, both FGFR2 and JAM3 exhibited
377 up-regulation, while ERBB3 and SDC4 were notably down-regulated. These four
378 genes are cataloged within the ligand-receptor pairs in the CellChat database, as
379 referenced in **Supplementary Table 4**. Yet, when examining the crosstalk analysis
380 derived from the sc- and snRNAseq data of NAFLD patients, only the down-regulated
381 SDC4 (ANGPTL4-SDC4) and ERBB3 (NRG1-ERBB3) emerged as significant receptors in
382 these pairs (see **Supplementary Tables 4 and 5**). The ANGPTL4-SDC4 pairing occurs
383 between liver hepatocytes and hepatocyte/cholangiocyte, as depicted in **Figures 4D**
384 **and 4G**. Similarly, the NRG1-ERBB3 interaction is evidenced between liver
385 hepatocytes and cholangiocytes, highlighted in **Figures 4E and 4G**. An accompanying
386 violin plot reveals expression patterns of the three crucial down-regulated secreted

387 genes from the crosstalk analysis. Intriguingly, Most of them are predominantly
388 found in hepatocytes, as demonstrated in **Figure 4H**. It may intricate that the
389 significant down regulation of liver gene expression in bulk RNA sequencing may
390 mostly caused by the gene variation in hepatocyts since it is the biggest population in
391 the liver. And the down-regulation may caused by the ligands acted on the receptors
392 in the hepatocytes when the NAFLD severity developed.
393

394 ## Discussion

395 The inherent complexities of NAFLD pathogenesis consistently propel scientists to
396 extend the boundaries of traditional diagnostic and investigative methodologies,
397 thus underscoring the escalating importance of genomics technology within this
398 domain. Our research RPCA integrated publicly available scRNA-seq and snRNA-seq
399 datasets to identify potential biomarkers, thereby contributing to the enhanced
400 understanding of the intricate pathogenic processes driving NAFLD (27). Using this
401 approach, we corroborated the significant correlation between the expression ratio
402 of QSOX1/IL1RAP and the severity of NAFLD at the cellular level (16). Furthermore,
403 through an analysis of the integrated large-scale scRNA and snRNA sequencing data
404 to interpret potential biomarkers, we shed light on previously undiscovered facets of
405 NAFLD pathogenesis and intercellular crosstalk.
406

407 QSOX1 and its less common variant, QSOX2, are unique among sulfhydryl oxidases
408 because they are specifically found in the Golgi apparatus of animal cells (28).
409 Studies found elevated expression of QSOX1 in quiescent fibroblasts, and it is also
410 released into the bloodstream (29, 30). QSOX1 plays a crucial role in the proper
411 construction of the ECM, particularly elements similar to the basement membrane
412 (BM) in cell culture. The BM, an ECM layer interfacing bodily cavities or blood vessels
413 with underlying stromal fibroblasts, is largely composed of elements generated by
414 fibroblasts (31). The exact reason behind QSOX1's secretion—whether due to
415 overwhelming the Golgi's retention mechanism or some specialized secretion

416 pathway—remains unclear. Our findings highlight an increased expression of QSOX1
417 in fibroblasts within NAFLD-affected liver tissue, suggesting that Golgi function in
418 fibroblasts could be a relevant focus for NAFLD research (32).
419
420 Fibrogenesis is a key driver in NAFLD progression and is chiefly managed by
421 specialized cells called myofibroblasts (33, 34). While these cells are not present in a
422 healthy liver, they can quickly develop following liver damage. Emerging research has
423 established that these myofibroblasts predominantly originate from hepatic stellate
424 cells (HSCs) (24, 25) , which act mainly as lipid storage cells and bear resemblance to
425 fat cells, or adipocytes, under healthy conditions (26).
426
427 In our previous article (16), the GO analysis revealed significant enrichment of the
428 up-regulated genes involved in the fibrosis-related process, such as ECM organization.
429 Similar to QSOX1 in terms of expression pattern and location, COL6A3 is associated
430 with liver fibrosis and also plays a role in the ECM organization (35). Increased
431 production of ECM moieties by diverse liver cell types is tightly linked with hepatic
432 fibrosis. Research has demonstrated that adipocytes are affected by the mechanics of
433 their surrounding ECM, such as stiffness and architecture, independent of initial
434 ligand density. In response to a stiffened, modified ECM structure, adipocytes
435 enhance the expression of fibrotic genes and increase ECM deposition. This reaction
436 eventually results in a dysfunctional, fibrotic state (36, 37). This represents a
437 significant research trajectory in the ongoing investigation of NAFLD pathophysiology.
438
439 IL1RAP is a crucial co-receptor expressed in numerous tissues and cell types,
440 emphasizing its pivotal role in immune responses, cancers, and infectious diseases
441 (38). The IL1RAP serves as a subunit to the type I IL-1 receptor (IL-1RI), a specific
442 plasma membrane receptor, which facilitates the transmission of numerous
443 biological effects of the pro-inflammatory cytokine IL-1 on an array of target cells (39,
444 40). Research indicates that one of the isoforms of IL1RAP, known as the soluble

isoform of the IL-1 receptor accessory protein (sIL1RAP), has been found to have lower levels in the plasma of obese individuals when compared to their non-obese subjects (17). IL1RAP, enriched in the cytoplasm of liver cells and released into the bloodstream, shows variations in expression levels that correlate with the severity of NAFLD, as per our study. Nevertheless, its specific role in NAFLD progression remains an enigma.

CFHR3, displaying similar expression patterns and localization as IL1RAP, also warrants further investigation for its potential role in diagnosing NAFLD (30). The complementary DNAs (cDNAs) that encode CFHR3 were extracted from a human liver cDNA library. These have been located within the 'regulators of complement activation' (RCA) gene cluster situated on human chromosome 1 (41, 42). CFHR3 has been linked with some diseases, including hepatocellular carcinoma (HCC), although the underlying mechanism is still not clearly understood (43). Our research indicates that CFHR3 expression decreases in correlation with the severity of NAFLD, including both the NAS and fibrosis scores. This finding opens a new avenue of research into the pathophysiology of NAFLD.

The IL1 pathway, especially IL1B, has attracted significant attention and has been extensively researched within the realm of NAFLD (35, 44-47). IL-1β exacerbates liver steatosis, inflammation, and fibrosis by interacting with the IL-1 receptor, which is extensively expressed across various liver cell types (44). In 2021, Unamuno et al. discovered that in the liver of obese patients with NASH, there was a significant upregulation of mRNA levels for NLRP3 ($P < 0.05$) and its associated proinflammatory mediator IL1B ($P < 0.01$) (35). In 2018, Pan and colleagues showed that in NASH patients, IL-1β is primarily secreted by Kupffer cells, which are a type of monocyte in the liver. This secretion is driven by the activation of the NLRP3 inflammasome, stimulated by mitochondrial DNA (48). From our studies on cellular interactions within NAFLD, we have formed and to some extent validated a hypothesis that the IL-1β, when secreted, serves as a ligand affecting the receptors IL1R1 and IL1RAP in both hepatocytes and

cholangiocytes. This interaction between the ligand and its receptors seems to play a crucial role in modulating the expression of IL1RAP, which is closely associated with NAFLD severity. The dynamic interplay between these cellular components provides a deeper understanding of the disease's underlying mechanisms and offers potential avenues for NAFLD biomarkers and therapeutic interventions.

The utility of Seurat v5 in processing and integrating scRNA-seq and snRNA-seq data is apparent in our study. However, we must also acknowledge the inherent limitations of these techniques (49). Therefore, our findings must be interpreted with caution, and further confirmatory studies are required to validate the candidate biomarkers and the NAFLD pathogenesis.

Utilizing the 'CellChat' R package allowed for a streamlined analysis of the data. Given the vastness of the data, it was segmented into more digestible portions, sketching them into clusters of 5,000 cells each, ensuring a more organized scrutiny. An outstanding revelation from this analysis was the significant presence of ligand-receptor pairs of secreted signaling within the NAFLD group, which constituted a remarkable 61.8% of all categories. This prevalence underscored their crucial role in the interactions between liver and PBMC cells. The interaction of IL1B-(IL1R1+IL1RAP) between liver monocytes and hepatocyte/cholangiocyte was of particular interest, suggesting a plausible explanation for the detected decline in IL1RAP expression in the liver - a potential indicator of NAFLD severity. Referring back to our prior research, the crosstalk analysis, built on sc- and snRNAseq data from NAFLD patients, spotlighted SDC4 (in association with ANGPTL4) and ERBB3 (aligned with NRG1) as receptors of importance. The relationships between ANGPTL4-SDC4 and NRG1-ERBB3 with liver hepatocytes and hepatocyte/cholangiocytes, respectively, reflect their likely significance in the liver's cellular dynamics. This proposition is bolstered by a violin plot that vividly illustrates the expression patterns of three crucial down-regulated secreted genes from the crosstalk analysis, with hepatocytes prominently featuring as their central hub.

Syndecan-4 (SDC4) is a transmembrane protein that is glycosylated and prevalent in numerous tissues (50). Essential cellular functions such as differentiation, adhesion, migration, and prognosis are facilitated by SDC4, given its ability to bind to a variety of external molecules and engage in signal transmission (51-53).

ERBB3, or Receptor tyrosine-protein kinase erbB-3, is a membrane-associated protein from the EGFR/ERBB family of receptor tyrosine kinases. Research has shown its involvement in numerous liver conditions, including liver fibrosis (54, 55).

In conclusion, our research sheds light on the intricacies of NAFLD pathogenesis and highlights the significance of newly identified biomarkers in tracking disease progression. These insights serve as a foundation for subsequent studies delving into the molecular intricacies of NAFLD, all aiming towards enhancing disease diagnosis, management, and the outcomes of patients.

Abbreviations

avg_log2FC, log fold-change of the average expression between the two groups; BM, basement membrane; cDC1, type 1 conventional dendritic cell; cDNAs, complementary DNAs; ECM, extracellular matrix; F, Fibrosis score; GEO, Gene Expression Omnibus; GO, Gene Ontology; GRCh37, Genome Reference Consortium Human Build 37; HCC, hepatocellular carcinoma; HSCs, hepatic stellate cells; IHC, immunohistochemistry staining; IL-1RI, type I IL-1 receptor; IL1RAP, Interleukin-1 receptor accessory protein; LECs, lymphatic endothelial cells; N, NAS score; NAFL, Non-alcoholic Fatty Liver; NAFLD, Non-alcoholic fatty liver disease; NAFLD-DB, NAFLD gene expression database; NAS, NAFLD activity scores; NASH, Non-alcoholic Steatohepatitis; NK cells, Natural killer cells; PBMC, Peripheral blood mononuclear cell; PCA, Principal components analysis; PCs, Principal Components; QSOX1, Quiescin sulfhydryl oxidase 1; RCA, regulators of complement activation; RNA-seq, RNA sequencing; RPCA, reciprocal principal components analysis; scRNA-seq,

534 single-cell RNA sequencing; sIL1RAP, the soluble isoform of the IL-1 receptor
535 accessory protein; SNN, shared nearest neighbor; snRNA-seq, Single-nucleus RNA
536 sequencing; SZTCMH, Shenzhen Traditional Chinese Medicine Hospital, China; UMAP,
537 Uniform Manifold Approximation and Projection.
538

539 **Financial support**

540 This research was funded by the Shenzhen Sanming Project of Medicine in Shenzhen,
541 China (grant nos. SZSM201612074).
542

543 **Authors' contributions**

544 WFM and LL designed the study and interpreted the data. The analysis strategy has
545 been developed by LL and WFM. WFM and LL collected, analyzed, and interpreted
546 the data. WFM drafted the manuscript. YLL, HG, and MBK revised the manuscript.
547 WFM and LL performed data analysis and interpretation. BQC, JRH, MLL, XZ, SMX,
548 BLZ, QL, QH, MQM contributed to the validation of the candidate biomarkers.
549 Technical support: LL and JRH. All authors reviewed and approved the final version of
550 the manuscript.
551

552 **Data and code availability statement**

553 The data and code that support the findings of this study are available at
554 https://github.com/cynthia139/NAFLD-scRNA-seq.
555

556 **Conflict of interest**

557 Henning Grønbæk has received research grants from Abbvie, Intercept, ARLA Food
558 for Health, ADS AIPHIA Development Services AG. Consulting Fees from Ipsen, NOVO,
559 Pfizer. Lecturer for AstraZeneca and EISAI; and on Data Monitoring Committee at
560 CAMURUS AB. All other authors have no conflicts of interest to declare.
561

562 ## Figure Legend

563

564 **Figure 1. Liver single-cell and nucleus RNA sequencing data integration**
565 A. Illustration of data processing.
566 B.Uniform Manifold Approximation and Projection (UMAP) based on the
567 unintegrated data.
568 C. UMAP based on the anchor-based reciprocal principal components analysis (RPCA)
569 integrated data.
570 D.UMAP based on the health and NAFLD groups after RPCA integration.
571 E. UMAP based on the library type after RPCA integration.
572 F. UMAP based on the input cell type after RPCA integration.
573

574 **Figure 2. Integrated liver transcriptome data annotation, analysis and verification**
575 A. Gene markers of liver cell types.
576 B. Total 12 distinct cell types were clustered according to the canonical markers.
577 C. Dot plot of 16 up-regulated secreting marker expressions in cell clusters.
578 D. Dot plot of 22 down-regulated secreting marker expressions in cell clusters.
579 E. Gene markers of HSC (hepatic stellate cell)/MFB (myofibroblast) cell types.
580 F. Total 2 distinct cell types were clustered according to the canonical markers.
581 G. Violin plot of QSOX1 gene expressions in the hepatic stellate cell and
582 myofibroblast.
583 H. Violin plot of IL1RAP, CFHR3 gene expressions in the Hepatocytes.
584 I.Representative immunohistochemical staining (IHC) images of QSOX1 in liver
585 biopsies from mild NAFLD patients (N0-4, F0-2) and severe NAFLD patients (N5-8,
586 F3-4) in our previous research (16).
587 J. Representative IHC images of IL1RAP in liver biopsies from mild NAFLD patients
588 (N0-4, F0-2) and severe NAFLD patients (N5-8, F3-4) in our previous research (16).
589 Statistical testing was performed using the Wilcoxon rank sum test, with p-values
590 shown in the plot.
591

592 **Figure 3. Tracing the Cellular Origins in Liver and PBMC Data Integration**
593 A. Uniform Manifold Approximation and Projection (UMAP) based on the
594 unintegrated data.Dot plot of 16 up-regulated secreting marker expressions in cell
595 clusters.
596 B. UMAP based on the anchor-based reciprocal principal components analysis (RPCA)
597 integrated data.
598 C. UMAP based on the health and NAFLD groups after RPCA integration.
599 D. UMAP based on the input cell type after RPCA integration.
600 E. Gene markers of PBMC cell types.
601 F. Total 6 distinct cell types were clustered according to the canonical markers.
602 G. Dot plot of 16 up-regulated secreting marker expressions in cell clusters.
603 H. Dot plot of 22 down-regulated secreting marker expressions in cell clusters.
604 I. Violin plot of QSOX1 gene expressions in the NK cell.

J. Violin plot of IL1RAP gene expressions in the NK cell.
Statistical testing was performed using the Wilcoxon rank sum test, with p-values shown in the plot.

Figure 4. The cell communication in the liver and PBMC transcriptomics data
A. UMAP based on the anchor-based reciprocal principal components analysis (RPCA) integrated liver and PBMC data.
B. Ligand-receptor pairs of secreted signaling, which constitutes 61.8% of all categories.
C. The interaction between IL1B-(IL1R1+IL1RAP) in liver monocytes and hepatocyte/cholangiocyte.
D. The ANGPTL4-SDC4 pairing occurs between liver hepatocytes and hepatocyte/cholangiocyte.
E. The NRG1-ERBB3 interaction is evidenced between liver hepatocytes and cholangiocytes
F. The interaction between IL1B-(IL1R1+IL1RAP) in liver monocytes and hepatocyte/cholangiocyte.
G. The ANGPTL4-SDC4 pairing occurs between liver hepatocytes and hepatocyte/cholangiocyte.
H. Violin plot of IL1B, IL1R1, IL1RAP, ANGPTL4, SDC4, NRG1, ERBB3 gene expressions in the liver and PBMC.

List of Supplementary Tables

https://drive.google.com/file/d/1Eji-PHKMxil246MClmnh3PZPICD-nlCA/view?usp=sharing

Supplementary Table 1

The scRNAseq and snRNA-seq data of liver and PBMC samples.

Supplementary Table 2

The detailed information regarding the data alignment of the scRNAseq and snRNA-seq data.

Supplementary Table 3

The canonical markers of cell types in the liver and PBMC.

Supplementary Table 4

The ligand-receptor pairs in the CellChat database.

Supplementary Table 5

The pathways in the CellChat database.

642 # Reference

643 1. Loomba R, Friedman SL, Shulman GI. Mechanisms and disease consequences of nonalcoholic
644 fatty liver disease. Cell 2021;184:2537-2564.
645 2. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. Lancet 2021;397:2212-2224.
646 3. Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. Lancet
647 2022;400:1345-1362.
648 4. Chen Y, Wang W, Morgan MP, Robson T, Annett S. Obesity, non-alcoholic fatty liver disease and
649 hepatocellular carcinoma: current status and therapeutic targets. Front Endocrinol (Lausanne)
650 2023;14:1148934.
651 5. Younossi ZM, Golabi P, Paik JM, Henry A, Van Dongen C, Henry L. The global epidemiology of
652 nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review.
653 Hepatology 2023;77:1335-1347.
654 6. Khairnar R, Islam MA, Fleishman J, Kumar S. Shedding light on non-alcoholic fatty liver disease:
655 Pathogenesis, molecular mechanisms, models, and emerging therapeutics. Life Sci 2023;312:121185.
656 7. Musso G, Saba F, Cassader M, Gambino R. Lipidomics in pathogenesis, progression and treatment
657 of nonalcoholic steatohepatitis (NASH): Recent advances. Prog Lipid Res 2023;91:101238.
658 8. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet
659 2019;20:631-656.
660 9. Andrews TS, Atif J, Liu JC, Perciani CT, Ma XZ, Thoeni C, Slyper M, et al. Single-Cell, Single-Nucleus,
661 and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal
662 Heterogeneity. Hepatol Commun 2022;6:821-840.
663 10. Alvarez M, Benhammou JN, Darci-Maher N, French SW, Han SB, Sinsheimer JS, Agopian VG, et al.
664 Human liver single nucleus and single cell RNA sequencing identify a hepatocellular
665 carcinoma-associated cell-type affecting survival. Genome Med 2022;14:50.
666 11. Xiong X, Kuang H, Ansari S, Liu T, Gong J, Wang S, Zhao XY, et al. Landscape of Intercellular
667 Crosstalk in Healthy and NASH Liver Revealed by Single-Cell Secretome Gene Analysis. Mol Cell
668 2019;75:644-660 e645.
669 12. Tamburini BAJ, Finlon JM, Gillen AE, Kriss MS, Riemondy KA, Fu R, Schuyler RP, et al. Chronic Liver
670 Disease in Humans Causes Expansion and Differentiation of Liver Lymphatic Endothelial Cells. Front
671 Immunol 2019;10:1036.
672 13. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, Portman JR, et
673 al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature 2019;575:512-518.
674 14. Deczkowska A, David E, Ramadori P, Pfister D, Safran M, Li B, Giladi A, et al. XCR1(+) type 1
675 conventional dendritic cells drive liver pathology in non-alcoholic steatohepatitis. Nat Med
676 2021;27:1043-1054.
677 15. Wang S, Li K, Pickholz E, Dobie R, Matchett KP, Henderson NC, Carrico C, et al. An autocrine
678 signaling circuit in hepatic stellate cells underlies advanced fibrosis in nonalcoholic steatohepatitis. Sci
679 Transl Med 2023;15:eadd3949.
680 16. Ma W, Huang J, Cai B, Shao M, Yu X, Kjær MB, Lv M, et al. A novel gene-screening approach
681 reveals QSOX1/IL1RAP as promising biomarkers for the severity of non-alcoholic fatty liver disease.
682 medRxiv 2023:2023.2007.2026.23293038.

683 17. Bozaoglu K, Attard C, Kulkarni H, Cummings N, Diego VP, Carless MA, Shields KA, et al. Plasma
684 levels of soluble interleukin 1 receptor accessory protein are reduced in obesity. J Clin Endocrinol
685 Metab 2014;99:3435-3443.
686 18. Gao R, Wang J, He X, Wang T, Zhou L, Ren Z, Yang J, et al. Comprehensive analysis of endoplasmic
687 reticulum-related and secretome gene expression profiles in the progression of non-alcoholic fatty
688 liver disease. Front Endocrinol (Lausanne) 2022;13:967016.
689 19. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, et al. Dictionary
690 learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol 2023.
691 20. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, et al. Integrated
692 analysis of multimodal single-cell data. Cell 2021;184:3573-3587 e3529.
693 21. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, et al. Inference and
694 analysis of cell-cell communication using CellChat. Nat Commun 2021;12:1088.
695 22. Fang Z, Tian Y, Sui C, Guo Y, Hu X, Lai Y, Liao Z, et al. Single-Cell Transcriptomics of Proliferative
696 Phase Endometrium: Systems Analysis of Cell-Cell Communication Network Using CellChat. Front Cell
697 Dev Biol 2022;10:919731.
698 23. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, Smillie C, et al. A
699 single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat Med
700 2020;26:792-802.
701 24. Yang W, He H, Wang T, Su N, Zhang F, Jiang K, Zhu J, et al. Single-Cell Transcriptomic Analysis
702 Reveals a Hepatic Stellate Cell-Activation Roadmap and Myofibroblast Origin During Liver Fibrosis in
703 Mice. Hepatology 2021;74:2774-2790.
704 25. Rosenthal SB, Liu X, Ganguly S, Dhar D, Pasillas MP, Ricciardelli E, Li RZ, et al. Heterogeneity of
705 HSCs in a Mouse Model of NASH. Hepatology 2021;74:667-685.
706 26. She H, Xiong S, Hazra S, Tsukamoto H. Adipogenic transcriptional regulation of hepatic stellate
707 cells. J Biol Chem 2005;280:4959-4967.
708 27. Hundertmark J, Berger H, Tacke F. Single Cell RNA Sequencing in NASH. Methods Mol Biol
709 2022;2455:181-202.
710 28. Reznik N, Fass D. Disulfide bond formation and redox regulation in the Golgi apparatus. FEBS Lett
711 2022;596:2859-2872.
712 29. Coppock DL, Kopman C, Scandalis S, Gilleran S. Preferential gene expression in quiescent human
713 lung fibroblasts. Cell Growth Differ 1993;4:483-493.
714 30. Uhlen M, Karlsson MJ, Hober A, Svensson AS, Scheffel J, Kotol D, Zhong W, et al. The human
715 secretome. Sci Signal 2019;12.
716 31. Grossman I, Alon A, Ilani T, Fass D. An inhibitory antibody blocks the first step in the
717 dithiol/disulfide relay mechanism of the enzyme QSOX1. J Mol Biol 2013;425:4366-4378.
718 32. Ilani T, Reznik N, Yeshaya N, Feldman T, Vilela P, Lansky Z, Javitt G, et al. The disulfide catalyst
719 QSOX1 maintains the colon mucosal barrier by regulating Golgi glycosyltransferases. EMBO J
720 2023;42:e111869.
721 33. Kisseleva T, Brenner D. Molecular and cellular mechanisms of liver fibrosis and its regression. Nat
722 Rev Gastroenterol Hepatol 2021;18:151-166.
723 34. Kisseleva T. The origin of fibrogenic myofibroblasts in fibrotic liver. Hepatology
724 2017;65:1039-1043.

725 35. Unamuno X, Gomez-Ambrosi J, Ramirez B, Rodriguez A, Becerril S, Valenti V, Moncada R, et al.
726 NLRP3 inflammasome blockade reduces adipose tissue inflammation and extracellular matrix
727 remodeling. Cell Mol Immunol 2021;18:1045-1057.
728 36. Kim M, Lee C, Seo DY, Lee H, Horton JD, Park J, Scherer PE. The impact of endotrophin on the
729 progression of chronic liver disease. Exp Mol Med 2020;52:1766-1776.
730 37. Di Caprio N, Bellas E. Collagen Stiffness and Architecture Regulate Fibrotic Gene Expression in
731 Engineered Adipose Tissue. Adv Biosyst 2020;4:e1900286.
732 38. Frenay J, Bellaye PS, Oudot A, Helbling A, Petitot C, Ferrand C, Collin B, et al. IL-1RAP, a Key
733 Therapeutic Target in Cancer. Int J Mol Sci 2022;23.
734 39. Wesche H, Korherr C, Kracht M, Falk W, Resch K, Martin MU. The interleukin-1 receptor accessory
735 protein (IL-1RAcP) is essential for IL-1-induced activation of interleukin-1 receptor-associated kinase
736 (IRAK) and stress-activated protein kinases (SAP kinases). J Biol Chem 1997;272:7727-7731.
737 40. Greenfeder SA, Nunes P, Kwee L, Labow M, Chizzonite RA, Ju G. Molecular cloning and
738 characterization of a second subunit of the interleukin 1 receptor complex. J Biol Chem
739 1995;270:13757-13765.
740 41. Diaz-Guillen MA, Rodriguez de Cordoba S, Heine-Suner D. A radiation hybrid map of complement
741 factor H and factor H-related genes. Immunogenetics 1999;49:549-552.
742 42. Jozsi M, Meri S. Factor H-related proteins. Methods Mol Biol 2014;1100:225-236.
743 43. Wan Z, Li X, Luo X, Wang B, Zhou X, Chen A. The miR-590-3p/CFHR3/STAT3 signaling pathway
744 promotes cell proliferation and metastasis in hepatocellular carcinoma. Aging (Albany NY)
745 2022;14:5783-5799.
746 44. Mirea AM, Tack CJ, Chavakis T, Joosten LAB, Toonen EJM. IL-1 Family Cytokine Pathways
747 Underlying NAFLD: Towards New Treatment Strategies. Trends Mol Med 2018;24:458-471.
748 45. Tilg H, Effenberger M, Adolph TE. A role for IL-1 inhibitors in the treatment of non-alcoholic fatty
749 liver disease (NAFLD)? Expert Opin Investig Drugs 2020;29:103-106.
750 46. Gehrke N, Hofmann LJ, Straub BK, Ruhle F, Waisman A, Galle PR, Schattenberg JM. Hepatic
751 interleukin-1 receptor type 1 signalling regulates insulin sensitivity in the early phases of nonalcoholic
752 fatty liver disease. Clin Transl Med 2022;12:e1048.
753 47. He Y, Hwang S, Ahmed YA, Feng D, Li N, Ribeiro M, Lafdil F, et al. Immunopathobiology and
754 therapeutic targets related to cytokines in liver diseases. Cell Mol Immunol 2021;18:18-37.
755 48. Pan J, Ou Z, Cai C, Li P, Gong J, Ruan XZ, He K. Fatty acid activates NLRP3 inflammasomes in
756 mouse Kupffer cells through mitochondrial DNA release. Cell Immunol 2018;332:111-120.
757 49. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data
758 across different conditions, technologies, and species. Nat Biotechnol 2018;36:411-420.
759 50. Tanino Y, Wang X, Nikaido T, Misa K, Sato Y, Togawa R, Kawamata T, et al. Syndecan-4 Inhibits the
760 Development of Pulmonary Fibrosis by Attenuating TGF-beta Signaling. Int J Mol Sci 2019;20.
761 51. Becsky D, Gyulai-Nagy S, Balind A, Horvath P, Dux L, Keller-Pinter A. Myoblast Migration and
762 Directional Persistence Affected by Syndecan-4-Mediated Tiam-1 Expression and Distribution. Int J
763 Mol Sci 2020;21.
764 52. Chronopoulos A, Thorpe SD, Cortes E, Lachowski D, Rice AJ, Mykuliak VV, Rog T, et al. Syndecan-4
765 tunes cell mechanics by activating the kindlin-integrin-RhoA pathway. Nat Mater 2020;19:669-678.
766 53. Lambert J, Makin K, Akbareian S, Johnson R, Alghamdi AAA, Robinson SD, Edwards DR.
767 ADAMTS-1 and syndecan-4 intersect in the regulation of cell migration and angiogenesis. J Cell Sci
768 2020;133.

54. Scheving LA, Zhang X, Threadgill DW, Russell WE. Hepatocyte ERBB3 and EGFR are required for maximal CCl4-induced liver fibrosis. Am J Physiol Gastrointest Liver Physiol 2016;311:G807-G816.

55. Xiong J, Ni J, Chen C, Wang K. miR-148a-3p regulates alcoholic liver fibrosis through targeting ERBB3. Int J Mol Med 2020;46:1003-1012.

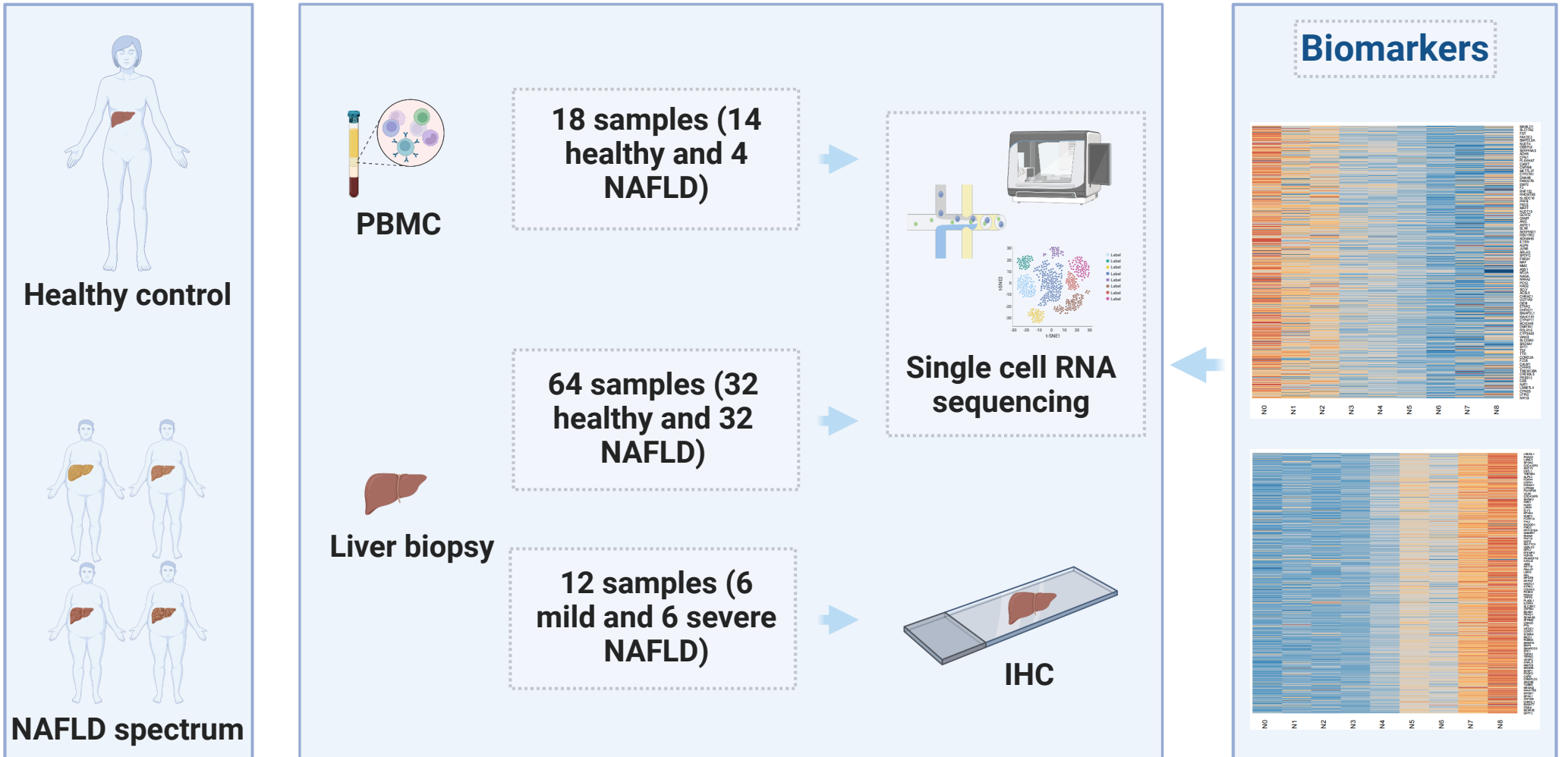
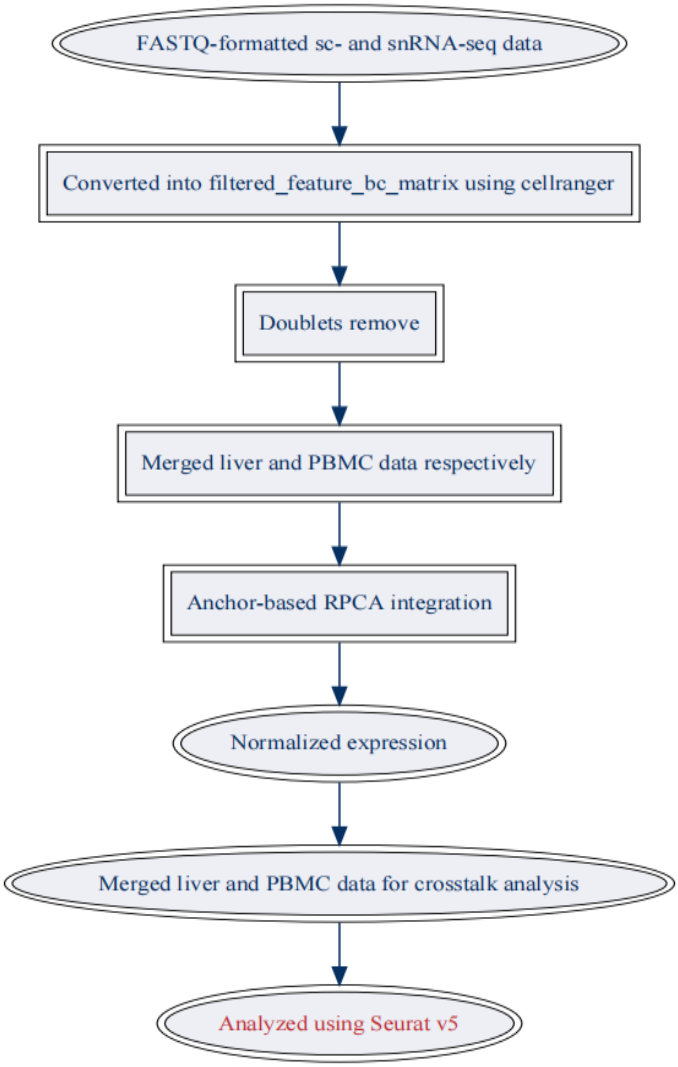
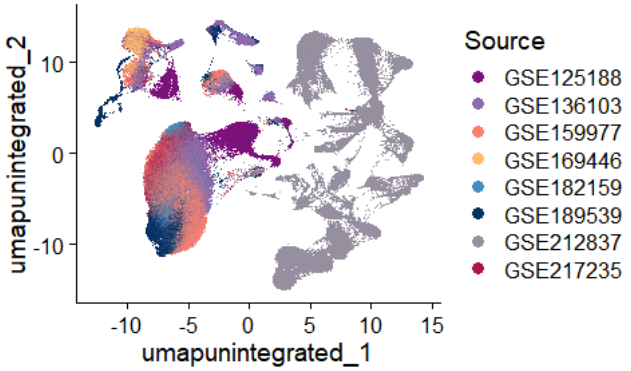

Healthy control
NAFLD spectrum
PBMC
Liver biopsy
18 samples (14 healthy and 4 NAFLD)
64 samples (32 healthy and 32 NAFLD)
12 samples (6 mild and 6 severe NAFLD)
Single cell RNA sequencing
IHC
Biomarkers
Created with BioRender.com

Figure1

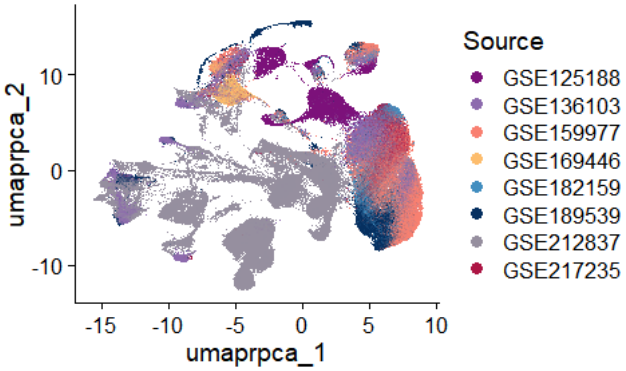
A



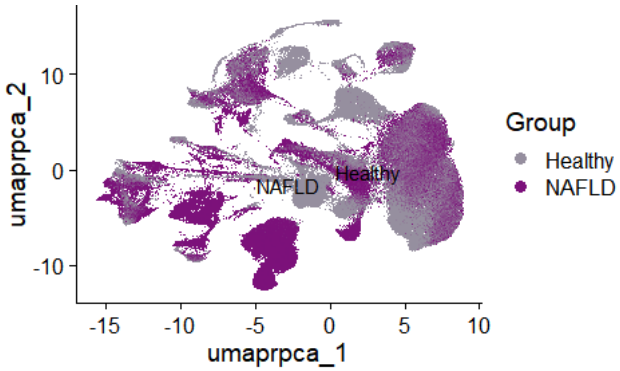
B



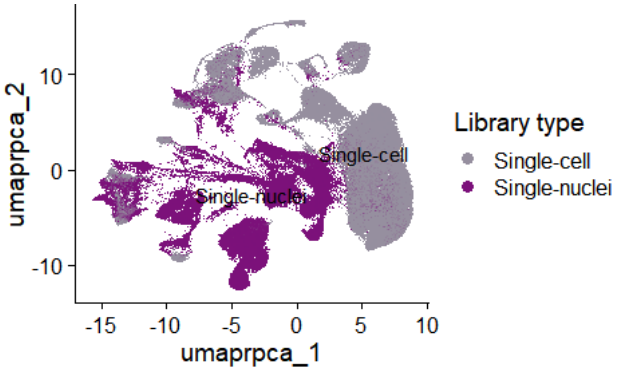
C



D



E



F

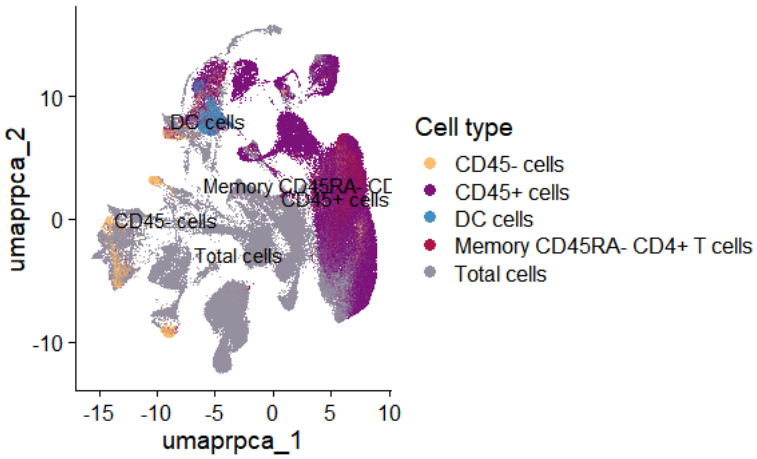


Figure 2

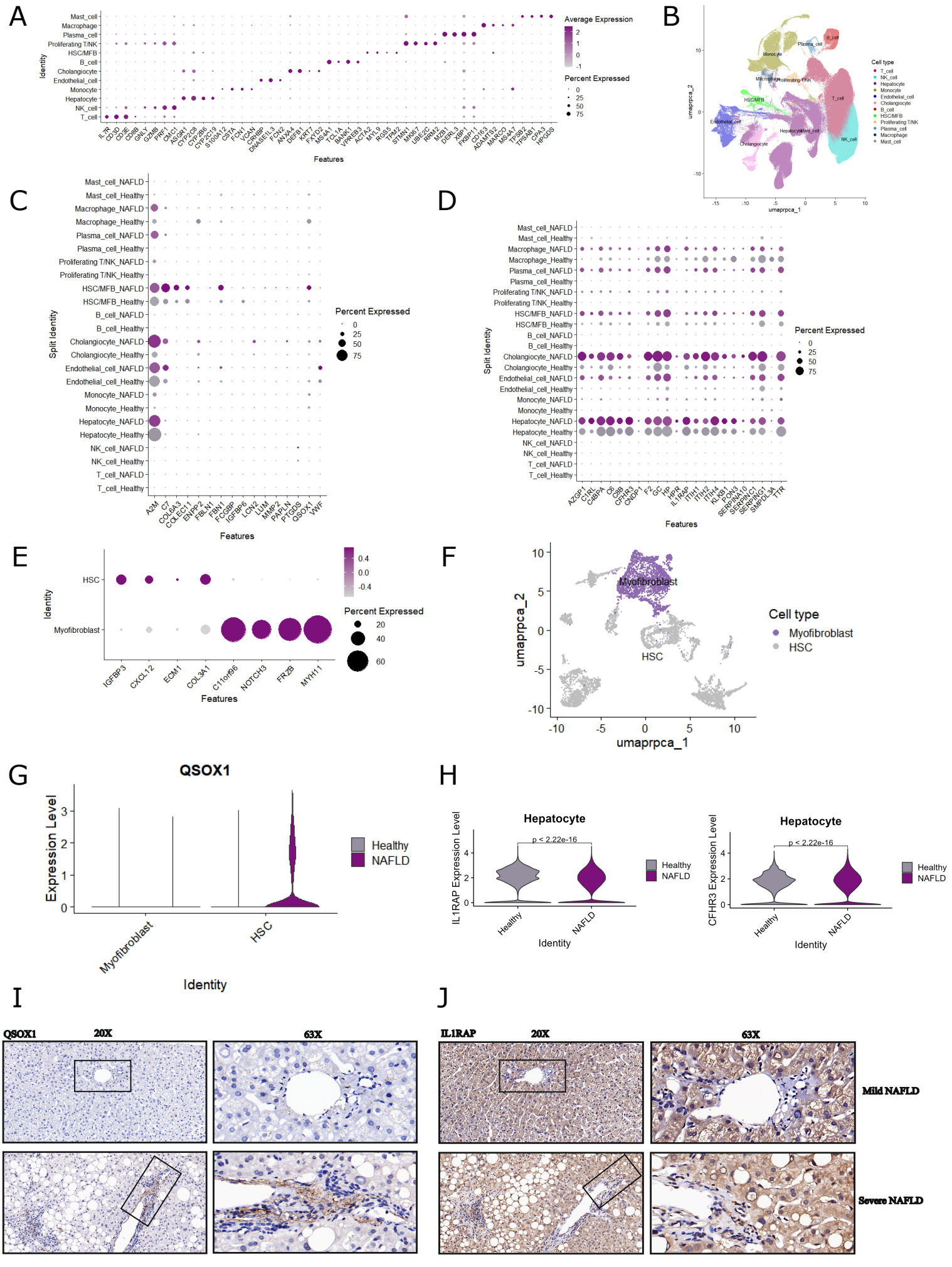

A
Mast_cell
Macrophage
Plasma_cell
Proliferating T/NK
HSC/MFB
B_cell
Cholangiocyte
Endothelial_cell
Monocyte
Hepatocyte
NK_cell
T_cell
Identity
Features
Average Expression
Percent Expressed
B
Cell type
T_cell
NK_cell
Hepatocyte
Monocyte
Endothelial_cell
Cholangiocyte
B_cell
HSC/MFB
Proliferating T/NK
Plasma_cell
Macrophage
Mast_cell
umaprrpca_1
umaprrpca_2
C
Mast_cell_NAFLD
Mast_cell_Healthy
Macrophage_NAFLD
Macrophage_Healthy
Plasma_cell_NAFLD
Plasma_cell_Healthy
Proliferating T/NK_NAFLD
Proliferating T/NK_Healthy
HSC/MFB_NAFLD
HSC/MFB_Healthy
B_cell_NAFLD
B_cell_Healthy
Cholangiocyte_NAFLD
Cholangiocyte_Healthy
Endothelial_cell_NAFLD
Endothelial_cell_Healthy
Monocyte_NAFLD
Monocyte_Healthy
Hepatocyte_NAFLD
Hepatocyte_Healthy
NK_cell_NAFLD
NK_cell_Healthy
T_cell_NAFLD
T_cell_Healthy
Split Identity
Features
Percent Expressed
D
Split Identity
Features
Percent Expressed
E
HSC
Myofibroblast
Identity
Features
Percent Expressed
F
Myofibroblast
HSC
Cell type
umaprrpca_1
umaprrpca_2
G
QSOX1
Expression Level
Healthy
NAFLD
Myofibroblast
HSC
Identity
H
Hepatocyte
IL1RAP Expression Level
CFHR3 Expression Level
p < 2.22e-16
Healthy
NAFLD
Identity
I
QSOX1
20X
63X
J
IL1RAP
20X
63X
Mild NAFLD
Severe NAFLD

Figure 3

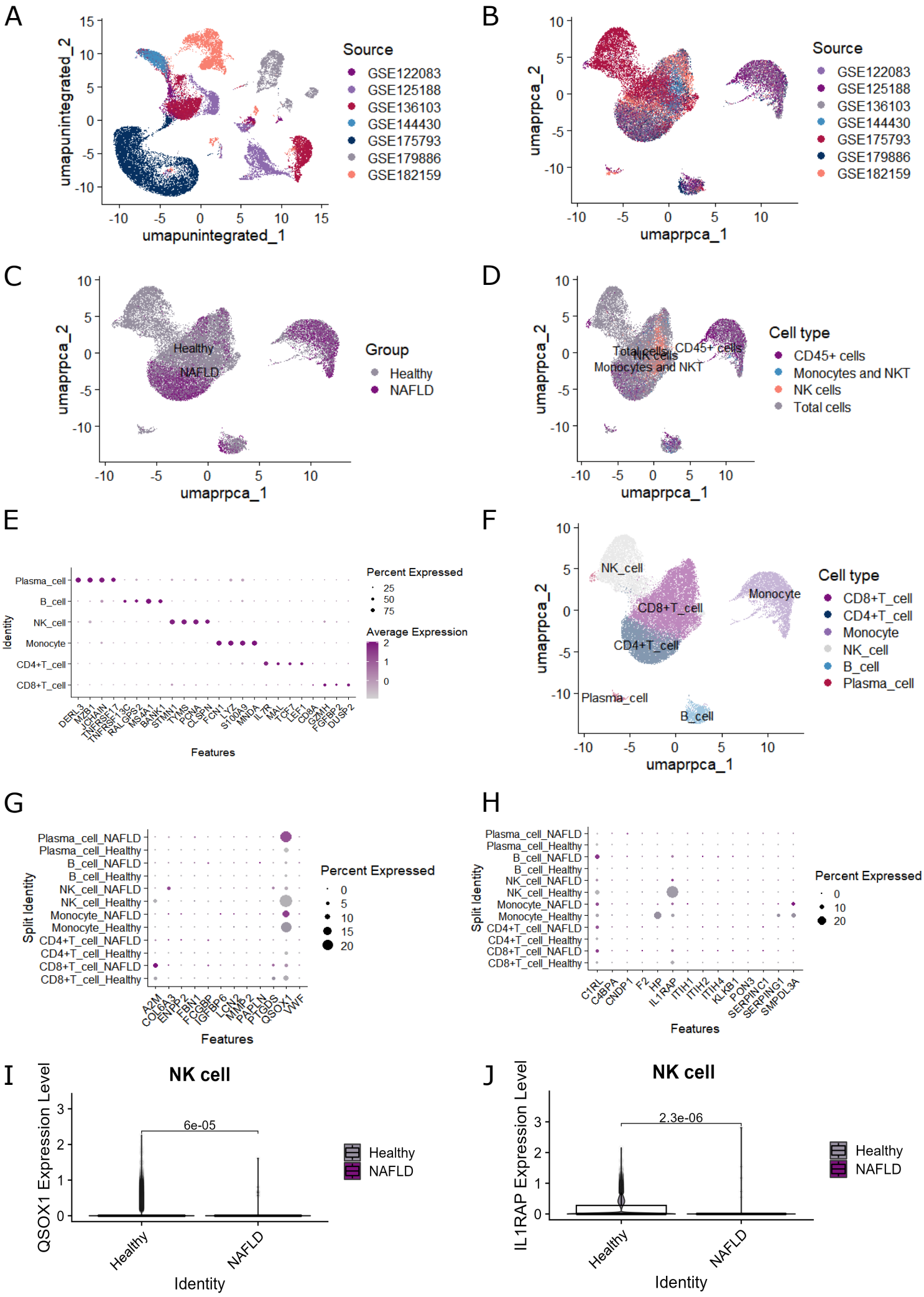
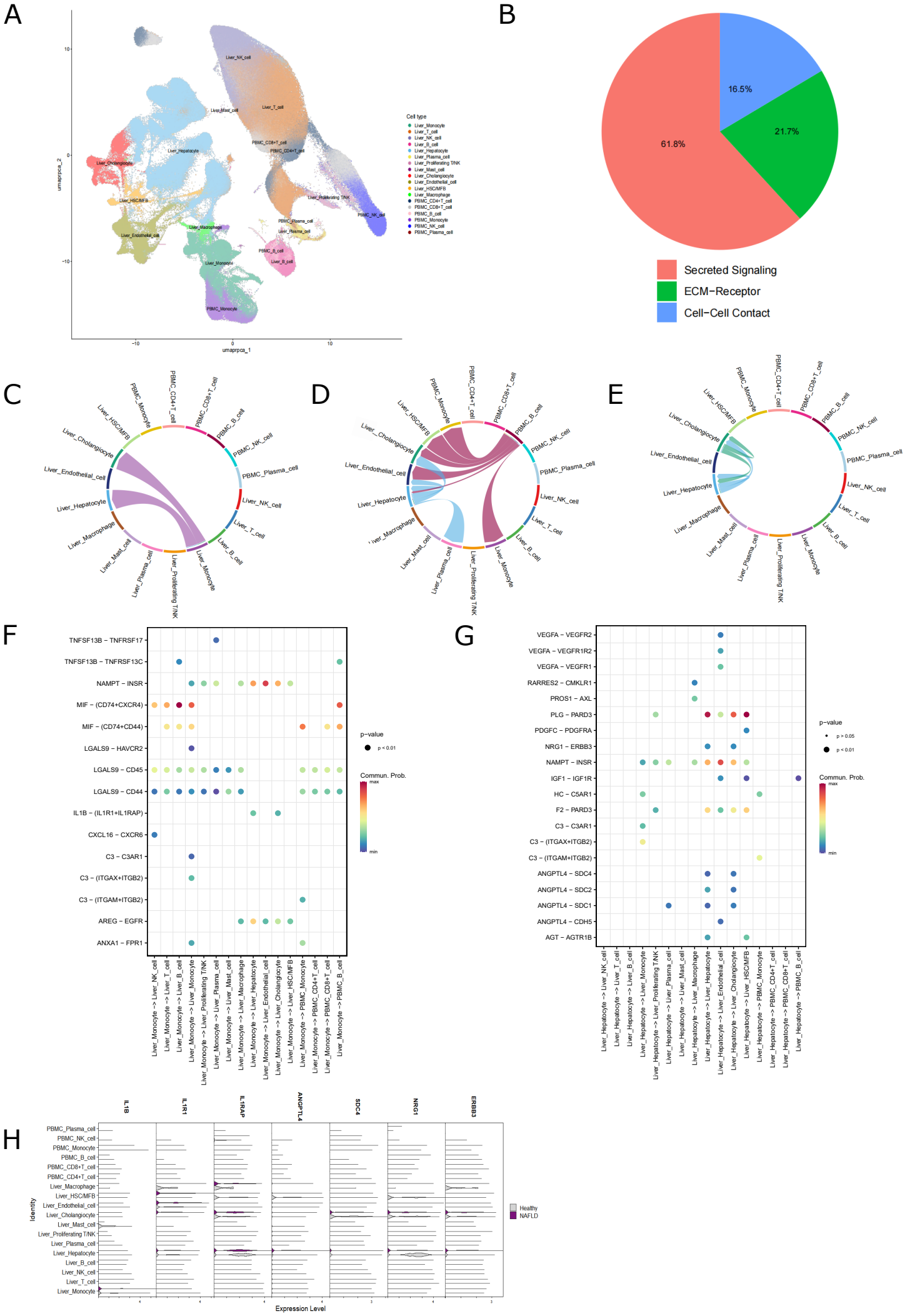

A
umapunintegrated_2
umapunintegrated_1
Source
GSE122083
GSE125188
GSE136103
GSE144430
GSE175793
GSE179886
GSE182159
B
umaprpca_2
umaprpca_1
Source
GSE122083
GSE125188
GSE136103
GSE144430
GSE175793
GSE179886
GSE182159
C
umaprpca_2
umaprpca_1
Healthy
NAFLD
Group
Healthy
NAFLD
D
umaprpca_2
umaprpca_1
Total cells
NK cells
CD45+ cells
Monocytes and NKT
Cell type
CD45+ cells
Monocytes and NKT
NK cells
Total cells
E
Plasma_cell
B_cell
NK_cell
Monocyte
CD4+T_cell
CD8+T_cell
Identity
DERL3
MZB1
HCHA1N
TNFRSF17
TNFRSF13C
RALGPS2
MS4A1
BANK1
STMN1
TYMS
PCNA
CLSPN
FCN1
LYZ
S100A9
MNDA
IL7R
MAL
TCF7
LEF1
CD8A
GZMH
FGFBP2
DUSP2
Features
Percent Expressed
25
50
75
Average Expression
2
1
0
F
umaprpca_2
umaprpca_1
NK_cell
CD8+T_cell
CD4+T_cell
Monocyte
Plasma_cell
B_cell
Cell type
CD8+T_cell
CD4+T_cell
Monocyte
NK_cell
B_cell
Plasma_cell
G
Plasma_cell_NAFLD
Plasma_cell_Healthy
B_cell_NAFLD
B_cell_Healthy
NK_cell_NAFLD
NK_cell_Healthy
Monocyte_NAFLD
Monocyte_Healthy
CD4+T_cell_NAFLD
CD4+T_cell_Healthy
CD8+T_cell_NAFLD
CD8+T_cell_Healthy
Split Identity
A2M
COL6A3
ENPP2
FBN1
FCGBP
IGFBP6
LCN2
MMP2
PAPLN
PTGDS
QSOX1
VWF
Features
Percent Expressed
0
5
10
15
20
H
Plasma_cell_NAFLD
Plasma_cell_Healthy
B_cell_NAFLD
B_cell_Healthy
NK_cell_NAFLD
NK_cell_Healthy
Monocyte_NAFLD
Monocyte_Healthy
CD4+T_cell_NAFLD
CD4+T_cell_Healthy
CD8+T_cell_NAFLD
CD8+T_cell_Healthy
Split Identity
C1RL
C4BPA
CNDP1
F2
HP
IL1RAP
ITIH1
ITIH2
ITIH4
KLKB1
PON3
SERPINC1
SERPING1
SMPDL3A
Features
Percent Expressed
0
10
20
I
NK cell
QSOX1 Expression Level
6e-05
Healthy
NAFLD
Identity
J
NK cell
IL1RAP Expression Level
2.3e-06
Healthy
NAFLD
Identity

Figure 4


A
B
16.5%
21.7%
61.8%
Secreted Signaling
ECM−Receptor
Cell−Cell Contact
C
D
E
F
G
H
IL1B
IL1R1
IL1RAP
ANGPTL4
SDC4
NRG1
ERBB3
Expression Level
Healthy
NAFLD