Identification of hepatocyte immune response in Autoimmune Hepatitis from human plasma cfChIP-seq

Gavriel Fialkoff[1,2], Ami Ben Ya'akov[3], Israa Sharkia[1,2], Ronen Sadeh[1,2], Jenia Gutin[1,2], Chen Goldstein[1], Abed khalaileh[4], Ashraf Imam[4], Rifaat Safadi[5], Yael Milgrom[5], Eithan Galun[6], Eyal Shteyer[3,#,✉], Nir Friedman[1,2,#,✉]

[1] The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel

[2] Institute for Medical Research Israel-Canada, The Hebrew University-Hadassah Medical School, Jerusalem, Israel

[3] The Juliet Keidan Institute of Pediatric Gastroenterology, Shaare Zedek Medical Center, Jerusalem, Israel

[4] Department of General surgery and Transplantation Unit, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

[5] The Liver Unit Institute of Gastroenterology and Liver Diseases, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

[6] The Goldyne Savad Institute for Gene Therapy, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

[#] equal contribution

[✉] corresponding authors

Abstract

Autoimmune hepatitis (AIH) is a self-perpetuating inflammatory liver disease with significant morbidity and mortality risks. Patients undergo liver biopsy to confirm diagnosis and affirm subsequent remission. Advances in liquid biopsies show promise to replace tissue biopsy in cancer, however little research has been done in liver disease. Here, we use plasma chromatin immunoprecipitation and sequencing (cfChIP-seq) to analyze cell-free nucleosomes carrying an active histone modification which reports on gene transcription in the dying cells. Comparing plasma samples from pediatric AIH patients to a control group we identify immune-related transcriptional processes activated in hepatocytes of AIH patients. We devise a classifier that based on cfChIP-seq profiles distinguishes AIH from other conditions involving increased liver damage. Our work demonstrates the potential of plasma cfChIP-seq as a non-invasive diagnostic tool for AIH, which could replace the need for liver biopsy, aid accurate diagnoses, and enable further scientific exploration of AIH pathogenesis.

Introduction

Autoimmune hepatitis (AIH) is a rare chronic self-perpetuating inflammatory liver disease, characterized by immune-mediated damage to hepatocytes. The clinical presentation of AIH is heterogeneous and includes elevated serum transaminases and seropositivity of autoantibodies and immunoglobulin G, yet the final diagnosis requires histological evidence of hepatic inflammation and interface hepatitis with increased plasma cell which entails liver biopsy (reviewed in [1,2]). Several lines of evidence suggest that hepatocyte damage in AIH is mediated by CD4+ T-cells, particularly the Th17 cells[3], though the underlying mechanisms are not fully understood. Immunosuppression and liver transplantation in severe cases of liver failure or cirrhosis, are the sole therapeutic alternatives. Normalization of transaminase levels along with IgG levels and negative auto-antibodoies define biochemical remission of the disease. Biochemical remission is usually a sufficient indication of successful response to treatment but does not always correlate with histological remission[4]. Thus, in most cases liver biopsy is needed to confirm histological remission to allow stopping medications.

Cell-free DNA liquid biopsies have emerged in the past two decades as a powerful tool for diagnosing and monitoring diseases and enabled their introduction into clinical practice mainly in the field of cancer[5]. However, little research has been performed on the use of cfDNA in non cancerous liver diseases, autoimmune maladies, and AIH. We recently reported chromatin immunoprecipitation and sequencing of cell-free nucleosomes from human plasma (cfChIP-seq) to infer the transcriptional programs by genome-wide mapping of plasma cell free-nucleosomes carrying specific histone modifications[6]. Specifically, tri-methylation of histone 3 lysine 4 (H3K4me3) is a well characterized histone modification, marking transcription start sites (TSS) of genes that are poised or actively transcribed, and predictive of gene expression [6–9].

We hypothesized that understanding sources of cfDNA in AIH would provide additional insights on the pathogenesis of the disease and assist in the challenging diagnosis process.

Results

Elevated liver-derived cfDNA in AIH plasma samples

Recently we described cfChIP-seq, a method for performing chromatin immunoprecipitation and sequencing from plasma[6]. Here, we used cfChIP-seq with H3K4me3-specific antibody which enriches for poised and active transcription start sites (TSS), on plasma samples from

37 plasma samples from pediatric patients with autoimmune hepatitis (n=27 patients) - either at diagnosis, with elevated liver transaminases or in biochemical remission (ALT and AST liver enzymes within the normal range) under immunosuppressive therapy. As control, we also included an additional cohort of 14 self-reported healthy donors (six children and eight adults) and a cohort of 58 samples from 56 patients with other liver diseases (**Fig. 1A; Supplementary tables 1-2**).

For quality control we examine the yield of the assay and its specificity. The average yield of the cfChIP-seq samples was 2.8 and 1.4 million unique reads for the AIH and healthy samples respectively (**Fig. S1A; Supplementary table 3**), presumably reflecting elevated cfCDNA levels in the AIH samples. The specificity of cfChIP-seq is defined as the proportion of reads that map to gene promoters vs. reads that are non-specific background. The average specificity of the samples is 70% (**Supplementary table 3,** Methods).

The self reported healthy control cohort consists of samples from children and adults. The means of the two groups were highly correlated (R = 0.99), and individual samples were also highly correlated (R > 0.95, median R = 0.975; **Fig. S1B-C**) and were therefore treated as one unified control group for downstream analysis.

The results of cfChIP-seq are analyzed at the level of genes. Briefly, reads are mapped to the genome and the number of normalized reads mapping to every gene's TSS regions was computed, resulting in gene counts resembling RNA-seq transcription counts (Methods). We then compared the gene counts of plasma samples from AIH patients to a healthy baseline reference (Methods) and found hundreds of genes that were significantly increased in AIH patients with active disease, many of which were shared among several samples. In plasma samples from patients in remission, we observed a smaller group of genes elevated compared to healthy, and some samples seemed identical to healthy plasma with no genes significantly elevated (**Fig. 1B-C; S1D**).

To identify the tissues and cell types that contribute to the elevated gene signal in AIH patients, we compared the profile of these genes to a comprehensive reference atlas of 182 H3K4me3 ChIP-seq samples from 36 tissues and cell-types, including solid tissues and immune cells [10,11]. Examining the set of genes which were significantly elevated in at least 3 AIH samples in the reference data, exhibits a low coverage in the immune cells which are the main source of cfDNA in healthy individuals [6,12]. A subset of the genes has some coverage in all solid tissues, and the majority of genes are marked by H3K4me3 only in the liver samples (**Fig. 1D**). Enrichment tests of this gene set finds a strong enrichment of the liver (EnrichR human gene atlas $q<10^{-80}$; gene overlap 140/618), reestablishing the identification of the liver as a major source of cfDNA in AIH samples.

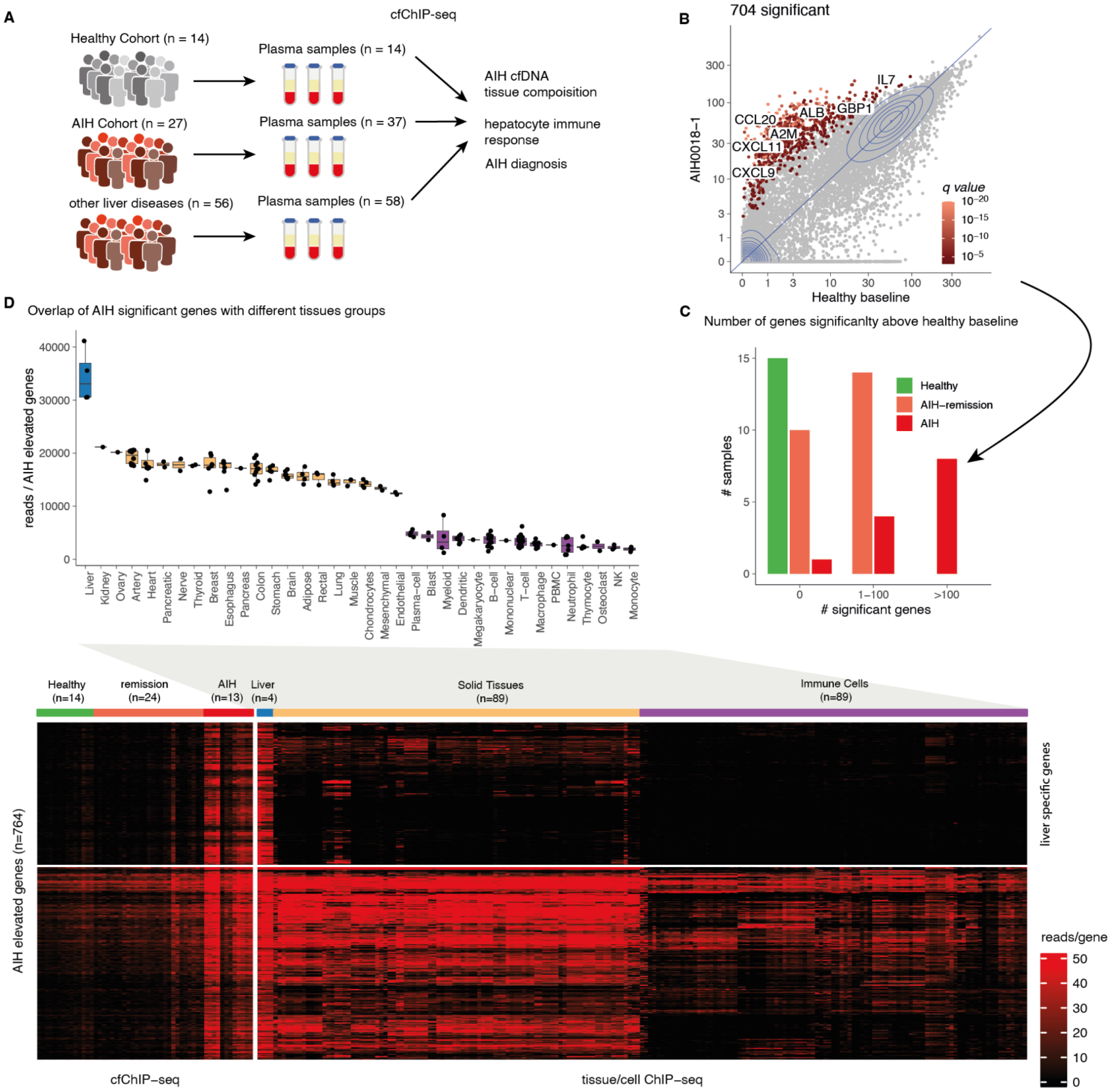

Figure 1: cfChIP-seq displays elevated liver-derived cfDNA in AIH plasma samples

A. Study outline. Plasma samples were collected from a cohort of healthy individuals, patients with AIH and patients with other liver diseases. cfChIP-seq was performed on 1ml of plasma to recover AIH cfDNA tissue-of-origin, infer transcriptional programs in dying cells and assist AIH diagnosis.

B. Detection of genes with significantly elevated coverage in a representative AIH plasma sample. For each gene, the normalized promoter coverage in the sample (y-axis) was compared to the mean of a reference healthy cohort (x-axis). Significance indicates whether the observed number of reads in the sample is significantly higher than expected based on the mean and variance in healthy samples (Methods).

C. Distribution of the number of genes with significantly elevated promoter coverage in sample compared to healthy baseline (Methods). Samples were binned to three groups: samples with 0, 1-100 and more than 100 genes with significant coverage

above healthy. AIH samples were grouped by Alanine Transferase (ALT) to active (> 40 U/L) and remission (<40 U/L) states.

D. Bottom: Hierarchical clustering of genes with significant coverage in at least 3 AIH samples (764 genes, rows) across cfChIP-seq samples (left columns) and reference tissue and cell types ChIP-seq samples (right columns). Color bar of reference data represents tissue category (liver, solid tissue or immune cells) as elaborated in the boxplot above. Color scale represents the normalized promoter coverage. Rows were hierarchically clustered and split at the root of the hierarchy (depth = 1).
Top: Median and distribution of cumulative signal of the same genes across reference tissues and cell types (right side of heatmap). The genes elevated in the AIH samples are significantly enriched for the liver.

cfChIP-seq recovers AIH cell-free DNA cell-of-origin

To quantify the relative contribution of liver-derived cfDNA in the circulation and achieve a systematic view of other tissues contributing to the circulation, we used a linear regression deconvolution of samples to their composing cell-types (Methods). Examining the results we find that in healthy donors, the major components of the cfDNA are the peripheral blood mononuclear cell (neutrophils, megakaryocytes, B cells and monocytes) which is in agreement with previous studies [6,13], while the liver constitutes less than 1% of the cfDNA an average. In contrast, in samples of AIH patients with active disease, the liver accounts for 15%-65% of the cfDNA, and increased levels are observed also in some of the patients in remission (*t-test P*=0.0002 and 0.01 in the active and remission AIH samples respectively). An additional, more subtle elevation is observed also in the T cell fraction of some AIH samples (**Fig. 2A, S2A**). Note that the estimated fractions represent the relative contribution of the tissues to the circulation, and not the absolute cell death of these tissues. Thus, an increase in liver fraction must be compensated by reduction of fractions of other tissues even if their absolute levels remain the same (**Fig. S2B**).

To further test whether cfChIP-seq can provide clues as to the AIH specific cell-type of origin within the liver, we used gene signatures derived from a single cell RNA-seq liver atlas[14], since no such ChIP-seq data exist. Across the 10 liver cell-types examined, including hepatocyte, cholangiocyte, endothelial, stellate and immune cells, the AIH samples are enriched specifically for hepatocyte marker genes such as *HPX* (hemopexin), *F12* (coagulation factor XII) and *APOB* (Apolipoprotein B) (**Fig. 2B-C**). The remarkably positive correlation of the hepatocyte marker genes and the liver fraction (R = 0.98; **Fig. S2C**) further corroborates the finding that the hepatocytes are indeed the major source of liver cfDNA in AIH plasma samples.

Comparison of the estimated liver fraction to the liver alanine transferase (ALT) levels measured in time-matched blood samples, displays a good agreement between the two modalities despite the differences in half-life of these analytes [15,16](R = 0.87; p = 7.3^{-12}; **Fig. 2D**). When performing principal component analysis (PCA) of the samples (Methods) the

first principal component (which accounts for 25% of variability) is highly correlated with estimated liver fraction (R = 0.93, p < 1x10^{15}. **Fig. 2E-F, S2D**).

Overall, these findings show that the predominant abnormality in AIH circulating DNA is an increase in hepatocyte contribution. The auto-immune nature of the disease suggests that this increase is due to immune attack on hepatocytes.

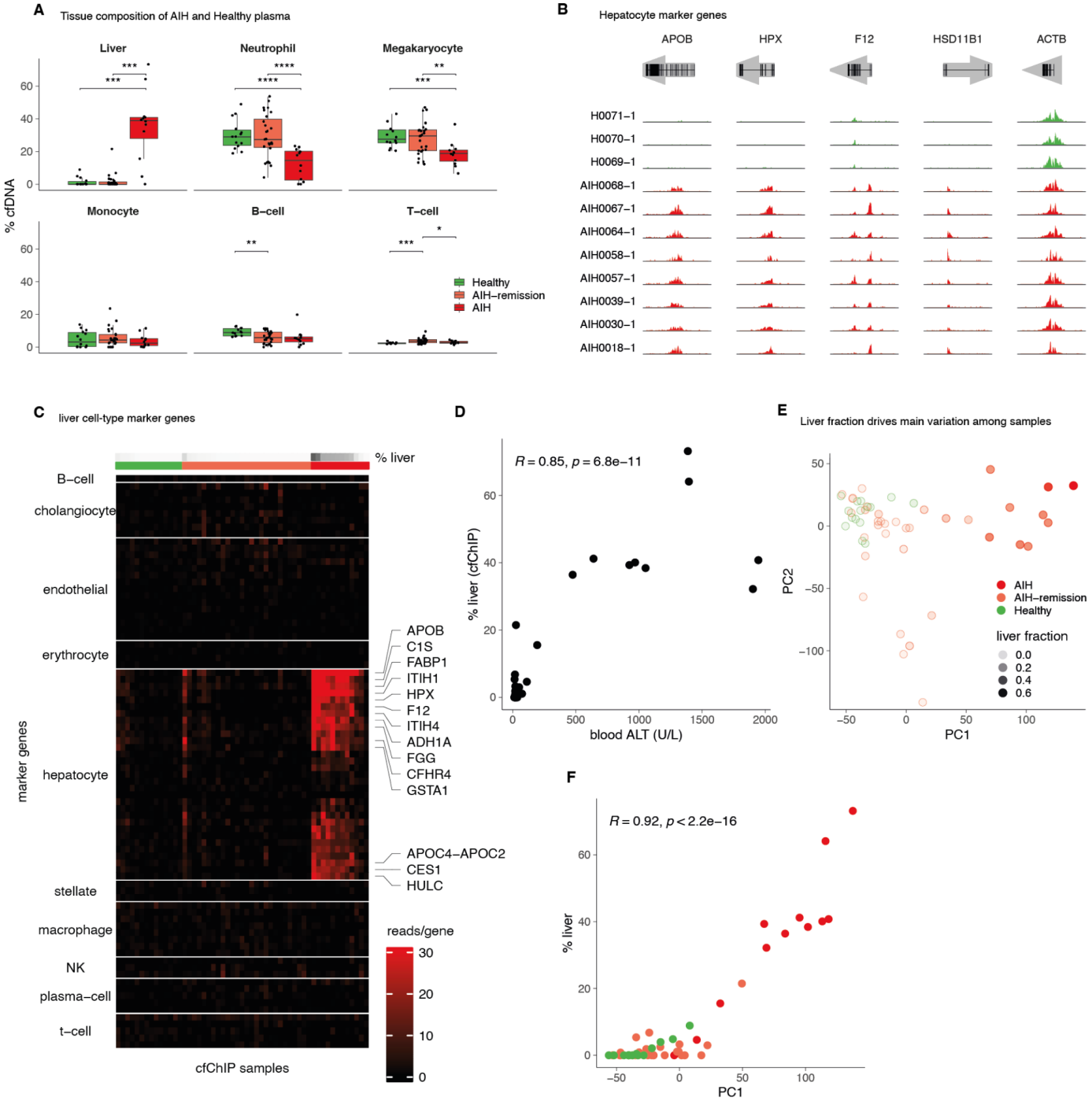

Figure 2: cfChIP-seq identifies hepatocyte as a major source of cfDNA in AIH plasma

A. Tissue composition median and distribution of healthy and AIH samples. y-axis values indicate relative contribution of tissue to

the circulation. Neutrophils, monocytes and megakaryocytes are observed in both groups, while liver has a pronounced contribution only in the AIH samples. */**/***: *t-test P* < 0.01/0.001/0.0001. Only significant comparisons are shown in figure.

B. Genome browser view of cfChIP-seq signal in hepatocytes marker genes (*APOB*, *HPX, F12, HSD11B1*) and *ACTB* as control. Green and red tracks represent healthy and AIH cfChIP-seq samples respectively.
C. Heatmap showing patterns of cfChIP-seq samples (columns) coverage across liver cell-type marker genes (rows). Color represents normalized coverage at the gene promoter. Of all liver cell type markers, a noticeable signal in the AIH samples is observed only in the hepatocyte marker genes.
D. Pearson correlation of cfChIP-seq liver fraction (y-axis) and time-matched blood ALT levels (x-axis).
E. Principal components 1 and 2 of AIH (red and orange) and healthy (green) samples. Transparency indicates liver fraction as in A. Principal components analysis was computed over all Refseq genes (~25,000).
F. Pearson correlation of PC1 and liver fraction displayed in D. Liver fraction differences across samples explains the largest variation observed in the data.

cfChIP-seq identifies hepatocyte immune response

Histone modifications are intimately related to the activity of RNA polymerase [17]. H3K4me3 in particular, is a histone modification associated with transcription initiation and transcriptional pause-release [7,18]. Thus levels of H3K4me3 are representative of the amount of such events in the cells that contribute to the circulating cfDNA pool. AIH is characterized by a complex process that involves activation of CD4+ effector and regulatory T-cells, cytokine and chemokine production and more (reviewed in [19]). We thus seek to explore whether cfChIP-seq can detect such processes in AIH plasma samples.

To distinguish changes within specific cell-types on the background of changes in cell-type composition, we used the following strategy: First, we used deconvolution to estimate cell-type composition of a sample. We then construct composition-informed reference for the specific sample taking into account the relative composition and the estimate of mean and variance gene levels in each cell type. Comparing this reference to the observed values we can identify genes that are significantly above or below the revised reference (**fig. 3A;** Methods).

Applying this model to the AIH samples we find that the vast majority of observed gene counts (99.98%) do not significantly deviate from the composition-informed reference (**fig. S3A)**. Focusing on the 774 genes with significantly elevated signal in the AIH samples compared to healthy, we find that here too the majority of genes (97%) do not significantly deviate from composition-informed reference suggesting that they reflect the normal transcription patterns of the liver (**fig. 3B**). However, a close inspection reveals a set of genes with coverage significantly above expected in several samples (**fig. 3C; fig. S3A**). This group includes the *CXCL9-11* (C-X-C motif chemokine ligand) genes, which are expressed in inflamed hepatocytes and play a role in the AIH immune response and liver fibrosis[20–23]. Importantly, many of these genes, which have high coverage in AIH samples, carry no H3K4me3 signal in normal liver nor in any of the other tissues represented in the

reference atlas, indicating that they reflect activation of an abnormal transcription program occurring in the patients with AIH (**fig. 3D-E; S3B**).

To rule out the possibility that these genes reflect a transcriptional program in other cell types other than the liver, we tested the correlation between the gene levels and the estimated fraction across all tissues composing the samples. This analysis revealed that these genes are positively correlated to the estimated liver fraction in the AIH samples and negatively correlated to all other tissues (**fig. S3C-D**). We conclude that the activity of these genes is strictly coupled to hepatocyte death, most likely reflecting transcription in the hepatocyte cells of patients with AIH.

Taken together, these results demonstrate that plasma H3K4me3 cfChIP-seq reliably identifies a hepatocyte immune process taking place in the AIH patients.

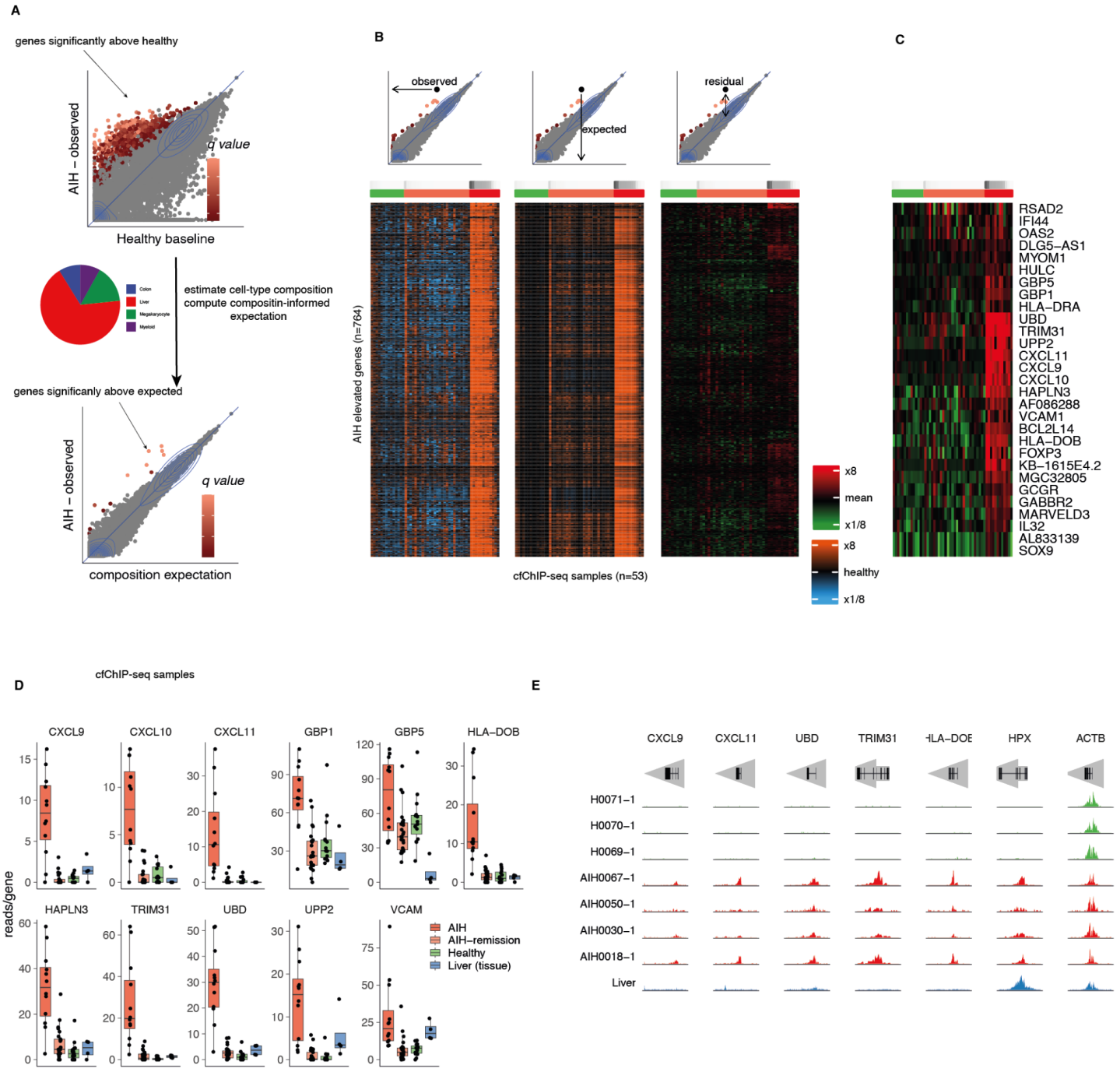

Figure 3: cfChIP-seq identifies hepatocyte immune response in AIH plasma

A. Concept of statistical model. AIH samples with genes significantly above healthy baseline (left) are deconvoluted to their composing cell types. The gene signals are then compared to the expected signal based on the sample's cell-types composition using reference data and tested whether they are significantly above expected given the mean and variance of the expected pattern.

B. Clustering of ~600 genes shown in figure 1D. Schematic representation of observed, expected and residual appears above heatmaps. Left, cfChIP-seq actual values relative to healthy mean. Middle, expected levels based on the samples compositions. Right, residual signal (log2(1+actual) - log2(1+expected)) of samples. genes (rows) and samples (columns) are clustered by residual heatmap and order is identical in the three plots.

C. Zoom in on genes that are significantly above expected in at least 3 AIH samples.

D. Median and distribution of normalized promoter coverage of selected genes shown in C in cfChIP-seq of AIH and healthy samples and ChIP-seq of liver tissue from reference atlas.

E. Genome browser of view of selected genes from C, with *HPX* as a liver specific control gene and *ACTB* as a general control gene.

Plasma based classifier for AIH diagnosis and monitoring

We next sought to test whether this realization can be utilized in the clinical setting in assisting the diagnosis and treatment management of patients with AIH.

The most prominent finding described so far indicates an elevation of liver derived cfDNA in patients with AIH compared to healthy control. Indeed, the single attribute of liver fraction suffices for discriminating between these two groups. In the clinical setting, however, the challenge is often differentiating AIH from other liver diseases and conditions that involve liver damage, such as drug induced liver injury or infections. To identify signals specific to AIH and to design a classifier aimed at distinguishing AIH from other diseases, we made use of previously published cfChIP-seq samples [6] (n= 18) with elevated liver derived cfDNA from patients with various diseases. We performed cfChIP-seq on two additional cohorts of adult (n=30) and pediatric (n=10) patients with various liver-related diseases. The adult cohort includes patients with nonalcoholic steatohepatitis (NASH), fatty liver, hepatitis B and C, drug induced liver injury, Cholestatic liver disease, primary biliary cholangitis and patients that underwent liver transplant. The pediatric cohort included patients that underwent liver biopsy due to elevated liver enzymes and were diagnosed with metabolic diseases, fatty liver, hypobetalipoproteinemia and with non-specific finding in the liver biopsies that were not compatible with AIH or any other disease.

As above, we neutralize the variable relative contribution of different tissues by computing residual signals — the differences between observed signal and the expected signal given the specific cell-type composition of the sample. Comparing the residual signal of the non-AIH and AIH samples over the group of genes that significantly deviate from the composition-informed reference described above, we find that 15 of the 29 genes are significantly elevated in the AIH group (*t*-test, $q < 0.1$ after false discovery rate (FDR) correction). Many of these genes lack signal completely in almost all non-AIH samples, supporting the role of these genes in an AIH unique immune response, as described in the previous section. Based on the differential genes, we define an ‘AIH score’ as the cumulative signal of the genes elevated in the AIH groups (**fig. 4A**). After computing this score of all samples, a clear distinction is apparent between the AIH and non-AIH samples (**fig. 4B**). Testing the effect of the liver derived cfDNA fraction on the ‘AIH score’ exhibits a linear relationship between the two in the AIH group. In the non-AIH group in contrast, this phenomenon is much less pronounced, reflecting the fact that this signature captures a transcription program typically inactive in the liver (**fig. 4C**). Finally, using a classifier based on the AIH-score, demonstrates the capability to accurately discriminate between the AIH and non-AIH plasma samples (AUC = 0.914; **fig. 4D**).

These results suggest that cfChIP-seq can fill an unmet need in assisting AIH diagnosis in a limited-invasive manner directly from plasma.

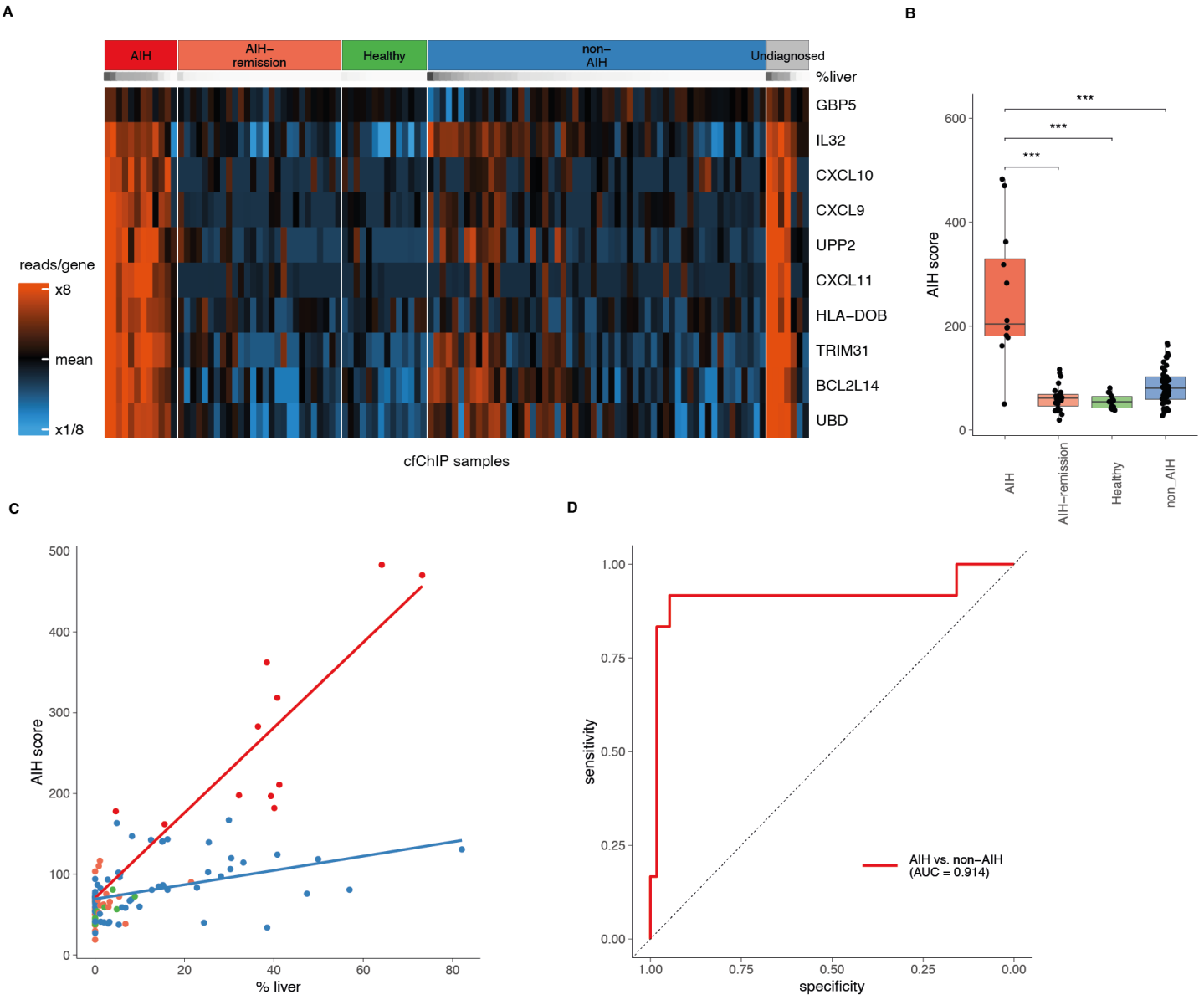

Figure 4: plasma based classifier for AIH diagnosis

A. Heatmap showing patterns of the cfChIP-seq coverage in genes that are significantly elevated in the AIH samples compared to the non-AIH samples (rows) on the various cfChIP-seq samples (columns). Columns were grouped by category (represented in the color bar) and ranked within the group by the relative liver contribution of every sample to the cfDNA. Color scale represents the normalized coverage of the genomic region after log-transformation (log_2(1+coverage)) and adjustment to zero mean for each gene across the samples. The undiagnosed group consists of samples collected from patients in the pediatric liver clinic that are currently under observation due to elevated liver enzymes, and are not finally diagnosed to date.

B. Median and distribution of the AIH-score across all samples. The score (y-axis) indicates the cumulative converge of the samples on the 'AIH score' genes shown in A.

C. AIH score (y-axis) as a function of liver percent (x-axis). cfChIP-seq samples (dots) are colored by group as in C and the red and blue lines indicate the linear fit of the AIH and non-AIH samples respectively. The diverging slopes demonstrate the ability of the AIH score to distinguish between AIH and non-AIH samples in plasma samples with liver contribution greater than 5%.

D. ROC curve of AIH vs. non-AIH classification.

Discussion

Autoimmune hepatitis (AIH) is a chronic disease that results in liver damage caused by autoantibodies. Despite progress in understanding the immune response mechanisms involved in this process, diagnosing AIH remains challenging, and traditionally requires invasive liver biopsy to obtain histological specimens[33,34]. Liquid biopsy is an emerging field of medical research and much effort has been made in the past two decades to develop assays that allow replacement of tissue biopsy with non-invasive liquid biopsy alternatives. However, most of the breakthroughs in the past few years are in the particular fields where changes in DNA sequence are available such as oncology and prenatal genetic screening[5,35]. More recent methods demonstrate the ability of cfDNA to report on abnormal tissue death also in somatic cells, but very little is known from liquid biopsies on specific transcriptional programs in the cells of origin, specifically in the context of AIH.

Here, we make use of plasma cfChIP-seq which reports on the promoter state of cell-free chromatin to reveal the cellular sources and transcription patterns of dying cells in AIH. We find that cfChIP-seq identifies elevated death of hepatocyte cells in patients with active disease. Moreover, applying a statistical model to explain away expected liver epigenetic landscape highlights abnormal transcription patterns taking place in the liver. Some of the genes with significantly elevated levels that were identified such as the C-X-C motif chemokine ligand family (*CXCL9/10/11*), have been associated with AIH response in hepatocytes and liver fibrosis[20,23]. Other genes are known to participate in inflammatory processes and liver related diseases, but were not reported in the AIH context to the best of our knowledge. These include the interferon induced Guanylate binding proteins 1/5 (*GBP1/5*) that induce liver injury and inflammation in different types of hepatitis and other liver diseases[27–29]; *UBD* (Ubiquitin D) and *TRIM31* (Tripartite Motif Containing 31) which are induced by pro inflammatory cytokines; *HLA-DOB* (Major histocompatibility complex, class II, Do Beta) which was identified as affecting the occurrence and development of hepatitis B (HBV)[24–26] and *HULC* (highly upregulated liver cancer) which is expressed in normal hepatocytes but strongly induced in hepatocellular carcinoma and HBV infection and involved in inflammatory injury in rats with cirrhosis[30–32]. An additional group of genes (e.g. *FOXP3, IL32)* exhibits coverage above expected not only in the AIH cohort but also in the cfChIP-seq samples of patients with other liver diseases and they presumably reflects a general stress response of hepatocytes which is not unique for AIH. Our control group is not intended for evaluating differential diagnosis, and includes a wide variety of liver-related conditions beyond those relevant for that task. Systematic identification of genes that are specific for AIH and devising a robust classifier based on them requires a larger and more

diverse control group. These examples, however, suggest the widespread applicability of cfChIP-seq for research of liver diseases, and as a potential method for liver liquid biopsy in the clinical setup and precision medicine.

Samples from AIH patients in remission show milder levels of liver-derived cfDNA, which is attributed to successful treatment. Importantly, in one case where there was discrepancy between liver enzyme levels and cfChIP-seq results, while liver enzymes were normal - both liver histology and cfChIP-seq results showed the patient had active disease. These results suggest that cfChIP-seq can be a valuable tool in monitoring the progression of the disease and optimizing treatment. However, this requires extensive longitudinal sampling, which is beyond the scope of the current study.

In summary, our results highlight the potential importance of cfChIP-seq in diagnosis and pathogenesis of liver disease, particularly AIH. Clearly, further studies are needed to establish and validate the clinical performance of cfChIP-seq in wider contexts. The untargeted nature of cfChIP-seq, which provides genome-wide transcription patterns of dying cells, can further serve researchers as a rich source of information to better understand the AIH pathogenesis, particularly cell death involved in this disease.

Methods

Patients

Plasma samples of patients and healthy controls were collected in the pediatric Gastroenterology institute at Shaare Zedek Medical Center (SZMC). The samples were taken from patients under various clinical conditions: (1) patients undergoing liver biopsy due to persistent elevation of liver enzymes to exclude AIH or other liver diseases or (2) to establish histological remission in patients with established AIH under treatment or (3) patients with established AIH under treatment with no adjacent liver biopsy. The control group comprises patients with normal liver biopsy or with no elevation of liver enzymes nor other liver disease. The study was approved by the Ethics Committees of the SZMC of Jerusalem (0269-19-SZMC). Informed consent was obtained from all individuals or their legal guardians before blood sampling.

Plasma cfChIP-seq

Immunoprecipitation, NGS library preparation, and sequencing

Sample collection and handling, Immunoprecipitation, library preparation and sequencing were performed by Senseera LTD. as previously reported[6], with certain modifications that increase capture and signal to background ratio. Briefly, ChIP antibodies were covalently immobilized to paramagnetic beads and incubated with plasma. Barcoded sequencing adaptors were ligated to chromatin fragments and DNA was isolated and next-generation sequenced.

Sequencing Analysis (assay yield and specificity)

Reads were aligned to the human genome (hg19) using bowtie2 (2.4.2) with 'no-mixed' and 'no-discordant' flags. We discarded fragments reads with low alignment scores (-q 2) and duplicate fragments.

Preprocessing of sequencing data was performed as previously described. Briefly, the human genome was segmented into windows representing TSS, flanking to TSS, and background (rest of the windows). The fragments covering each of these regions were quantified and used for further analysis. Non-specific fragments were estimated per sample and extracted resulting in the specific signal in every window. Counts were normalized and scaled to 1 million reads in healthy reference accounting for sequencing depth differences. Detailed information regarding these steps can be found at supplementary note in [6]. See **Supplementary Table 3** for full alignment statistics.

Statistical analysis

Differential genes compared to healthy

Statistical analysis of differential genes was performed as previously reported[6]. Briefly, for every gene in every sample we test whether the observed gene coverage is higher than expected according to the healthy mean/variance estimated from a control group of 26 self-reported healthy donors. Using the background rate of every sample and the scaling factor accounting for the sequencing depth, we define an expected distribution and estimate the probability of the observed coverage under the null hypothesis that the sample came from the healthy population. Genes with a FDR corrected P-value below 0.001 are reported as significantly elevated in the sample.

Cell type composition of samples (deconvolution)

To estimate the tissue composition of every sample, we used a non-negative least square model as implemented in the '*nnls'* R package (1.4). Given reference matrix $X^{\epsilon KxG}$ of the genes in K cell types and vector $Y^{\epsilon G}$ of observed gene counts in a sample, the objective is identifying non-negative coefficients $\widehat{\beta}$ (cell-type proportion) by solving $argmin_{\beta}||X^{T}\beta - Y||_{2}^{2}$ subject to $\beta_i \geq 0$ and $\sum_i \beta_i = 1$. For reference tissue atlas we used 182 samples from the Roadmap and Blueprint H3K4me3 ChIP-seq data. Estimated coefficients of similar cell-types were summed and the final composition across 36 distinct cell-types is shown. These results were reproducible when using different features for the regression and with other regression models. A full list of tissues and cell-types used as reference data can be found in supplementary table 5.

Principal component analysis (PCA) was performed on the AIH and healthy plasma cfChIP-seq Refseq gene counts as implemented in the 'prcomp' function of the R 'stats' package (4.2.2). Scree plot was generated using the 'fviz_eig' function of the R 'factoextra' package (1.0.7).

Liver single cell signatures

Identification of liver specific cell-types genes was done based on liver specific marker genes from published liver scRNA-seq data[14]. To increase the specificity of cell-type signature in the cfDNA context, we exclude genes with mean above 2 reads/progmoer in the healthy reference assuming their promoter is marked by H3K4me3 in non-liver cells contributing to the circulation. In addition we exclude genes where the 95 percentile in all non-liver tissues and cell types is above 50 reads/promoter. These filtering steps resulted in a reduced number of marker genes - particularly of the liver immune cells.

Expected, residual and unexplained genecounts

For every sample we define the expected gene counts to be the mean gene counts of the composing cell types weighted by the contribution fraction of the cell-type as described above $(X \cdot \widehat{\beta})$. To overcome misleading results due to missing tissues in the reference atlas, we added to the atlas an additional healthy profile derived from a large cohort of healthy cfChIP-seq samples.

The residual is defined as $log_2(1 + observed) - log_2(1 + expected)$. To further account for inter-tissue variability, we estimate the expected variance of every gene based on the weighted empirical variance observed in the replicates of the tissues composing the samples and test whether the null hypothesis that the observed counts are negative binomial distributed with that mean and variance can be rejected.

Formally, given a set of genes G, let:

$S_1, S_2 \ldots S_n$ - set of cfChIP-seq samples

$Y_{i,g}$ - coverage of gene g in samples i

$B_{i,g}$ - background reads in gene g of sample i

Q_i - normalization factor of sample i (sequencing rate)

$k = 1,\ldots K$ - set of reference cell-types

$X_{k,g}$ - coverage of gene g in cell-type k

$\hat{\beta}_i \in R^k_{\geq 0}$ - estimated fractions of cell-types composing sample i

$\mu_{k,g}, \sigma_{k,g}$ - mean and standard deviation of gene g in cell-type k, were $\mu_{k,g}, \sigma_{k,g}$ are estimated as previously described[6].

For every sample i the objective is to estimate the distribution:

$$p(Y_{i,g} | B_{i,g}, Q_i, \hat{\beta}_i, \mu_{k,g}, \sigma_{k,g}).$$

Due to discrete sampling in library preparation and sequencing we assume that $Y_{i,g}$ is Poisson distributed depending on the expected counts and sequencing depth.

$$Y_{i,g} \sim Poisson(\frac{1}{Q_i}\eta_{i,g} + B_{i,g}).$$

where $\eta = X \cdot \hat{\beta}$.

We approximate the distribution of $Y_{i,g}$ as negative binomial $NB(\mu_{i,g}, \sigma^2_{i,g})$. Using linearity of expectation and the law of total variation we can match the mean and variance of the negative binomial to that of the exact distribution:

$$\mu_{i,g} = \frac{1}{Q_i} E[\eta_{i,g}] + B_{i,g} = \frac{1}{Q_i} \sum_k \beta_{i,k} \mu_{k,g} + B_{i,g}$$

$$\sigma^2_{i,g} = \mu_{i,g} + \frac{1}{Q_i^2} Var[\eta_{i,g}] = \mu_{i,g} + \frac{1}{Q_i^2} \sum_k \beta^2_{i,k} \sigma^2_{k,g}$$

For every gene in every sample we compute the probability of $p_{NB}(x_{i,g} \geq Y_{i,g} | \mu_{i,g}, \sigma^2_{i,g})$. Unexplained genes were defined as genes where the FDR-corrected q-value was less than 0.001 in at least 3 AIH samples.

Acknowledgements

We thank the members of the Friedman lab for discussions and comments on this manuscript. This work was supported by the European Research Council's AdG Grant cfChIP 101019560 (to N.F.) and Israel Science Foundation IPMP Grant 3751/21 (to N.F and E.G.).

Data availability

All datasets used in this study are in the process of being deposited to public repositories.

Code availability

All script files used in the analysis in this manuscript will be available online at publication.

References

1. Anon. Autoimmune hepatitis: A comprehensive review. J Autoimmun 2013;41:126–139.
2. Mieli-Vergani G, Vergani D, Czaja AJ, et al. Autoimmune hepatitis. Nat Rev Dis Primers 2018;4:18017.
3. Liberal R, Longhi MS, Mieli-Vergani G, et al. Pathogenesis of autoimmune hepatitis. Best Pract Res Clin Gastroenterol 2011;25:653–664.
4. Feld JJ, Dinh H, Arenovich T, et al. Autoimmune hepatitis: effect of symptoms and cirrhosis on natural history and outcome. Hepatology 2005;42:53–62.
5. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic — implementation issues and future challenges. Nat Rev Clin Oncol 2021;18:297–312.

6. Sadeh R, Sharkia I, Fialkoff G, et al. ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. Nat Biotechnol 2021;39:586–598.

7. Karlić R, Chung H-R, Lasserre J, et al. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A 2010;107:2926–2931.

8. Weiner A, Hsieh T-HS, Appleboim A, et al. High-resolution chromatin dynamics during a yeast stress response. Mol Cell 2015;58:371–386.

9. Liu CL, Kaplan T, Kim M, et al. Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol 2005;3:e328.

10. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–330.

11. Anon. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell 2016;167:1145–1149.

12. Moss J, Magenheim J, Neiman D, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018;9:5068.

13. Loyfer N, Magenheim J, Peretz A, et al. A DNA methylation atlas of normal human cell types. Nature 2023;613:355–364.

14. MacParland SA, Liu JC, Ma X-Z, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun 2018;9:1–21.

15. Giannini EG, Testa R, Savarino V. Liver enzyme alteration: a guide for clinicians. CMAJ 2005;172:367–379.

16. Lo YM, Zhang J, Leung TN, et al. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet 1999;64:218–224.

17. Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell 2007;128:707–719.

18. Wang H, Fan Z, Shliaha PV, et al. H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. Nature 2023:1–10.

19. Webb GJ, Hirschfield GM, Krawitt EL, et al. Cellular and Molecular Mechanisms of Autoimmune Hepatitis. Annual Review of Pathology: Mechanisms of Disease 2018;13:247–292. Available at: http://dx.doi.org/10.1146/annurev-pathol-020117-043534.

20. Czaja AJ. Review article: chemokines as orchestrators of autoimmune hepatitis and potential therapeutic targets. Aliment Pharmacol Ther 2014;40:261–279.

21. Ikeda A, Aoki N, Kido M, et al. Progression of autoimmune hepatitis is mediated by IL-18-producing dendritic cells and hepatic CXCL9 expression in mice. Hepatology 2014;60:224–236.

22. Christen U, Hintermann E. Immunopathogenic Mechanisms of Autoimmune Hepatitis: How Much Do We Know from Animal Models? Int J Mol Sci 2016;17. Available at: http://dx.doi.org/10.3390/ijms17122007.

23. Zeremski M, Dimova R, Astemborski J, et al. CXCL9 and CXCL10 Chemokines as Predictors of Liver Fibrosis in a Cohort of Primarily African-American Injection Drug Users With Chronic Hepatitis C. J Infect Dis 2011;204:832–836.

24. Raasi S, Schmidtke G, Giuli R de, et al. A ubiquitin-like protein which is synergistically inducible by interferon-γ and tumor necrosis factor-α. European Journal of Immunology 1999;29:4030–4036. Available at: http://dx.doi.org/10.1002/(sici)1521-4141(199912)29:12<4030::aid-immu4030>3.0.co;2-y.

25. Li H, Liu S, Lin Y, et al. Identification and Verification of Ubiquitin D as a Gene Associated with Hepatitis C Virus-Induced Hepatocellular Carcinoma. INT 2022;65:195–205.

26. Jia X, Zhao C, Zhao W. Emerging Roles of MHC Class I Region-Encoded E3 Ubiquitin Ligases in Innate Immunity. Front Immunol 2021;12:687102.

27. Xiao J, Tipoe GL. Inflammasomes in non-alcoholic fatty liver disease. Front Biosci 2016;21:683–695.

28. Ding K, Li X, Ren X, et al. GBP5 promotes liver injury and inflammation by inducing hepatocyte apoptosis. FASEB J 2022;36:e22119.

29. Milks MW, Cripps JG, Lin H, et al. The role of Ifng in alterations in liver gene expression in a mouse model of fulminant autoimmune hepatitis. Liver Int 2009;29:1307–1315.

30. Matouk IJ, Abbasi I, Hochberg A, et al. Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis. Eur J Gastroenterol Hepatol 2009;21:688–692.

31. Panzitt K, Tschernatsch MMO, Guelly C, et al. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology 2007;132:330–342.

32. Zhu Y, Chen X, Zheng C, et al. Down-regulation of LncRNA UCA1 alleviates liver injury in rats with liver cirrhosis. Int J Clin Exp Pathol 2019;12:455–465.

33. Manns MP, Czaja AJ, Gorham JD, et al. Diagnosis and management of autoimmune hepatitis. Hepatology 2010;51:2193–2213.

34. Anon. International Autoimmune Hepatitis Group Report: review of criteria for diagnosis of autoimmune hepatitis. J Hepatol 1999;31:929–938.

35. Alix-Panabières C, Pantel K. Liquid Biopsy: From Discovery to Clinical Application. Cancer Discovery 2021;11:858–873. Available at: http://dx.doi.org/10.1158/2159-8290.cd-20-1311.

Supplementary information

identifying information statement

The encoded patient IDs (e.g., AIH0018, AIH0019, etc.) that appear in the supplementary tables are unknown to anyone outside the research group.

supplementary tables

- Supplementary Table 1 - Individuals and samples information
- Supplementary Table 2 - AIH clinical information
- Supplementary Table 3 - Sequencing statistics for samples sequenced in this study
- Supplementary Table 4 - genes with elevated signal in at least 3 AIH cfChIP-seq samples compared to healthy baseline (figures 1D and 3C)
- Supplementary Table 5 - reference tissues and cell types ChIP-seq used in this work. Reference data was achieved from the encode project, the Blueprint epigenome and the Roadmap epigenomics consortium.

  (https://www.encodeproject.org/, https://egg2.wustl.edu/roadmap/web_portal/, http://dcc.blueprint-epigenome.eu/#/home)

Supplementary figures

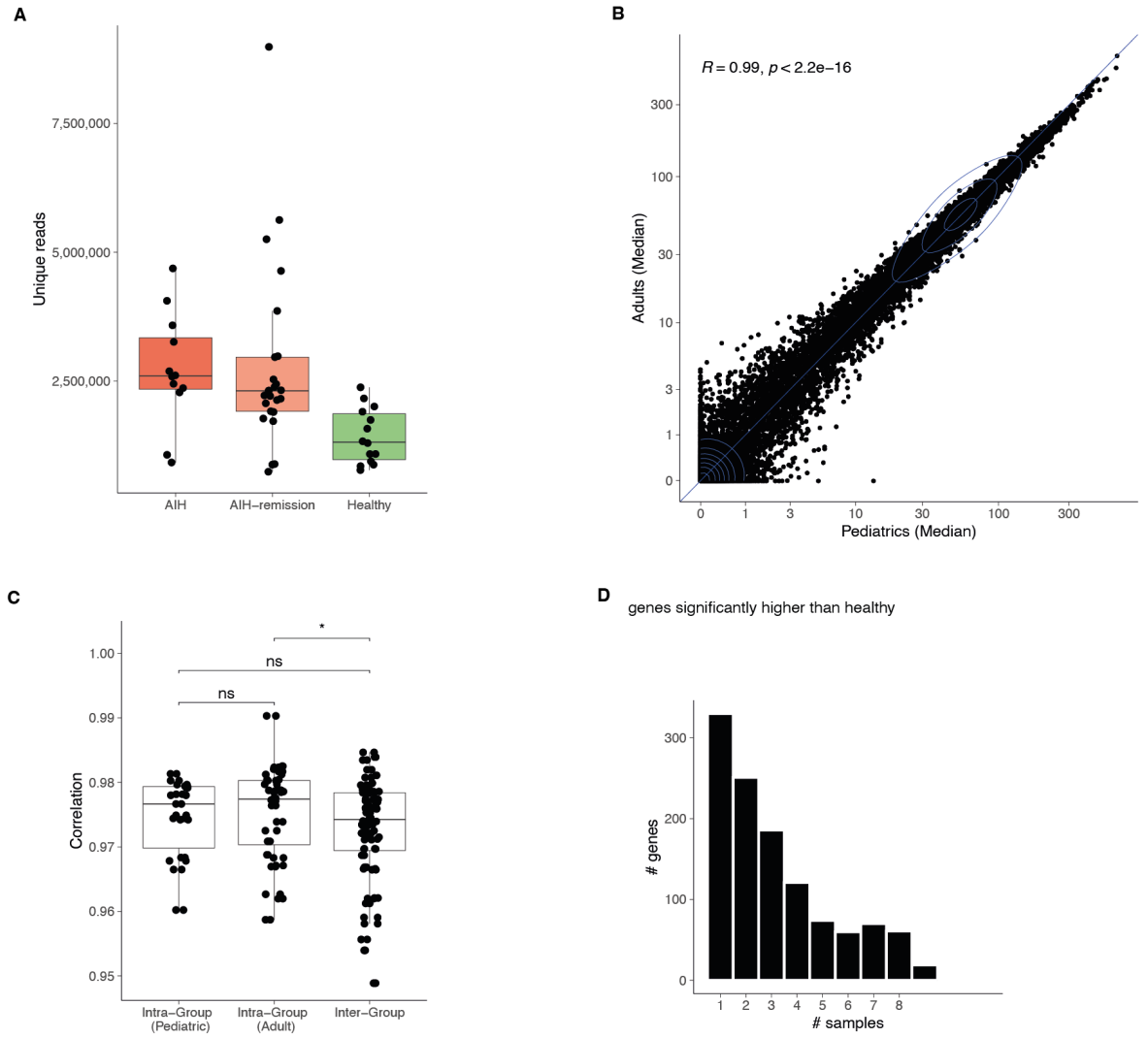

Figure S1: cfChIP-seq yield and healthy samples correlations

A. cfChIP-seq yield of AIH and healthy plasma samples. Values (y-axis) indicate the number of unique reads mapped to the genome (after duplicates removal).
B. Evaluation of similarity of plasma samples from healthy adults (y-axis) and healthy pediatrics (x-axis). Points represent normalized gene counts (reads/prompter) across the ~25,000 Refseq genes.
C. Gene counts pairwise Pearson correlation within healthy pediatric samples, within healthy adults samples and between adults and pediatrics healthy samples.
D. Frequency of genes with significantly elevated promoter coverage in AIH samples compared to healthy baseline shown in figure 1B.

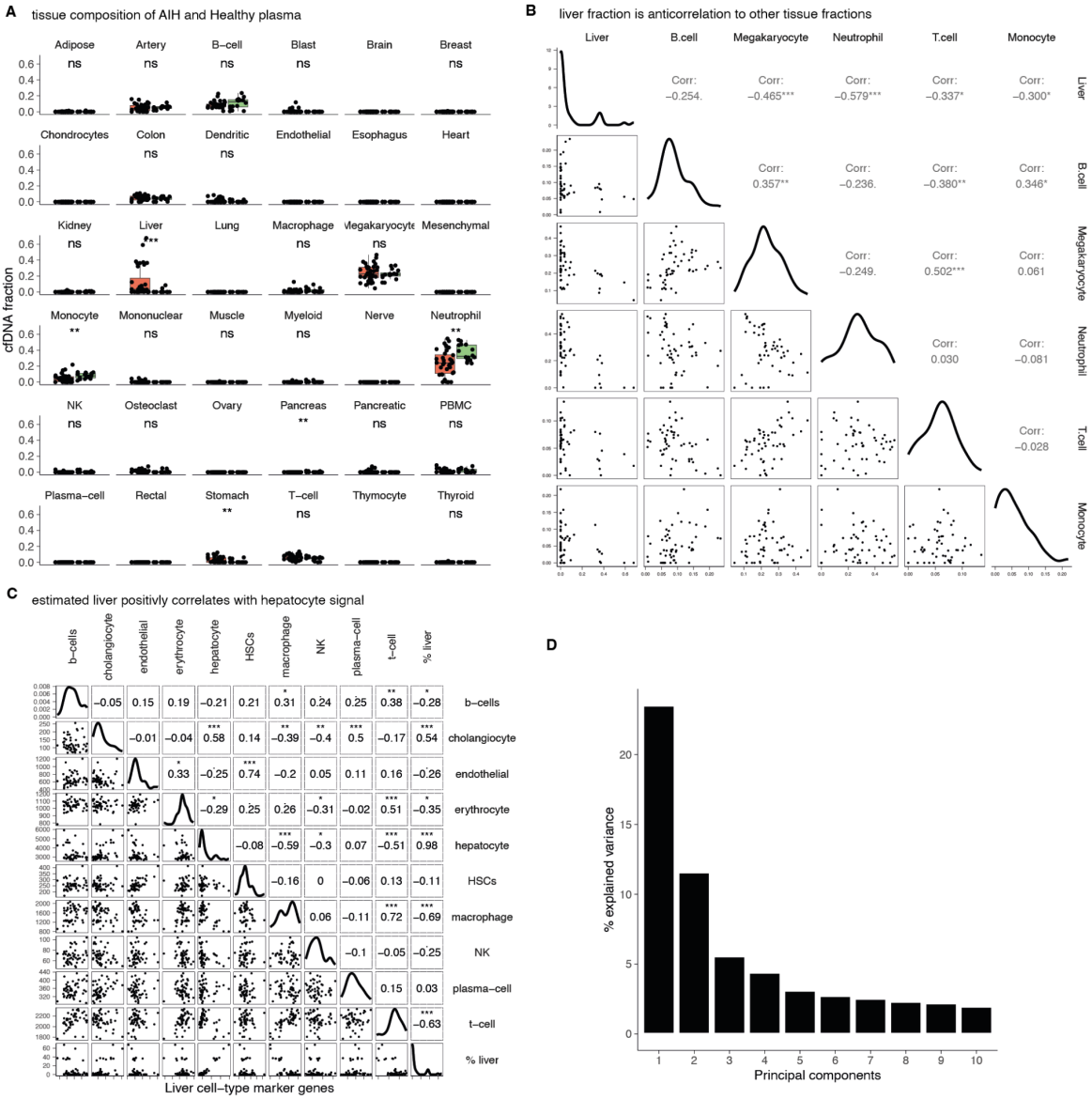

Figure S2: Tissue composition of AIH and Healthy plasma samples

A. Tissue composition median and distribution of healthy and AIH samples. Same as figure 2A for all tissues.
B. Pearson correlation of tissue fractions shown in A. liver fractions are negatively correlated with all other tissues, reflecting the relative nature of the tissue fraction.
C. Pearson correlation of estimated liver percent and cumulative signature of liver cell-type marker genes shown in 2C. Liver proportion is perfectly correlated with hepatocyte score, suggesting that the hepatocytes are the liver cell contributing to the circulation.
D. Scree plot of PCA plot shown in figure 2D.

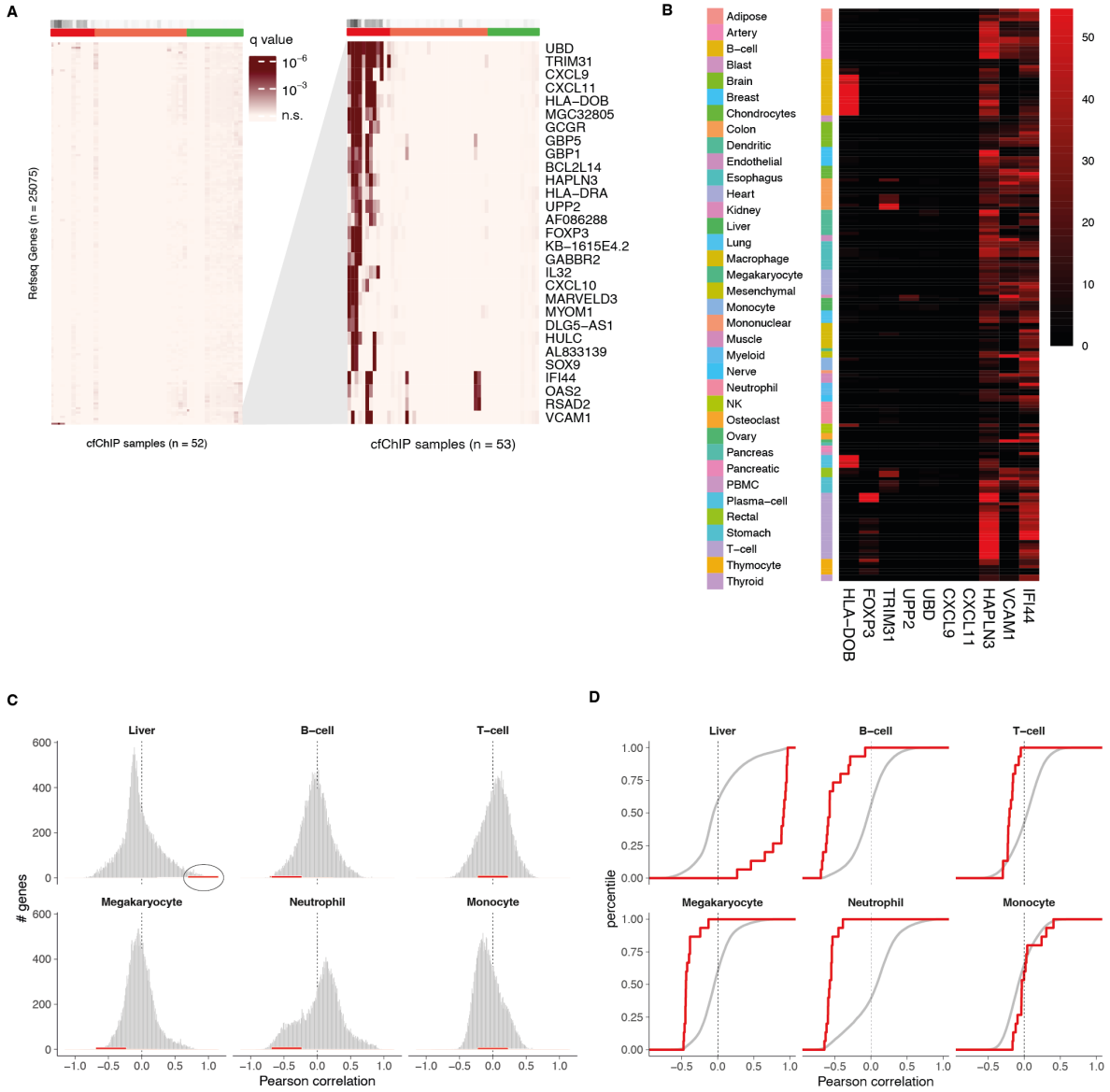

Figure S3: residual cf-ChIP-seq signal unexplained by tissue composition

A. Statistical significance of the observed signal given the composition informed expectation for all Refseq genes (rows) across healthy and AIH cfChIP-seq samples (columns). Color scale represents FDR corrected p-value (Methods).

B. Patterns of promoter coverage of reference tissues and cell types (rows) over genes with elevated signal unexplained by tissue composition (columns). Color represents normalized coverage over the promoter region of the genes.

C. Correlation of promoter coverage and estimated tissue fraction in the AIH samples. Gray background represents the correlation of all Refseq genes and the red histogram represents correlation of the 11 unexplained genes shown in A.

D. cumulative distribution function of correlation shown in C. black and red lines represent all Refseq genes and unexplained genes respectively.