Flexible and high-throughput simultaneous profiling of gene expression and chromatin accessibility in single cells

Volker Soltys[1,4*#], Moritz Peters[1], Dingwen Su[1], Marek Kučka[1,2], Yingguang Frank Chan[1,3#]

1 Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen, Germany
2 Department of Translational Genomics, University of Cologne, 50931 Cologne, Germany
3 University of Groningen, Groningen Institute of Evolutionary Life Sciences, 9747 AG Groningen, Netherlands
4 Current Address: Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

First author
* Corresponding authors
volker.soltys@eva.mpg.de; frank.chan@rug.nl

Abstract

Gene regulation consists of several highly interconnected layers including gene expression and chromatin accessibility. To understand cell-type specific regulatory circuits, both of these processes therefore need to be profiled simultaneously at single-cell level. However, these joint measurements are challenging and current methodologies suffer from limited applicability. Here, we present easySHARE-seq, an elaboration of SHARE-seq, for the simultaneous measurement of ATAC- and RNA-seq in single cells. By streamlining the barcode design, easySHARE-seq libraries have a usable sequence of up to 300bp (+200%), making it suitable for e.g. investigation of allele-specific signals or variant discovery. Furthermore, easySHARE-seq libraries do not require a dedicated sequencing run thus saving costs. Lastly, recovery of transcripts per cell is substantially boosted (> +50%) while retaining high scalability and low cost. To showcase these changes, we generated joint profiles for 19,664 murine liver nuclei, identified cell types, exploited the multi-omic measurements to link *cis*-regulatory elements to their target genes and investigated liver-specific micro-scale changes. easySHARE-seq produces high-quality multiomic data sets and is applicable to a wide range of study designs.

Introduction

Gene expression and chromatin state together influence fundamental processes such as gene regulation or cell fate decisions [1–3]. A better understanding of these mechanisms and their interactions will be a major step toward decoding developmental trajectories or reconstructing cellular taxonomies in both health and disease. Over the last decades, many technologies such as ChIP-seq, ATAC-seq, ChRO-seq [4–6] have emerged to profile different aspects of these processes and more recently, techniques such as single-cell (sc) ATAC-seq, scCUT&TAG or scChIP-seq profile many of these at the level of individual cells [7–9]. Perhaps the most widely used single-cell assays profile either the transcriptome (RNA-seq) or chromatin accessibility (ATAC-seq), which can resolve cell-type specific transcriptomic or epigenetic states in complex tissues [10,11,12]. Many such methodologies are able to profile hundreds of thousands of cells simultaneously by relying on combinatorial indexing for single-cell barcoding, where sequential rounds of barcodes combine to create unique cellular barcode combinations, a concept that theoretically can be scaled up to profile millions of cells at once [13,14]. Yet, measuring these modalities separately, even at this scale, does not fully capture their convoluted relationships and limits our ability to directly assess the complex molecular hierarchy between the transcriptome and epigenome or how genetic variation impacts cellular function and phenotype [15]. For example, prior studies have argued that chromatin state is often predictive of gene expression and can also prime cells toward certain lineage decisions or even induce tissue regeneration [16–18]. However, these studies depend on the computational integration of separately measured modalities. By assuming a shared biological state, this restricts the discovery of novel and potentially fine-scale differences and renders it challenging to identify the root cause of erroneous cell states[19].
To address these issues so-called 'multi-omic' techniques have recently become available, which capture two or more information layers simultaneously in the same cell [15]. Compared to single-modality assays, multi-omic technologies are therefore relatively new and thus still suffer from drawbacks. For example, commercially available solutions tend to be expensive and custom technologies limited by sensitivity such that multi-omic studies tend to have limited sample sizes [20–22]. However, few technologies such as SHARE-seq [23] both have high throughput and produce high-quality datasets. Yet, further technical limitations still prevent the widespread use of this and other technologies. For example, SHARE-seq is not only molecularly complex but aims to produce a singular sample with ultra-high cell numbers. In contrast though, studies frequently require many samples and replicates, highlighting the need for a protocol that takes these requirements into consideration, is less complex and less time consuming. Additionally, due to the highly unusual sequencing requirements, SHARE-seq has very limited applicability to a variety of study designs. For example, studies quantifying allele-specific signals or relying on genetic variants of any kind are difficult because only short lengths of the insert can be sequenced (up to 100 bp), leading to lower coverage of genetic variants and thus less power for these analyses. Furthermore, the unusual sequencing configuration requires a dedicated sequencing run for the libraries, adding substantial overall cost. To address these limitations we developed easySHARE-seq, a multi-omic single-cell (sc) RNA- and scATAC-seq protocol that allows for more widespread applicability, flexibility and ease of use making it suitable for a wide range of study designs.

Results

Technical improvements of easySHARE-seq

To develop easySHARE-seq, we implemented several improvements to SHARE-seq[23] while retaining the broad overall workflow (**Fig. 1A**). First, we streamlined the barcode design. We retain a combinatorial indexing scheme but shorten each barcode segment to 7 nucleotides (nt), connected by 3nt overhangs (hamming distance between barcode segments >= 2). This shortens the index sequences of the libraries to 17nt (from 99nt in SHARE-seq, Index1) and 8nt (Index2), respectively (**Fig 1B, Suppl. Fig. 1A** shows a direct comparison). By using two segments of 192 barcodes each, we achieve a total of 192*192*96 barcode combinations (~ 3.5 million) in the space of 25nt. The advantage of this design is that it allows sequencing of up to 300bp of the insert if desired (compared to 100nt in SHARE-seq or 90nt in Chromium Multiome libraries). These longer read lengths result in higher coverage of DNA variants, which is crucial in detecting allele-specific signals or cell-specific variant discovery, e.g., in F1 hybrids or cancer cells[24]. Furthermore, this design allows simultaneous sequencing of easySHARE-seq libraries with standard Illumina libraries ('multiplexing'), which can strongly reduce sequencing costs, especially in smaller pilot experiments. A second improvement is the shortening and optimisation of the bench protocol. The single-cell barcoding step is reduced by 3h (to 1.5h) due to replacing all hybridization steps with a ligation round as well as the lack of need for blocking oligos. Additionally, after barcoding cells are aliquoted into as many 'sub-libraries' of ~3.5k cells as desired. This results in higher flexibility for the fine-tuning of cell numbers, the cost-effective sequencing of e.g. a pilot sub-library to assess overall data quality or the storing of sub-libraries for potential later use. Lastly, fixation conditions and enzymatic reactions (e.g. Reverse Transcription) have been optimised resulting in higher recovery of transcripts per cell (see *Performance of easySHARE-seq in murine primary liver cells*). All in all, easySHARE-seq reagents cost ~0.056$/cell, which is similar to the original SHARE-seq protocol. A comparison, including cost, between the two protocols as well as the most commonly used commercial solution can be found in **Table 1**. For a detailed description of the flexibility of easySHARE-seq, instructions on how to modify and incorporate the framework into new designs, how to convert it to a single-channel assay as well as critical steps to assess when planning to use easySHARE-seq see *Supplementary Notes*.

easySHARE-seq reliably labels both transcriptome and accessible chromatin in individual cells

To evaluate the accuracy and cell-specificity of the barcoding, we first performed easySHARE-seq on a mixed pool between human and murine cell lines (HEK and OP-9, respectively). This allows the identification of two or more cells sharing the same barcode ('doublets'; **Fig. 1C**, left). After sequencing, we recovered a total of 3,808 cells. Both chromatin and transcriptome profiles separated well within each cell (**Fig. 1C**, middle), with cDNA showing a lower accuracy with increasing transcript counts, likely because of the more conserved cDNA sequences between human and mouse leading to less precise read mapping. We identified a total of 124 doublets (**Fig, 1C**, right), which results in a final doublet rate of 6.34% factoring in undetectable intra-species doublets. For comparison, a 10X Chromium Next GEM experiment with 10,000 cells has a doublet rate of ~7.9% (www.10xgenomics.com). Importantly, easySHARE-seq doublet rates can easily be lowered further by aliquoting fewer cells within each sub-library. To summarise, easySHARE-seq

provides a high-throughput and flexibility framework for accurately measuring chromatin accessibility and gene expression in single cells.

Performance of easySHARE-seq in murine primary liver cells

To assess data quality and overall performance of easySHARE-seq, we focused on murine liver. The liver consists of a diverse set of defined primary cell types, ranging from large and potentially multinucleated hepatocytes to small non-parenchymal cell types such as Liver Sinusoidal Endothelial Cells [25] (LSECs). We generated matched high-quality joint chromatin and gene expression profiles for 19,664 adult liver nuclei across four age-matched mice (2 male, 2 female), amounting to a recovery rate of 70.2% (28,000 input cells). Per nucleus, we recovered on average 3,629 unique transcripts (unique molecular identifiers; UMIs) and 2,213 unique fragments with 1,798 genes being expressed and 2,011 peaks accessible (74% of total RNA-seq reads were cDNA, the majority of which (69.6%) mapped to intronic regions, similarly to SHARE-seq nuclei performance; 55.9% mean ATAC-seq fragments in peaks; **Suppl. Fig. 1B-E;** how sequencing depth relates to UMI/fragment recovery see **Suppl. Fig. 1F & G,** an example data track can be found in **Suppl. Fig. 1L**). The scATAC-seq modalities displayed the characteristic banding pattern with reads being highly enriched at transcription start sites (TSS; **Suppl. Fig. 1H & I**). easySHARE-seq data was also highly reproducible between both sub-libraries and biological replicates, highlighting the accuracy of the technique (**Suppl. Fig. 1J & K**).

We next compared the performance of easySHARE-seq to other multi-omic and representative single channel assays. To do so, we made use of publicly available datasets (when possible we matched sample type) and downsampled them to a common sequencing depth (**Fig. 2A & B**, not all datasets provided all data needed for downsampling, see *Methods* for a detailed description and figure legend for tissue type and study). This showed that easySHARE-seq outperformed all other previously published multi-omic techniques in terms of recovery of UMIs per cell including SHARE-seq and performed similar to single-channel assays, likely due to improvements in efficiency of the enzymatic reactions or fixation (e.g. Reverse Transcription and Template Switching). Regarding recovery of DNA fragments in peaks per cell, easySHARE-seq performed similarly to other published multi-omic assays (**Fig. 2B**) but recovered less than SHARE-seq. This is likely because we chose to prioritise nuclei integrity and RNA-seq quality and therefore performed a stronger fixation (0.35% PFA compared to 0.2% in SHARE-seq), which typically leads to less efficient transposition reactions.

Simultaneous scATAC-seq and scRNA-seq profiling in murine primary liver cells

To visualise and identify cell types, we first projected the ATAC- and RNA-seq modalities separately into 2D Space and then used Weighted Nearest Neighbor[26] (WNN) integration to combine both modalities into a single UMAP visualisation (**Fig. 2C)**. Importantly, the same cells independently clustered together when clustering scRNA- and scATAC-seq separately (**Suppl. Fig. 2A&B**). We then annotated previously published cell types based on gene expression of previously established marker genes [27,28]. Marker gene expression was highly specific to the clusters (**Fig. 2D, Suppl. Fig. 2F**) and we recovered all expected cell types (**Suppl. Fig.2C**). Importantly, the same cell types were identified using each modality independently, showcasing high congruence between the scATAC- and scRNA-seq modalities (**Fig. 2E**). Altogether, our results show that easySHARE-seq generates high quality and reproducible joint cellular profiles of chromatin accessibility and gene expression within primary tissue, expanding our toolkit of multi-omic protocols.

Uncovering the *cis*-regulatory landscape of key regulators through peak-gene associations

The power of multiomic technologies lies in directly investigating the relationship between modalities, thereby allowing to further probe their complex relationships. For example, connecting a *cis*-regulatory element (CREs) to its target gene is difficult with separate measurements but it's an important step to fully understand how gene regulation functions and can change, both in health and disease. Measuring chromatin accessibility and gene expression simultaneously can be a potentially powerful and scalable approach to this issue.

To showcase the power of easySHARE-seq in doing so, we adopted the analytical framework from Ma et al. [23], which queries if an increased expression within a cell is significantly correlated with chromatin accessibility at a peak while controlling for GC content and accessibility strength. Focusing on LSECs (1,501 cells), a cell type with relatively low mRNA contents (**Suppl. Fig. 2D & E),** we calculated associations between putative CREs (pCREs, defined as peaks with a significant peak–gene association) and each expressed gene, considering all peaks within ± 500kb of the TSS. We identified 40,957 significant peak–gene associations (45% of total peaks, $P < 0.05$, FDR = 0.1) with 15,061 genes having at least one association (**Suppl. Fig. 3A,C**). In rare cases (2.9%), these pCREs were associated with five or more genes. This drops to 0.03% when considering only pCREs within ± 50 kb of a TSS (**Suppl. Fig. 3B,D**). These pCREs tended to cluster to regions of higher expressed gene density (2.15 mean expressed genes within 50kbp vs 0.93 for all global peaks) and their associated genes were enriched for biological processes such as mRNA processing, histone modifications and splicing (**Suppl. Fig. 3H**), possibly reflecting loci with increased regulatory activity.

We then ranked genes based on their number of associated pCREs (**Fig. 2F**). Within the top 1% of genes with the most pCRE associations were many key regulators and transcription factors. Examples include *Taf5,* which directly binds the TATA-box [29] and is required for initiation of transcription, or *Gata4,* which has been identified as the master regulator for LSEC specification during development as well as controlling regeneration and metabolic maturation of liver tissue in adult mice [30,31]. As such, it incorporates a variety of signals and its expression needs to be strictly regulated, which is reflected in its many pCREs associations (**Fig. 2H**). Similarly, *Igf1* also integrates signals from many different pCREs [32] (**Suppl. Fig. 3G**). Notably, pCRES are significantly enriched at transcription start sites (TSS), even relative to background enrichment (**Fig. 2G**). To summarise, easySHARE-seq has sufficient data quality to directly investigate the relationship between chromatin accessibility and gene expression and link putative *cis*-regulatory elements to their target genes at genomic scale, even in small cell types with relatively low mRNA contents.

***De Novo* identification of open chromatin regions and genes displaying zonation in LSECs**

Micro-scale changes in e.g. gene regulation and/or gene expression can often have major implications, for example during embryonic development or in certain diseases [33,34]. To demonstrate the ability of easySHARE-seq to capture and assess these micro-scale changes, we investigated the process of zonation in LSECs. The liver consists of hexagonal units called lobules where blood flows from the portal vein and arteries toward a central vein [35,36] (**Fig. 3A)**. The central–portal (CP) axis is characterised by a morphogen gradient, e.g. from *Wnt2*, secreted by central vein LSECs, with the resulting micro-environment giving rise to spatial division of labour among hepatocytes [37–39]. Studying zonation in non-parenchymal cells such as LSECs is challenging as these are small cells with low mRNA content (**Suppl.**

Fig. 2D,E), lying below the detection limit of current spatial transcriptomic techniques. As a result, only very few studies assess zonation in LSECs on a genomic level[40]. However, LSECs are critical to liver function as they line the artery walls, clear and process endotoxins, play a critical role in liver regeneration and secrete morphogens themselves to regulate hepatocyte gene expression [41–43], rendering their understanding a prerequisite for tackling many diseases.

We therefore asked if we can recover known zonation gradients and potentially identify novel marker genes and open chromatin regions displaying zonation using the resolution of easySHARE-seq and therefore ordered LSECs along pseudotime (**Fig. 3B**). To confirm that this faithfully recovers zonation gradients, we assessed gene expression and chromatin accessibility gradients for major known zonation marker genes along pseudotime [40] (**Fig. 3C,D**). For example, *Wnt2* expression decreased strongly along the CP axis as did chromatin accessibility of all three peaks at the *Wnt2* locus (**Fig. 3C**). We also recovered the expected zonation profiles for several more known pericentral (increasing along the CP-axis, e.g. *Rspo3* & *Kit*) and periportal (decrease along the CP-axis, e.g. *Efnb2* & *Meis1*) marker genes as well as chromatin accessibility gradients at their associated open chromatin regions (**Fig. 3D**). Knowing that the imputed pseudotime faithfully ordered cells along their zonation gradients, we sought to identify novel genes and open chromatin regions displaying zonation in LSECs based on the decrease or increase of the rolling mean in expression or accessibility along pseudotime. In total, we classified 179 genes and 198 open chromatin regions as pericentral and 371 genes and 546 open chromatin regions showed periportal zonation profiles (**Fig. 3D**). The list of markers contained many genes regulating epithelial growth and angiogenesis (e.g. *Efna1, Nrg2, Jag1, Bmp2, Bmper, Nrp2*) [44–47], related to regulating hepatocyte functions and communication (e.g. *Dpp4, Foxo1, Foxp1, Insr, Ephb4*) [48–51] as well as immunological functions (e.g. *Il1a, Pigr, Tgfb1)* [52,53], suggesting that these processes show variation along the PC axis. As dysregulation of LSEC zonation is implicated in multiple illnesses such as liver cirrhosis or non-alcoholic fatty liver disease [54,55], these genes are potential new biomarkers for their identification and the open chromatin regions starting points for investigating the role of gene regulation in their emergence. Ultimately, this highlights the sensitivity and applicability of easySHARE-seq to investigate micro-scale changes even in cell types with low overall expression levels.

Discussion

Understanding complex processes such as gene regulation or disease states requires the integration of multiple layers of information. easySHARE-seq is a high-quality, high-throughput and flexible platform for joint profiling of chromatin accessibility and gene expression within single cells. This technology is built upon SHARE-seq but has multiple important improvements. First, RNA-seq data quality is improved. More importantly, an improved barcode structure substantially reduces sequencing costs and results in longer insert sizes. This leads to increased coverage of DNA variants and adds immense flexibility to the protocol by removing hurdles for incorporating multi-omic single-cell assays into a variety of study designs and making their use more wide-spread. Combined with shorter experimental times (~12h total) and the inclusion of sub-libraries, easySHARE-seq might be particularly suited for studies where higher sample sizes are required or ones that rely on identification of genomic variants, e.g., in diverse, non-inbred individuals or in cancer. Yet in studies where a dataset with high cell numbers from a singular sample is needed, the original protocol might be more convenient. In terms of costs per cell, easySHARE-seq performs similarly to standard SHARE-seq with ~0.056 cents/cell, which is before factoring in the reduction in sequencing costs compared to standard SHARE-seq, and a fraction of the costs (<25%) of commercially available platforms without even considering their specialised instrument costs. A comparison between the two technologies as well as the most frequently used commercial alternative can be found in **Table 1**.

We showcase the power of easySHARE-seq by generating a dataset of 19,664 primary murine liver nuclei. We demonstrate that easySHARE-seq generates accurate and reproducible data, that both modalities can be used to identify the same cell types and how it compares to other technologies in terms of data quality. We then demonstrate the data quality and accuracy of easySHARE-seq in multiple ways. First, we leverage the simultaneous measurements to identify peak–gene interactions in LSECs which showed that key transcriptional regulators tend to have the most complex *cis*-regulatory landscape. Second, we assess zonation across both gene expression and chromatin accessibility. Even in a cell type with relatively low mRNA content such as LSECs, we recovered known zonation profiles of e.g., transcription factors, which are often expressed at a low level, and identified novel genes and open chromatin regions displaying zonation. Altogether, this shows that easySHARE-seq is not only applicable to a variety of study designs but also that the simultaneous measurements can be used to disentangle the complex molecular hierarchy and relationships between different regulatory layers. We envision easySHARE-seq as an important technological step toward the more widespread and diverse use of multiomic techniques. Ultimately, these techniques will play an important role in understanding gene regulation in health and disease, differentiation or lineage commitment, and determining genetic variants affecting those processes.

Methods

Animal Model & Tissue preparation

Mice

All animal experimental procedures were carried out under the licence number EB 01-21M at the Friedrich Miescher Laboratory of the Max Planck Society in Tübingen, Germany. The procedures were reviewed and approved by the Regierungspräsidium Tübingen, Germany. Liver was collected from both male and female wild-type C57BL/6 and PWD/PhJ mice aged between 9 to 11 weeks.

Study design

From each strain, we generated easySHARE-seq libraries for one male and one female mice (four total). For each individual, we sequenced two sub-libraries, resulting in 8 easySHARE-seq libraries.

Cell Culture

For the species-mixing experiment, HEK Cells were cultured in media containing DMEM/F-12 with GlutaMAX™ Supplement, 10% FBS and 1% Penicillin-Streptomycin (PenStrep) at 37°C and 5% CO_2. Cells were harvested on the day of the experiment by simply pipetting them off the plate and were then spun down for 5 min at 250G.
Murine OP9-DL4 cells were cultured in alpha-MEM medium containing 5% FBS and 1% PenStrep. On the day of the experiment, the cells were harvested by aspirating the media and adding 4 ml of Trypsin, followed by an incubation at 37°C for 5 min. Then, 5ml of media was added and cells were spun down for 5 min at 250G. After counting both cell lines using TrypanBlue and the Evos Countess II, equal cell numbers were mixed.

Liver Nuclei

The liver was extracted, rinsed in HBSS, cut into small pieces, frozen in liquid nitrogen and stored in the freezer at -80 °C for a maximum of two weeks. On the day of the experiment, 1 ml of ice cold Lysis Solution (0.1% Triton-X 100, 1mM DTT, 10mMM Tris-HCl pH8, 0.1mM EDTA, 3mM $Mg(Ac)_2$, 3mM $CaCl_2$ and 0.32M sucrose) was added to the tube. The cell suspension was transferred to a pre-cooled Douncer and dounced 10x using Pestle A (loose) and 15x using Pestle B (tight). The solution was added to a thick wall ultracentrifuge tube on ice and topped up with 4ml ice cold Lysis Solution. Then 9 ml of Sucrose solution (10mM Tris-HCl pH8.0, 3mM $Mg(Ac)_2$, 3mM DTT, 1.8M sucrose) was carefully pipetted to the bottom of the tube to create a sucrose cushion. Samples were spun in a pre-cooled ultracentrifuge with a SW-28 rotor at 24,400rpm for 1.5 hours at 4 °C. Afterwards, all supernatant was carefully aspirated so as not to dislodge the pellet at the bottom and 1 ml ice cold DEPC-treated water supplemented with 10µl SUPERase & 15µl Recombinant RNase Inhibitor was added. Without resuspending, the tube was kept on ice for 20 min. The pellet was then resuspended by pipetting ~15 times slowly up and down followed by a 40 µm cell straining step. Counting of the nuclei using DAPI and the Evos Countess II was immediately followed up by fixation.

easySHARE-seq protocol

Preparing the barcoding oligonucleotides

There are two barcoding rounds in easySHARE-seq with 192 unique barcodes distributed across two 96-well plates in each round (see **Suppl. Table 1** for a full list of oligonucleotide sequences). Each barcode (BC) is pre-annealed as a DNA duplex for improved stability. The first round of barcodes contains two single-stranded linker sequences at its ends as well as a 5' phosphate group to ligate the different barcodes together. The first single-stranded overhang links the barcode to a complementary overhang at the 5' end of the cDNA molecule or transposed DNA molecule, which originates either from the RT primer or the Tn5 adapter. The second overhang (3bp) is used to ligate it to the second round of barcodes (**Fig.1B**). Each duplex needs to be annealed prior to cellular barcoding, preferably on the day of the experiment. No blocking oligos are needed.

The Round1 BC plates contain 10µl of 4µM duplexes in each well and Round2 BC plates contain 10µl of 6µM barcode duplexes in each well, all in Annealing Buffer (10mM Tris pH8.0, 1mM EDTA, 30mM KCl). Pre-aliquoted barcoding plates can be stored at -20 °C for at least three months. On the day of the experiment, the oligo plates were thawed and annealed by heating plates to 95 °C for 2 min, followed by cooling down the plates to 20 °C at a rate of -2 °C per minute. Finally, the plates were spun down. Until the annealed barcoding plates are needed, they should be kept on ice or in the fridge.

This barcoding scheme is very flexible and currently supports a throughput of ~350,000 cells (assuming 96 indexing primers) per experiment, limited only by sequencing cost and availability of indexing primer. The barcodes were designed to have at least a Hamming distance of 2. See *Supplementary Notes* for further details on the barcoding system and flexibility.

Tn5 *preparation*

Tn5 was expressed in-house as previously described [57]. Two differently loaded Tn5 are needed for easySHARE-seq, one for the tagmentation, loaded with an adapter for attaching the first barcodes (termed Tn5-B2S), and one for library preparation with a standard illumina sequencing adapter (termed Tn5-A-only). See Supplementary Table 1 for all sequences.

To assemble Tn5-B2S, two DNA duplexes were annealed: 20 µM Tn5-A oligo with 22 µM Tn5-reverse and 20 µM Tn5-B2S with 22 µM Tn5-reverse, all in 50 mM NaCl and 10mM Tris pH8.0. Oligos were annealed by heating the solution to 95 °C for 30 s and cooling it down to 20 °C at a rate of 2 °C/min. An equal volume of duplexes was pooled and then 200 µl of unassembled Tn5 was mixed with 16.5 µl of duplex mix. The Tn5 was then incubated at 37 °C for 1 hour, followed by 4 °C overnight. The Tn5 can then be stored at -20 °C. In our hands, Tn5 did not show a decrease in activity after 10 months of storage.

To assemble Tn5-A-only, 10 µM of Tn5-A and 10.5 µM Tn5-reverse was annealed using the same conditions as described above. Again, 200 µl of unassembled Tn5 was mixed with 16.5 µl of Tn5-A duplex and incubated at 37 °C for 1 hour, followed by 4 °C overnight. The Tn5 can then be stored for later and repeated use for more than 10 months at -20 °C. We observed an increase in all Tn5 activity during the first months of storage, possibly due to continued transposome assembly in storage.

Fixation

One million liver nuclei ("cells" for short) were added to ice-cold PBS for 4 ml total. After mixing, 87 µl 16% formaldehyde solution (0.35%; for liver nuclei) or 25 µl 16% formaldehyde solution (0.1%; for HEK and OP9 cells) was added and the suspension was mixed by pipetting up and down exactly 3 times with a P1000 pipette set to 700 µl. The suspension was incubated at room temperature for 10 min. Fixation was stopped by adding ice-cold Stop-Mix (224 µl 2.5M glycine, 200 µl 1M Tris-HCl pH8.0, 53 µl 7.5% BSA in PBS). The suspension was mixed exactly 3 times with a P1000 pipette set to 850 µl and incubated on ice for 3 min followed by a centrifugation at 500G for 5 min at 4°C. Supernatant was removed and the pellet was resuspended in 1 ml Nuclei Isolation Buffer (NIB; 10mM Tris pH8.0, 10mM NaCl, 2mM $MgCl_2$, 0.1% NP-40) and kept on ice for 3 min followed by straining the suspension with a 40 µm cell strainer. It was then spun down at 500G for 3 min at 4°C and re-suspended in ~100-200µl PBSi (1x PBS + 0.4 U/µl Recombinant RNaseInhibitor, 0.04% BSA, 0.2 U/µl SUPERase, freshly added), depending on the amount of input cells. Cells were then counted using DAPI and the Countess II and concentration was adjusted to 2M cells/ml using PBSi.

Tagmentation

In a typical easySHARE-seq experiment for this study, 8 tagmentation reactions with 10,000 cells each followed by 3 RT reactions were performed. This results in sequencing libraries for around 30,000 cells. To increase throughput, simply increase the amount of tagmentation and RT reactions accordingly. No adjustment is needed to the barcoding. Each tube and PCR strip until the step of Reverse Crosslinking was coated before use by rinsing it with PBS+0.5% BSA to maximize cell recovery.

For each tagmentation reaction, 5 µl of 5X TAPS-Buffer, 0.25µl 10% Tween, 0.25µl 1% Digitonin, 3 µl PBS, 1 µl Recombinant RNaseInhibitor and 9µl of H2O was mixed. TAPS Buffer was made by first making a 1M TAPS stock solution in H_2O, followed by adjustment of the pH to 8.5 by titrating 10M NaOH. Then, 4.25ml H_2O, 500µl 1M TAPS pH8.5, 250µl 1M $MgCl_2$ and 5ml N-N-Di-Methyl-Formamide (DMF) was mixed on ice and in order. When adding DMF, the buffer heats up so it is important to be kept on ice. The resulting 5X TAPS-Buffer can then be stored at 4°C for short term use (1-2 months) or for long-term storage at -20°C (> 6 months).

Then, 5 µl of cell suspension at 2M cells/ml in PBSi was added to the tagmentation mix for each reaction, mixed thoroughly and finally 1.5µl of Tn5-B2S was added. The reaction was incubated on a shaker at 37°C for 30 min at 850 rpm. Afterwards, all reactions were pooled on ice into a pre-cooled 15ml tube. The reaction wells were washed with ~30 µl PBSi which was then added to the pooled suspension in order to maximize cell recovery. The suspension was then spun down at 500G for 3 min at 4°C. Supernatant was aspirated and the cells were washed with 200µl NIB followed by another centrifugation at 500G for 3 min at 4°C.

We only observed cell pellets when centrifuging after fixation and only when using cell lines as input material. Therefore, when aspirating supernatant at any step it is preferable to leave around 20-30µl liquid in the tube. Additionally, it is recommended to pipette gently at any step as to not damage and fracture the cells.

Reverse Transcription

As stated above, three tagmentation reactions were combined into one RT reaction. When increasing cells to more than 30,000 per RT reaction, we observed a steep drop in reaction

efficiency. The Master Mix for one RT reaction contained 3µl 100µM RT-primer, 2µl 10mM dNTPs, 6µl 5X MaximaH RT Buffer, 4.5µl 50% PEG6000, 1.5 µl H_2O, 1.5µl SUPERase and 1.66µl MaximaH RT. The RT primer contains a polyT tail, a 10bp UMI sequence, a biotin molecule and an adapter sequence used for ligating onto the first round of barcoding oligos.
The cell suspension was resuspended in 10µl NIB per RT reaction and added to the Master Mix for a total of 30µl. As PEG is present, it is necessary to pipette ~30 times up and down to ensure proper mixing. The RT reaction was performed in a PCR cycler with the following protocol: 52°C for 12 min; then 2 cycles of 8°C for 12s, 15°C for 45s, 20°C for 45s, 30°C for 30s, 42°C for 2min and 50°C for 3 min. Finally, the reaction was incubated at 52°C for 5 more minutes. All reactions were then pooled on ice into a pre-cooled and coated 15ml tube and the reaction wells were washed with ~40µl NIB, which was then added to the pooled cell suspension in order to maximise cell recovery. The suspension was then spun down at 500G for 3 min at 4°C. Supernatant was aspirated and the cells were washed in 150µl NIB and spun down again at 500G for 3min at 4°C. This washing step was repeated once more, followed by resuspension of the cells in 2ml Ligation Mix (400µl 10x T4-Buffer, 40µl 10% Tween-20, 1460µl Annealing Buffer and 100µl T4 DNA Ligase, added last).

Single-cell barcoding

Using a P20 pipette, 10µl of cell suspension in the ligation mix was added to each well of the two annealed Round1 BC plates, taking care as to not touch the liquid at the bottom of each well. The plates were then sealed, shaken gently by hand and quickly spun down (~ 8s) followed by an incubation on a shaker at 25°C for 30 min at 350 rpm. After 30 min, the cells from each well were pooled into a coated PCR strip using a P200 multichannel pipette set to 30µl. In order to pool, each row was pipetted up and down three times before adding the liquid to a PCR strip. After 8 columns were pooled into the strip, the suspension was transferred into a coated 5ml tube on ice. This process was repeated until both plates were pooled, taking care to aspirate most liquid from the plates. The cell suspension was then spun down for 3min at 500G at 4°C. Supernatant was aspirated and the cells were resuspended thoroughly in 2 ml new Ligation Mix. Now, 10µl of cell suspension was added into each well of the annealed Round2 barcoding plates using a P20 pipette, taking care as to not touch the liquid within each well. The plates were sealed, shaken gently by hand and spun down quickly followed by incubating them on a shaker at 25°C for 45 min at 350 rpm. The cells were then pooled again using the above-described procedure into a new coated 15ml tube. The cells were spun down at 500G for 3 min at 4°C. Supernatant was aspirated, the cells were washed with 150µl NIB and spun down again. Finally, the cells were resuspended in ~60µl NIB and counted. For counting, 5µl of cells were mixed with 5µl of NIB and 1x DAPI and counted on the Evos Countess II, taking the dilution into account. Sub-libraries of 3,500 cells were made and the volume was adjusted to 25µl by addition of NIB.
Using 3,500 cells results in a doublet rate of ~6.3%. The recovery rate of cells after sequencing depends on the input material (and QC thresholds), with cell lines recovering around 80% of input cells (~2,800-3,000 cells) and liver nuclei around 70% (~2,300-2,500 cells).

Reverse-Crosslinking

To each sub-library of 3,500 cells, 30µl 2x Reverse Crosslinking (RC) Buffer (0.4% SDS, 100mM NaCl, 100mM Tris pH8.0) as well as 5µl ProteinaseK was added. The sub-libraries were mixed and incubated on a shaker at 62°C for one hour at 800 rpm. Afterwards, they

were transferred to a PCR cycler into a deep well module set to 62°C (lid to 80°C) for an additional hour. Afterwards, each sub-library was incubated at 80°C for 10 min and finally 5µl of 10% Tween-20 to quench the SDS and 35µl of NIB was added for a total volume of 100µl. The lysates can be stored at this point at -20°C for at least two days, which greatly simplifies handling many sub-libraries at once. Longer storage has not been extensively tested.

Streptavidin Pull-Down

Each transcript contains a biotin molecule as the RT primers are biotinylated which is used to separate the scATAC-seq libraries from the scRNA-seq libraries. For each sublibrary, 50µl M280 Streptavidin beads were washed three times with 100µl B&W Buffer (5mM Tris pH8.0, 1M NaCl, 0.5mM EDTA) supplemented with 0.05% Tween-20, using a magnetic stand. Afterwards, the beads were resuspended in 100µl 2x B&W Buffer and added to the sublibrary, which were then shaken at 25°C for one hour at 900 rpm. Now all cDNA molecules are attached to the beads whereas transposed molecules are within the supernatant. The lysate was put on a magnetic stand to separate supernatant and beads. It likely is possible to reduce the number of M280 beads in this step, significantly reducing overall costs. However, this has not been extensively tested.

scATAC-seq library preparation

The supernatant from each sub-library was cleaned up with a Qiagen MinElute Kit and eluted twice into 30µl 10mM Tris pH8.0 total. PCR Mix containing 10µl 5X Q5 Reaction Buffer, 1µl 10mM dNTPs, 2µl 10µM i7-TruSeq-long primer, 2µl 10µM Nextera N5XX Indexing primer, 4.5µl H2O and 0.5µl Q5 Polymerase was added (All Oligo sequences in **Suppl. Table 1**). Importantly, in order to distinguish the samples, each sub-library needs to be indexed with a different N5XX Indexing primer. The fragments were amplified with the following protocol: 72°C for 6 min, 98°C for 1 min, then cycles of 98°C for 10s, 66°C for 20s and 72°C for 45s followed by a final incubation at 72°C for 2 min. The number of PCR cycles strongly depends on input material (Liver: 17 PCR cycles, Cell Lines: 15 PCR cycles). The reactions were then cleaned up with custom size selection beads with 0.55X as upper cutoff and 1.4X as lower cutoff and eluted into 25µl 10mM Tris pH8.0. Libraries were quantified using the Qubit HS dsDNA Quantification Kit and run on the Agilent 2100 bioanalyzer with a High Sensitivity DNA Kit.

cDNA library preparation

The beads containing the cDNA molecules were washed three times with 200µl B&W Buffer supplemented with 0.05% Tween-20 before being resuspended in 100µl 10mM Tris ph8.0 and transferred into a new PCR strip. The strip was put on a magnet and the supernatant was aspirated. The beads were then resuspended in 50µl Template Switch Reaction Mix: 10µl 5X MaximaH RT Buffer, 2µl 100µM TS-oligo, 5µl 10mM dNTPs, 3µl Enzymatics RNaseIn, 15µl 50% PEG6000, 14µl H2O and 1.25µl MaximaH RT. The sample was mixed well and incubated at 25°C for 30 min followed by an incubation at 42°C for 90 min. The beads were then washed with 100µl 10mM Tris while the strip was on a magnet and resuspended in 60µl H_2O. To each well, 40µl PCR Mix was added containing 20µl 5X Q5 Reaction Buffer, 4µl 10µM i7-Tru-Seq-long primer, 4µl 10µM Nextera N5XX Indexing primer, 2µl 10mM dNTPs, 9µl H2O and 2µl Q5 Polymerase. The resulting mix can be split into two 50µl PCR reactions or run in one 100µl reaction. The PCR involved initial incubation at 98°C for 1 min followed by PCR cycles of 98°C for 10s, 66°C for 20s and 72°C for 3 min with a final incubation at 72°C for 5 min. Importantly, in order to distinguish the samples, each

sub-library needs to be indexed with a different N5XX Indexing primer. The number of PCR cycles strongly depends on input material (Liver: 15 cycles, Cell lines: 13 cycles).
The PCR reactions were cleaned up with custom size selection beads using 0.7X as a lower cutoff (70µl) and eluted into 25µl 10mM Tris pH8.0. The cDNA libraries were quantified using the Qubit HS dsDNA Quantification Kit.

scRNA-seq library preparation

As the cDNA molecules are too long for sequencing (mean length > 700bp), they need to be shortened on one side. To achieve this, 25ng of each cDNA library was transferred to a new strip and volume was adjusted to 20µl using H_2O. Then 5µl 5X TAPS Buffer and 0.8µl Tn5-A-only was added and the sample was incubated at 55°C for 10 min. To stop the reaction, 25µl 1% SDS was added followed by another incubation at 55°C for 10 min. The sample was then cleaned up with custom size selection beads using a ratio of 1.3X and eluted into 30µl. Then 20µl PCR mix was added containing 10µl 5X Q5 reaction buffer, 1µl 10mM dNTPs, 2µl 10µM i7-Tru-Seq-long primer, 2µl 10µM Nextera N5XX Indexing primer (note: each sample needs to receive the **same** index primer as was used in the cDNA library preparation), 4.5µl H_2O and 0.5µl Q5 Polymerase. The PCR reaction was carried out with the following protocol: 72°C for 6 min, 98°C for 1 min, followed by 5 cycles of 98°C for 10s, 66°C for 20s and 72°C for 45s with a final incubation at 72°C for 2 min. Libraries were purified using custom size selection beads with a ratio of 0.5X as an upper cutoff and 0.8X as a lower cutoff. The final scRNA-seq libraries were quantified using the Qubit HS dsDNA Quantification Kit and run on the Agilent 2100 bioanalyzer with a High Sensitivity DNA Kit.

Sequencing

Both scATAC-seq and scRNA-seq libraries were sequenced simultaneously as they were indexed with different Index2 indices (N5XX). All libraries were sequenced on the Nova-seq 6000 platform (Illumina) using S4 2x150bp v1.5 kits (Read 1: 150 cycles, Index 1: 17 cycles, Index 2: 8 cycles, Read 2: 150 cycles). Libraries were partly multiplexed with standard Illumina sequencing libraries.

Custom Size selection beads

To make custom size selection beads, we washed 1ml of SpeedBeads on a magnetic stand in 1ml of 10mM Tris-HCl pH8.0 and re-suspended them in 50ml Bead Buffer (9g PEG8000, 7.3g NaCl, 500ul 1M Tris HCl pH8.0, 100ul 0.5M EDTA, add water to 50ml). The beads don't differ in their functionality from other commercially available ready-to-use size selection beads. They can be stored at 4°C for > 3 months.

Analysis

Gene annotations and Genomic variants

The reference genome and the Ensembl gene annotation of the C57BL/6J genome (mm10) were downloaded from Ensembl (Version GRCm38, release 102). Gene annotations for PWD/PhJ mice were downloaded from Ensembl. A consensus gene annotation set in mm10 coordinates was constructed by filtering for genes present in both gene annotations.

easySHARE-RNA-seq pre-processing

Fastq files were demultiplexed using a custom C-script, allowing one mismatch within each barcode segment. The reads were trimmed using cutadapt[58]. UMIs were then extracted from bases 1-10 in Read 2 using UMI-Tools [59] and added to the read name. Only reads with TTTTT at the bases 11-15 of Read 2 were kept (> 96%), allowing one mismatch. Lastly, the barcode was also moved to the read name.

Species-Mixing Experiments

RNA-seq reads were aligned to a composite hg38-mm10 genome using STAR [60]. The resulting bamfile was then filtered for uniquely mapping reads and reads mapping to chrM, chrY or unmapped scaffolds or containing unplaced barcodes were removed. Finally, the reads were deduplicated using UMItools [59]. ATAC-seq reads were also aligned to a composite genome using bwa [61]. Duplicates were removed with Picard tools and reads mapping to chrM, chrY or unmapped scaffolds were filtered out. Additionally, reads that were improperly paired or had an alignment quality < 30 were also removed.

The reads were then split depending on which genome they mapped to and reads per barcode were counted. Barcodes needed to be associated with at least 700 fragments and 500 UMIs in order to be considered a cell for the analysis. A barcode was considered a doublet when either the proportion of UMIs or fragments assigned to a species was less than 75%. This cutoff was chosen to mitigate possible mapping bias within the data.

easySHARE-RNA-seq processing and read alignment

We only used Read 1 for all our RNA-seq analyses as sequencing quality tends to drop after a polyT tail is sequenced in R2. This effect though can be largely mitigated bby reducing the fraction of RNA-seq libraries on a sequencing lane (<30%). Each sample was mapped to mm10 using the twopass mode in STAR [60] with the parameters --outFilterMultimapNmax 20 --outFilterMismatchNmax 15. We then processed the bamfiles further by moving the UMI and barcode from the read name to a bam flag, filtering out multimapping reads and reads without a definitive barcode. To determine if a read overlapped a transcript, we used featureCounts from the subread package [62]. UMI-Tools was used to collapse the UMIs of aligned reads, allowing for one mismatch and de-duplication of the reads. Finally, (single-cell) count matrices were created also using UMI-Tools.

easySHARE-ATAC-seq pre-processing and read alignment

Fastq files were demultiplexed using a custom C-script, allowing one mismatch within each barcode segment. The paired reads were trimmed using cutadapt [58] and the resulting reads were mapped to the mm10 genome using bwa mem [61]. Reads with alignment quality < Q30, unmapped, undetermined barcode, or mapped to mtDNA were discarded. Duplicates were removed using Picard tools. Open chromatin regions were called by subsampling the bamfiles from all samples to a common depth, merging them into a pooled bamfile and using the peak caller MACS2 [63] with the parameters -nomodel -keep-dup -min-length 100. The count matrices as well as the FRiP score was generated using featureCounts from the Subread package [62] together with the tissue-specific peak set.

Filtering, Integration & Dimensional reduction of scRNAseq data

The count matrices were loaded into Seurat [64] and cells were then filtered for >200 detected genes, >500 UMIs and < 20.000 UMIs. The sub-libraries coming from the same experiment were then merged together and normalised. Merged experiments from the same species

(one from male mouse, one from female mouse) were then integrated by first using SCTransform [65] to normalise the data, then finding common features between the two experiments using FindIntegrationAnchors() and finally integrated using IntegrateData(). Lastly, the integrated datasets from C57BL/6 and PWD/PhJ were again integrated using IntegrateData(). To visualise the data, we projected the cells into 2D space by UMAP using the first 30 principal components and identified clusters using FindClusters().

Filtering, Integration & Dimensional reduction of scATACseq data

Fragments per cell were counted using sinto and the resulting fragment file was loaded into Signac [66] alongside the count matrices and the peakset. We calculated basic QC statistics using base Signac and cells were then filtered for a FRiP score of at least 0.3, > 300 fragments, < 15.000 fragments, a TSS enrichment > 2 and a nucleosome signal < 4. Again, sub-libraries coming from the same experiment were merged. We then integrated all four experiments (C57BL/6 & PWD/PhJ, one male & one female mouse each) by finding common features across datasets using FindIntegrationAnchors() using PCs 2:30 and then integrating the data using IntegrateEmbeddings(). To visualise the data, we projected the cells into 2D space by UMAP.

Weighted-Nearest-Neigbor (WNN) Analysis & Cell type identification

In order to use data from both modalities simultaneously, we created a multimodal Seurat object and used WNN [26] clustering to visualise and leverage both modalities for downstream analysis. Afterwards, we assigned cell cycle scores and excluded clusters consisting of nuclei solely in the G2M-phase (2 clusters, 121 nuclei total). Cell types were assigned via expression of previously known marker genes, which allows subsetting the data into cell types.

Calculating Peak–Gene Associations

Peak–gene associations were calculated following the framework described by Ma et al [23]. In short, Spearman correlation was calculated for every peak–gene pair within a +-500kb window around the TSS of the expressed gene. To obtain a background estimation, we used chromVAR [67] (*getBackgroundPeaks*()) to generate 100 background peaks matched in GC bias and chromatin accessibility but randomly distributed throughout the genome. We calculated the Spearman correlation between every background-gene comparison, resulting in a null distribution with known population mean and standard deviation. We then calculated the z-score for the peak–gene pair in question ((correlation - population mean)/ standard deviation) and used a one-sided z-test to determine the p-value. This functionality is also implemented in Signac under the function *LinkPeaks*(). Increasing the number of background peaks to 200, 350 or 500 for each peak–gene pair does not impact the results (*data not shown*).

Analysis of LSEC zonation markers

To analyse gene expression and chromatin accessibility along LSEC zonation, we ordered LSECs along pseudotime in Monocle3 [68]. Gene expression and Chromatin Accessibility for marker genes was smoothed over pseudotime with local polynomial regression fitting (loess). To identify novel marker genes, we excluded genes with low expression, divided the pseudotime into 10 bins and calculated the moving average (for three bins) across them. We then required the moving average to continuously decrease (for pericentral marker genes) or increase (for periportal marker genes), allowing two exceptions. Lastly, we divided the

means for each gene by their maximum to normalise the values. Identification of *cis*-regulatory elements displaying zonation effects had equal requirements.

Gene Ontology Analysis

Gene Ontology Analysis was done using the R package clusterProfiler [69] with standard parameters.

Comparison to external datasets

In order to compare the performance of easySHARE-seq to other multiomic and single-channel technologies, we downloaded the raw data from Martin et al.[13] (sciRNA-seq3; GEO: GSE186824, first four samples, 9.974 murine E16.5 embryonic cells), Chen et al.[21] (SNARE-seq, GEO: GSE126074, first 4 replicates, 2.621 adult mouse brain cortex cells), Cao et al.[20] (sciCAR, GEO:GSE117089, 5968 murine kidney cells), Cusanovich et al.[11] (sciATAC-seq, GEO: GSE111586, 7.023 murine liver nuclei) Nikopolou et al.[56] (10x scATAC-seq, E-MTAB: E-MTAB-12706, 7.810 murine liver nuclei). We downsampled all datasets to a common sequencing depth (22.000 reads/cell for the RNA-seq, 34.000 reads/cell for the ATAC-seq). To determine fragments in peaks per cell, we used the peaksets provided by the authors. Su et al.[27] (10x 3' expression, 82.168 murine liver cells) did not provide all raw data needed for downsampling and we therefore used the authors provided count matrix. They report a sequencing depth of ~ 180.000 reads / cell. Unfortunately, Ma et al.[23] (SHARE-seq, 42.948 murine skin cells) did neither report full raw data, code nor a sequencing depth. We therefore used the authors count matrix as provided.

Data Availability

The easySHARE-seq data reported in this paper can be downloaded with the accession number GSE256434. All code used in data analysis is available at https://github.com/vosoltys/easySHARE_seq.git.

Acknowledgements

We thank members of the Chan and Jones lab for helpful discussions and critical reading of the manuscript. We are very grateful to Arnar Breevoort and Alex Pollen for sharing tissue preparation protocols and a very helpful research visit. We thank Sinja Mattes and all animal care takers at the Friedrich Miescher Laboratory for their work. We also thank the Genome Center in the Max Planck Institute for Biology Tübingen for providing support. The OP9-DL4 cells were a kind gift from Juan Carlos Zúñiga-Pflücker. M.P. is supported by an International Max Planck Research School fellowship. M.K. and Y.F.C. were supported European Research Council Starting Grant 639096 “HybridMiX" and Proof-of-Concept Grant 101069216 “Haplotagging”. The research was supported by the Max Planck Society.

Author Contributions

V.S. and Y.F.C. designed the experiments. V.S. and M.P. developed the barcoding framework for easySHAREseq. V.S. developed the rest of the protocol and performed experiments. V.S. performed the computational analyses advised by Y.F.C. V.S. drafted the manuscript. M.P., D.S., M.K. and Y.F.C. helped with experimental or computational support. All authors reviewed the manuscript.

Declaration of Interest

The authors declare no competing interests.

References

1. Anderson, E., Devenney, P. S., Hill, R. E. & Lettice, L. A. Mapping the Shh long-range regulatory domain. *Development* **141**, 3934–3943 (2014).
2. Nord, A. S. *et al.* Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell* **155**, 1521–1531 (2013).
3. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).
4. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213–1218 (2013).
5. Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669–680 (2009).
6. Chu, T. *et al.* Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat Genet* **50**, 1553–1564 (2018).
7. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060–1066 (2019).
8. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**, 825–835 (2021).
9. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
10. Tikhonova, A. N. *et al.* The bone marrow microenvironment at single-cell resolution. *Nature* **569**, 222–228 (2019).
11. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e18 (2018).
12. Kashima, Y. *et al.* Single-cell sequencing techniques from individual to multiomics analyses. *Exp Mol Med* **52**, 1419–1427 (2020).
13. Martin, B. K. *et al.* Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat Protoc* **18**, 188–207 (2023).
14. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
15. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494–515 (2023).
16. Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548.e16 (2018).
17. Zhang, C., Macchi, F., Magnani, E. & Sadler, K. C. Chromatin states shaped by an epigenetic code confer regenerative potential to the mouse liver. *Nat Commun* **12**, 4110 (2021).
18. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
19. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
20. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
21. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* **37**, 1452–1457 (2019).

22. Pourmorady, A. D. *et al.* RNA-mediated symmetry breaking enables singular olfactory receptor choice. *Nature* **625**, 181–188 (2024).
23. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116.e20 (2020).
24. Lähnemann, D. *et al.* Eleven grand challenges in single-cell data science. *Genome Biology* **21**, 31 (2020).
25. Aizarani, N. *et al.* A Human Liver Cell Atlas reveals Heterogeneity and Epithelial Progenitors. *Nature* **572**, 199–204 (2019).
26. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
27. Su, Q. *et al.* Single-cell RNA transcriptome landscape of hepatocytes and non-parenchymal cells in healthy and NAFLD mouse liver. *iScience* **24**, 103233 (2021).
28. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107.e17 (2018).
29. Chen, X. *et al.* Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, eaba8490 (2021).
30. Winkler, M. *et al.* Endothelial GATA4 controls liver fibrosis and regeneration by preventing a pathogenic switch in angiocrine signaling. *J Hepatol* **74**, 380–393 (2021).
31. Géraud, C. *et al.* GATA4-dependent organ-specific endothelial differentiation controls liver development and embryonic hematopoiesis. *J Clin Invest* **127**, 1099–1114.
32. Lara-Diaz, V. *et al.* IGF-1 modulates gene expression of proteins involved in inflammation, cytoskeleton, and liver architecture. *J Physiol Biochem* **73**, 245–258 (2017).
33. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer During Snake Evolution. *Cell* **167**, 633-642.e11 (2016).
34. Horn, S. *et al.* TERT Promoter Mutations in Familial and Sporadic Melanoma. *Science* **339**, 959–961 (2013).
35. Baratta, J. L. *et al.* Cellular Organization of Normal Mouse Liver: A Histological, Quantitative Immunocytochemical, and Fine Structural Analysis. *Histochem Cell Biol* **131**, 713–726 (2009).
36. Jungermann, K. & Kietzmann, T. Zonation of parenchymal and nonparenchymal metabolism in liver. *Annu Rev Nutr* **16**, 179–203 (1996).
37. Braeuning, A. *et al.* Differential gene expression in periportal and perivenous mouse hepatocytes. *The FEBS Journal* **273**, 5051–5061 (2006).
38. Planas-Paz, L. *et al.* The RSPO–LGR4/5–ZNRF3/RNF43 module controls liver zonation and size. *Nat Cell Biol* **18**, 467–479 (2016).
39. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2+ cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
40. Halpern, K. B. *et al.* Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nat Biotechnol* **36**, 962–970 (2018).
41. Knolle, P. A. & Wohlleber, D. Immunological functions of liver sinusoidal endothelial cells. *Cell Mol Immunol* **13**, 347–353 (2016).
42. Smedsrød, B. Clearance function of scavenger endothelial cells. *Comparative Hepatology* **3**, S22 (2004).
43. Rafii, S., Butler, J. M. & Ding, B.-S. Angiocrine functions of organ-specific endothelial cells. *Nature* **529**, 316–325 (2016).
44. Theilmann, A. L. *et al.* Endothelial BMPR2 Loss Drives a Proliferative Response to BMP (Bone Morphogenetic Protein) 9 via Prolonged Canonical Signaling. *Arteriosclerosis,*

Thrombosis, and Vascular Biology **40**, 2605–2618 (2020).
45. Russell, K. S., Stern, D. F., Polverini, P. J. & Bender, J. R. Neuregulin activation of ErbB receptors in vascular endothelium leads to angiogenesis. *American Journal of Physiology-Heart and Circulatory Physiology* **277**, H2205–H2211 (1999).
46. Vihanto, M. M. *et al.* Hypoxia up-regulates expression of Eph receptors and ephrins in mouse skin. *FASEB J* **19**, 1689–1691 (2005).
47. Duan, J.-L. *et al.* Endothelial Notch activation reshapes the angiocrine of sinusoidal endothelia to aggravate liver fibrosis and blunt regeneration in mice. *Hepatology (Baltimore, Md.)* **68**, 677 (2018).
48. Dong, X. C. *et al.* Inactivation of Hepatic Foxo1 by Insulin Signaling Is Required for Adaptive Nutrient Homeostasis and Endocrine Growth Regulation. *Cell Metabolism* **8**, 65–76 (2008).
49. Das, A. *et al.* Ephrin B2/EphB4 pathway in hepatic stellate cells stimulates Erk-dependent VEGF production and sinusoidal endothelial cell recruitment. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **298**, G908–G915 (2010).
50. Ghorpade, D. S. *et al.* Hepatocyte-secreted DPP4 in obesity promotes adipose inflammation and insulin resistance. *Nature* **555**, 673–677 (2018).
51. Zou, Y. *et al.* Forkhead Box P1 (FOXP1) Transcription Factor Regulates Hepatic Glucose Homeostasis. *J Biol Chem* **290**, 30607–30615 (2015).
52. Dewidar, B., Meyer, C., Dooley, S. & Meindl-Beinker, N. TGF-β in Hepatic Stellate Cell Activation and Liver Fibrogenesis—Updated 2019. *Cells* **8**, 1419 (2019).
53. Barbier, L. *et al.* Interleukin-1 Family Cytokines: Keystones in Liver Inflammatory Diseases. *Front Immunol* **10**, 2014 (2019).
54. Miyao, M. *et al.* Pivotal role of liver sinusoidal endothelial cells in NAFLD/NASH progression. *Laboratory Investigation* **95**, 1130–1144 (2015).
55. Su, T. *et al.* Single-Cell Transcriptomics Reveals Zone-Specific Alterations of Liver Sinusoidal Endothelial Cells in Cirrhosis. *Cellular and Molecular Gastroenterology and Hepatology* **11**, 1139–1161 (2021).
56. Nikopoulou, C. *et al.* Spatial and single-cell profiling of the metabolome, transcriptome and epigenome of the aging mouse liver. *Nat Aging* **3**, 1430–1445 (2023).
57. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
59. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
60. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
62. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
63. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
64. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).

65. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296 (2019).
66. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333–1341 (2021).
67. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975–978 (2017).
68. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381–386 (2014).
69. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).

Table 1

Table 1: Comparison between multiomic single-cell technologies

Protocol	easySHARE-seq	SHARE-seq	10x Multiome
Cost / Cell	0.056$	0.043$	0.26$
Additional costs?	None	Dedicated sequencing run	Specialised Instrument
Throughput (cells)	> 200,000	> 200,000	80,000
Max. sequencing length of insert	300bp	100bp	100bp
Processing time	~12h	~17h	~14.5h
Convertible to single-channel assay?	Yes	Not described	No

Figure 1

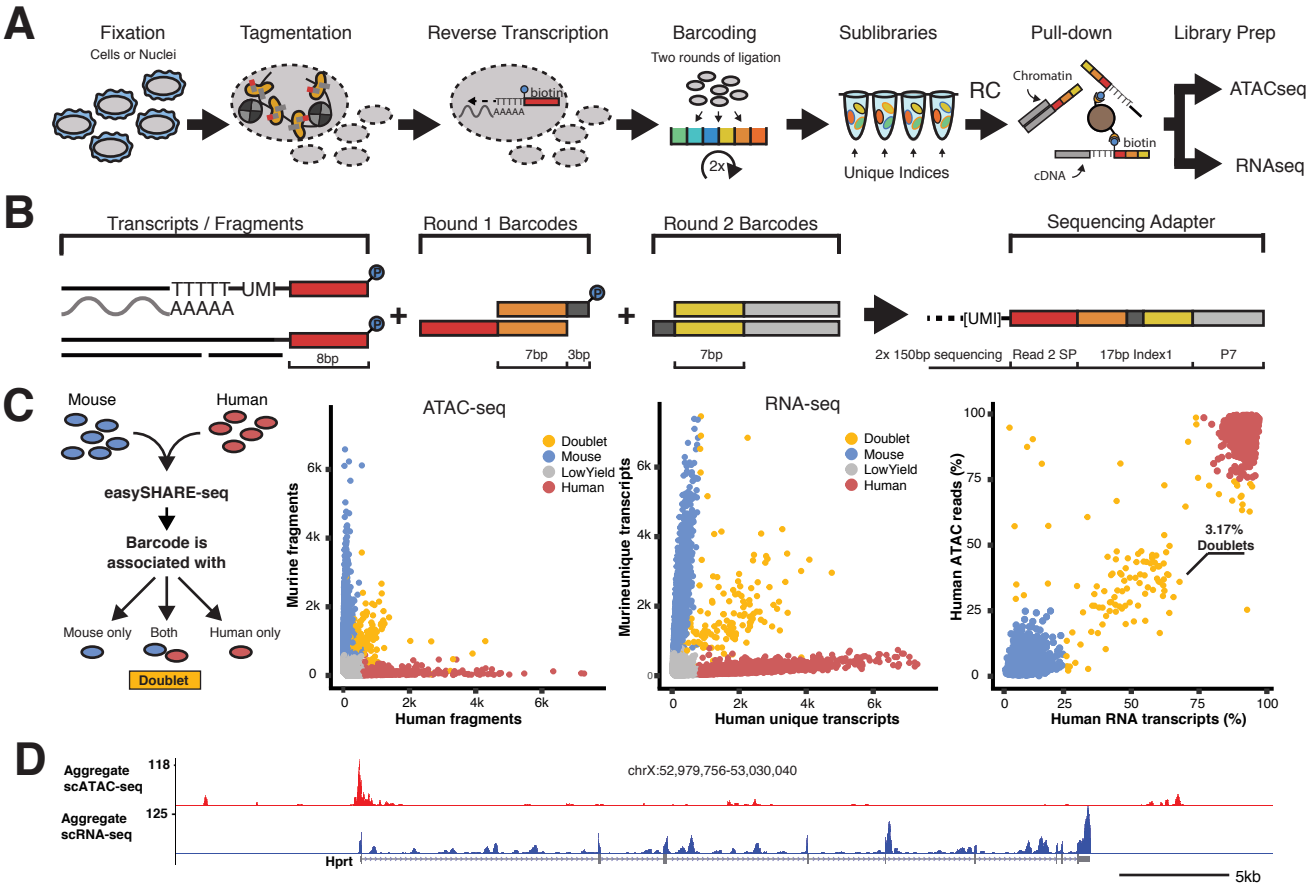
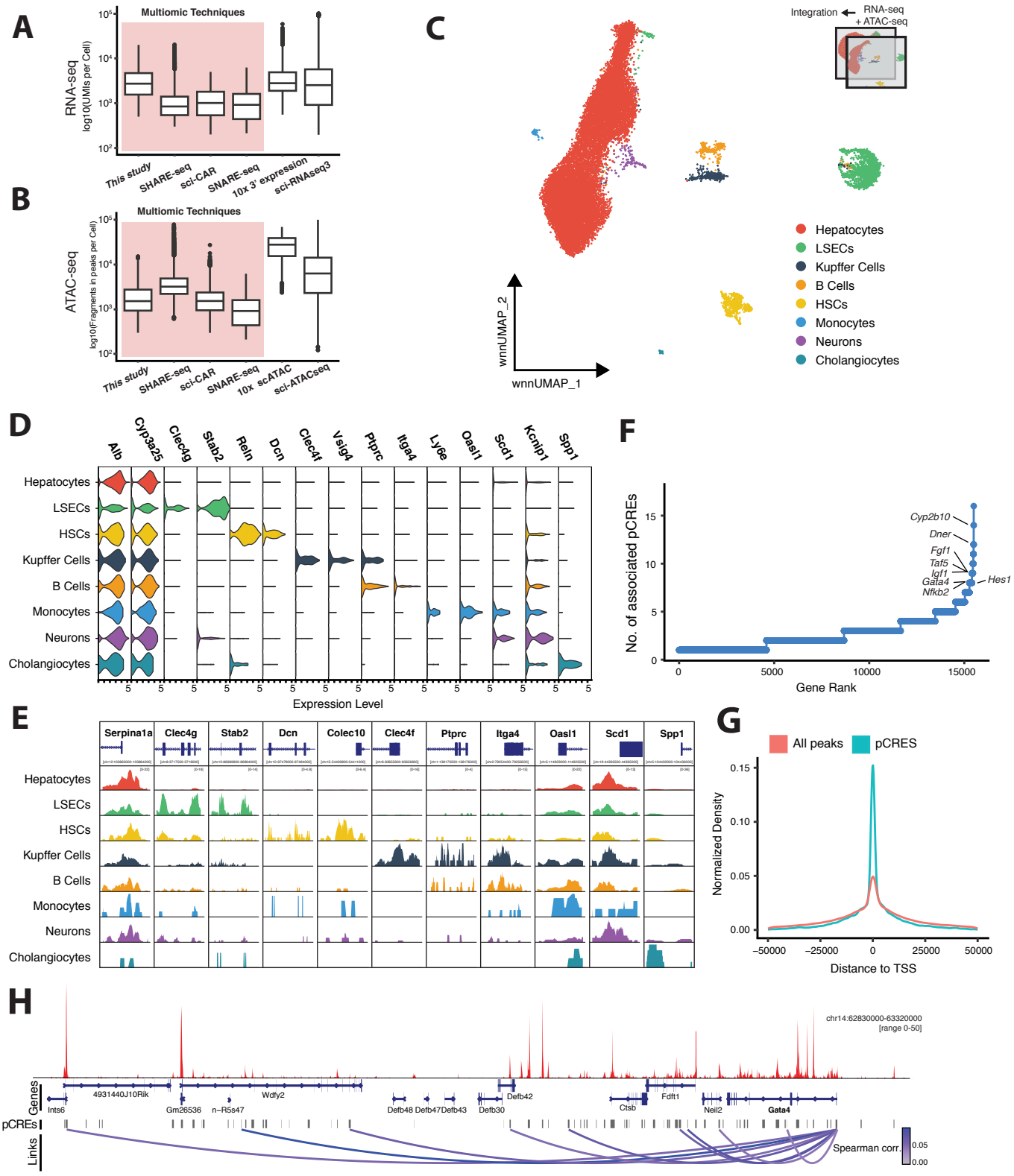


Figure 1: easySHARE-seq enables highly-accurate simultaneous scATAC-seq and scRNA-seq profiling

(A) Schematic workflow of easySHARE-seq.
(B) Generation and structure of the cell-specific barcode within Index 1.
(C) Principle of a species-mixing experiment. Cells are mixed prior to easySHARE-seq and sequences associated with each cell barcode are assessed for genome of origin (left panel). Unique ATAC fragments per cell aligning to the mouse or human genome (middle left). Cells are coloured according to their assigned origin (red: human; blue: mouse; orange: doublet). Middle right: Same plot but with RNA UMIs. Right: Percentage of ATAC fragments or RNA UMIs per cell relative to total sequencing reads mapping uniquely to the human genome. 3.17% of all observed cells classified as doublets. Accounting for same-species doublets, this results in a doublet rate of 6.34%.
(D) Aggregate chromatin accessibility (red) and expression-seq (blue) profile of OP-9 cells at the *Hprt* locus.

Figure 2



A
Multiomic Techniques
RNA-seq
log10(UMIs per Cell)
This study
SHARE-seq
sci-CAR
SNARE-seq
10x 3' expression
sci-RNAseq3
B
Multiomic Techniques
ATAC-seq
log10(Fragments in peaks per Cell)
This study
SHARE-seq
sci-CAR
SNARE-seq
10x scATAC
sci-ATACseq
C
Integration
RNA-seq
+ ATAC-seq
wnnUMAP_2
wnnUMAP_1
Hepatocytes
LSECs
Kupffer Cells
B Cells
HSCs
Monocytes
Neurons
Cholangiocytes
D
Alb
Cyp3a25
Clec4g
Stab2
Reln
Dcn
Clec4f
Vsig4
Ptprc
Itga4
Ly6e
Oasl1
Scd1
Kcnip1
Spp1
Hepatocytes
LSECs
HSCs
Kupffer Cells
B Cells
Monocytes
Neurons
Cholangiocytes
Expression Level
F
No. of associated pCREs
Gene Rank
Cyp2b10
Dner
Fgf1
Taf5
Igf1
Gata4
Nfkb2
Hes1
E
Serpina1a
Clec4g
Stab2
Dcn
Colec10
Clec4f
Ptprc
Itga4
Oasl1
Scd1
Spp1
Hepatocytes
LSECs
HSCs
Kupffer Cells
B Cells
Monocytes
Neurons
Cholangiocytes
G
All peaks
pCRES
Normalized Density
Distance to TSS
H
chr14:62830000-63320000
[range 0-50]
Genes
4931440J10Rik
Wdfy2
Defb42
Fdft1
Ints6
Gm26536
n-R5s47
Defb48
Defb47
Defb43
Defb30
Ctsb
Neil2
Gata4
pCREs
Links
Spearman corr.

Figure 2: Joint expression and chromatin accessibility profiling in primary liver nuclei

(A) Comparison of UMIs/cell across different single-cell technologies. Red shading denotes all multi-omic technologies. Datasets are this study, SHARE-seq [13] (murine skin cells), sci-CAR [11] (murine kidney nuclei), SNARE-seq [12] (adult & neonatal mouse cerebral cortex nuclei), 10x 3' Expression [17] (murine liver nuclei) and sci-RNAseq3[9] (E16.5 mouse embryo nuclei). Cells have been downsampled to a common sequencing depth where possible (see Methods).

(B) Comparison of unique fragments per cell across different single-cell technologies. Colouring as in (A). Datasets differing to (A) are 10x 3' scATAC[46] (murine liver nuclei) and sciATAC-seq[47] (murine liver nuclei). Cells have been downsampled to a common sequencing depth where possible (see Methods).

(C) UMAP visualisation of WNN-integrated scRNAseq and scATACseq modalities of 19,664 liver nuclei. Nuclei are coloured by cell types.

(D) Normalised gene expression of representative marker genes per cell type.

(E) Aggregate ATAC-seq tracks at marker accessibility peaks per cell type.

(F) Genes ranked by number of significantly correlated pCREs ($P < 0.05$, FDR = 0.1) per gene (±500kbp from TSS) in LSECs. Marked are transcription factors & regulators within the top 1% of genes with a critical role in LSECs.

(G) Significantly correlated pCREs are enriched for TSS proximity. Normalised density of all peaks versus pCREs within ±50kbp of nearest TSS.

(H) Aggregate scATAC-seq track of LSECs at the *Gata4* locus and 500kbp upstream region. Loops denote pCREs significantly correlated with *Gata4* and are coloured by Spearman correlation of respective pCRE–*Gata4* comparison

Figure 3

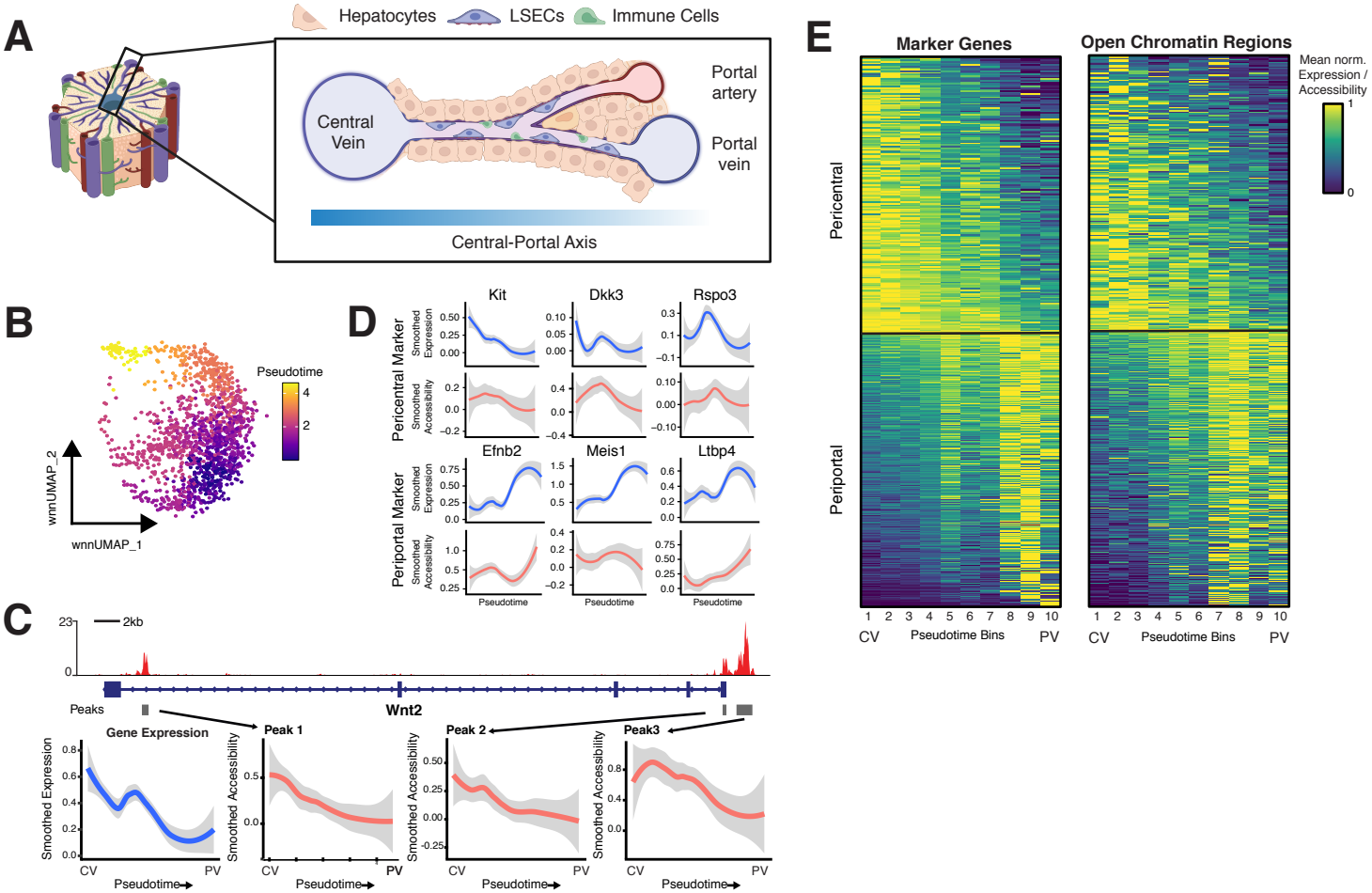


Figure 3: Zonation profiles in LSECs across gene expression and chromatin accessibility

(A) Schematic of a liver lobule. A liver lobule has a 'Central–Portal Axis' starting from the central vein to the portal vein and portal artery. The sinusoidal capillary channels are lined with LSECs.

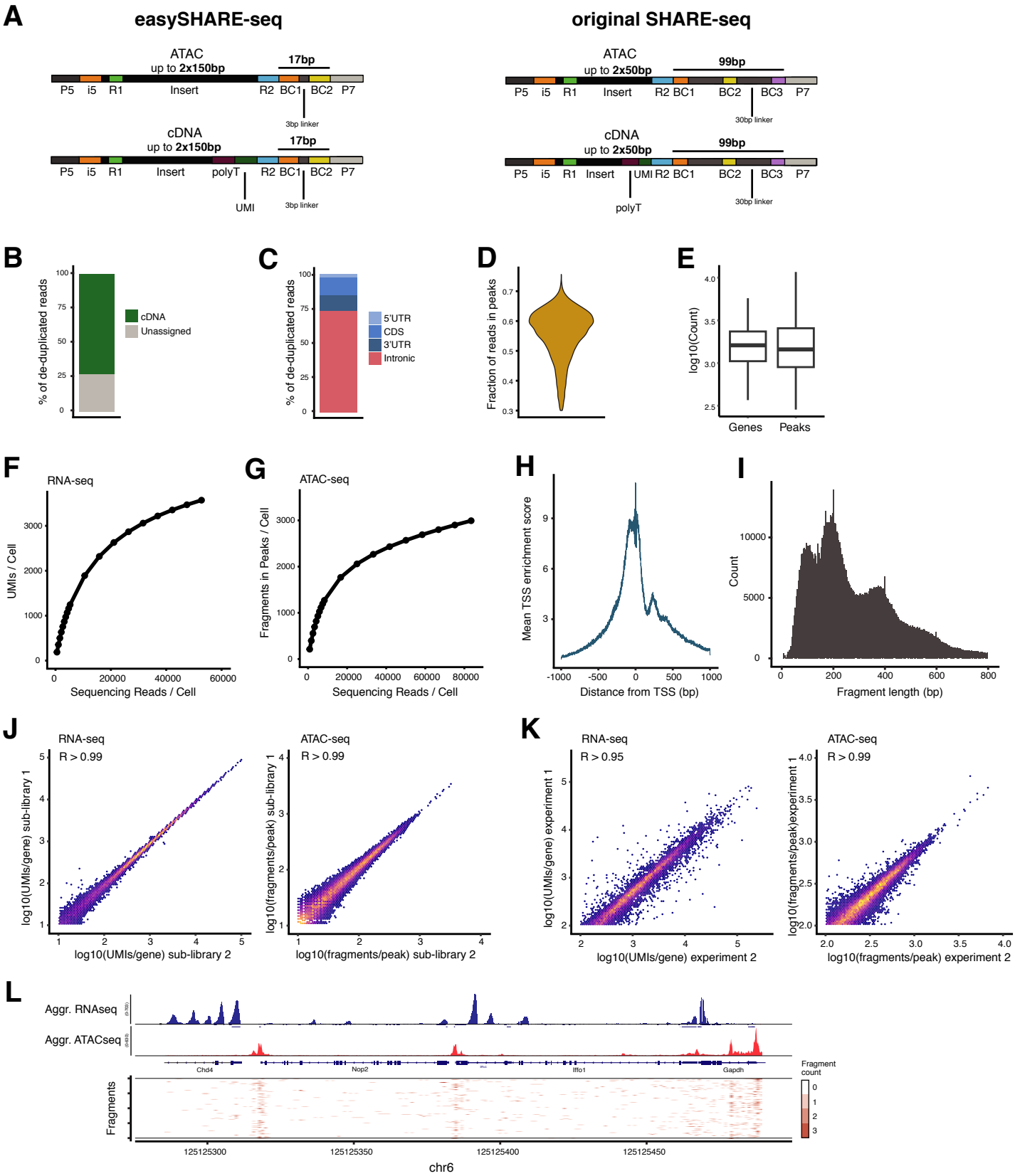
(B) UMAP of 1501 LSECs coloured by pseudotime.

(C) Changes along the Central–Portal Axis at the *Wnt2* locus. Top: Aggregate scATACseq profile (red) of LSECs at *Wnt2* locus. Grey bars denote identified peaks. Bottom: In blue, loess trend line of mean normalised *Wnt2* gene expression along the Central–Portal-Axis (central vein, CV; portal vein, PV). In red, loess trend line of mean normalised chromatin accessibility in peaks at the Wnt2 locus along the CP-axis.

(D) Loess trend line of mean normalised expression (blue) and mean normalised accessibility along the Central–Portal axis for pericentral markers (top, increased toward the central vein, *Kit*, *Dkk3* and *Rspo3*) and periportal markers (increased toward the portal vein, *Efnb2*, *Meis1* & *Ltbp4*)

(E) Left: Zonation profiles of 550 genes along the Central–Portal axis. Right: Zonation profiles of 744 open chromatin regions along the Central–Portal axis. All profiles are normalised by their maximum.

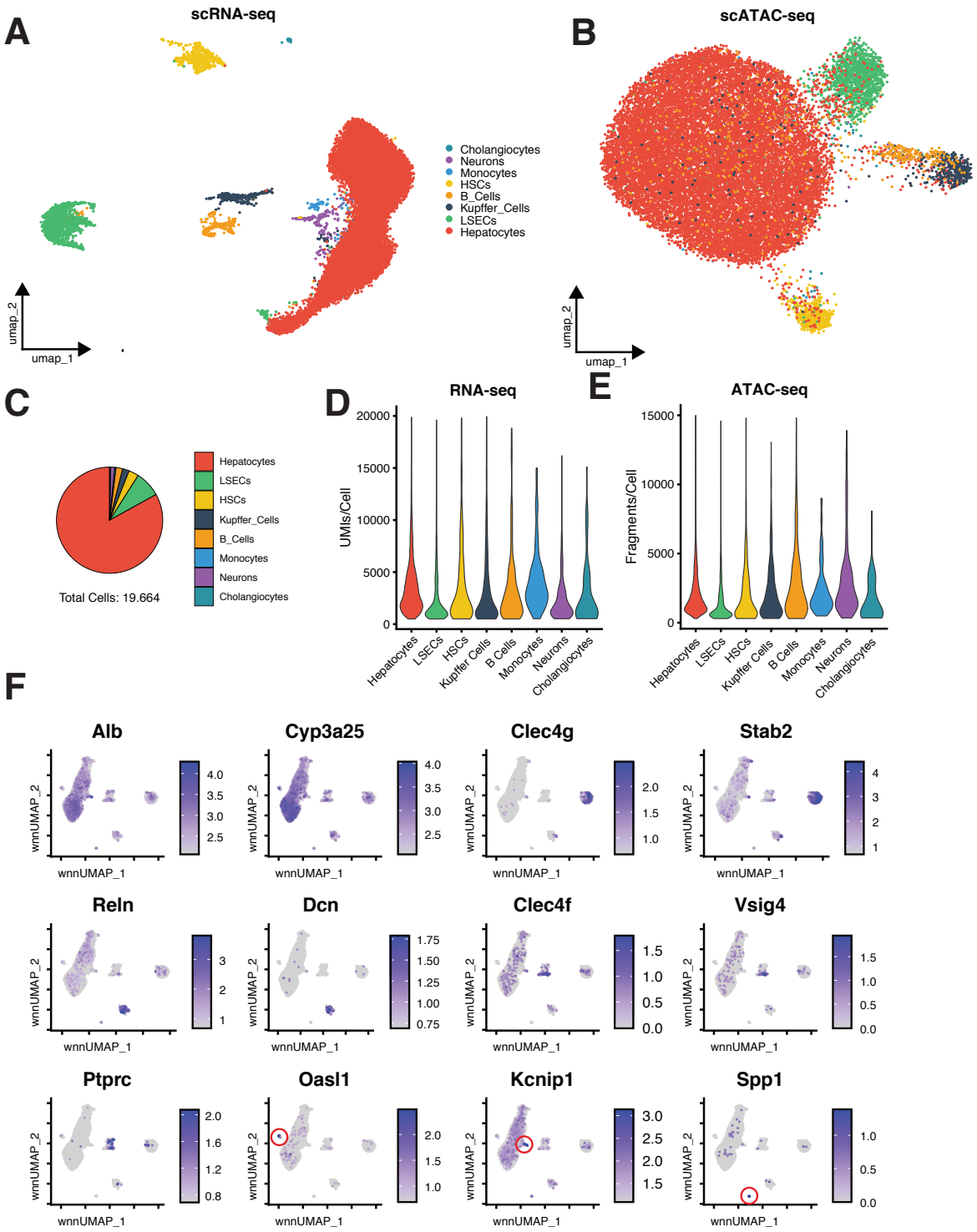
Supplemental Figure 1



Supplementary Figure 1: Barcode structure and summary of quality control measures in liver nuclei

(A) Structure of a scATAC-seq and scRNA-seq sequencing read in easySHARE-seq and the original protocol.
(B) Percentage of total scRNAseq sequencing reads containing cDNA fragments.
(C) Percentage of de-duplicated scRNA-seq sequencing reads overlapping an exon, intron, 5'UTR or 3'UTR.
(D) Distribution of fraction of reads in peaks (FRiP) per cell in the scATAC-seq data (mean: 0.55).
(E) Expressed genes and accessible peaks per cell (mean expressed genes: 1,798; mean accessible peaks: 1,983)
(F) Number of mean UMIs per cell recovered when subsampling to different raw sequencing depths.
(G) Number of mean fragments in peaks per cell recovered when subsampling to different raw sequencing depths.
(H) Mean TSS enrichment score per cell in relation to distance from nearest TSS in the scATAC-seq data.
(I) Histogram of fragment length in scATAC-seq sequencing reads
(J) Reproducibility of easySHARE-seq between sub-libraries
(K) Reproducibility of easySHARE-seq between biological replicates
(L) Top: Aggregate scRNA-seq (blue) and scATAC-seq (red) of all liver nuclei at *Nop2*/*Iffo2*/*Gapdh* locus. Bottom: Chromatin accessibility profiles of 100 individual cells.

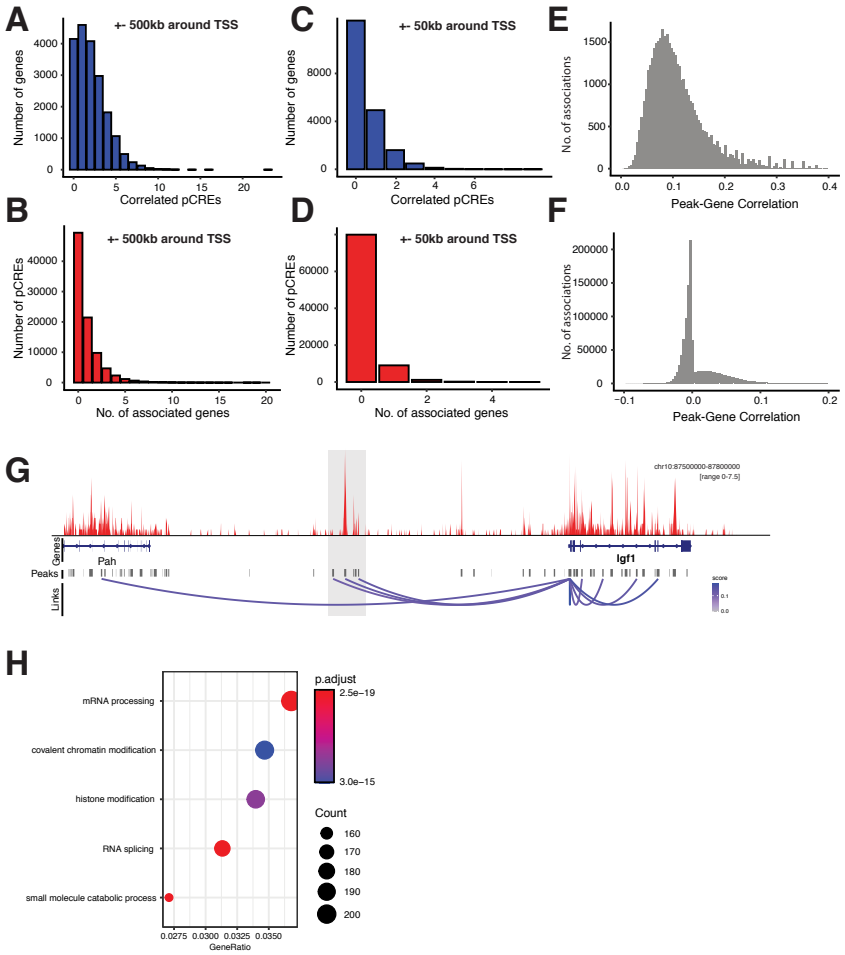
Supplementary Figure 2



Supplementary Figure 2: easySHAREseq robustly separates cell types

(A) UMAP visualisation of merged and integrated scRNA-seq data. Nuclei are coloured according to their cell type.
(B) UMAP visualisation of merged and integrated scATAC-seq data. Nuclei are coloured according to their cell type.
(C) Fraction of cell types recovered relative to total cells
(D) Distribution of UMIs per cell split by cell type. Some cell types (e.g. LSECs) consistently yield less UMIs.
(E) Distribution of unique fragments per cell split by cell types. Some cell types (e.g. LSECs) consistently yield less fragments.
(F) WNN-UMAPs with cells coloured according to the mean expression strength of a given marker gene. Red circles indicate the position of the cell population showing elevated expression for this marker gene.

Supplementary Figure 3



Supplementary Figure 3: Summary of peak–gene correlations

(A) Number of significantly correlated pCREs ($P < 0.05$, FDR = 0.1) per gene, considering all peaks ±500kbp of the TSS
(B) Number of genes a given pCREs is significantly correlated with ($P < 0.05$, FDR = 0.1), considering all peaks ±500kbp of the TSS
(C) Number of significantly correlated pCREs ($P < 0.05$, FDR = 0.1) per gene, considering all peaks ±50kbp of the TSS
(D) Number of genes a given pCREs is significantly correlated with ($P < 0.05$, FDR = 0.1), considering all peaks ±50kbp of the TSS
(E) Histogram of Spearman correlations of all significant peak–gene correlations ($P < 0.05$)
(F) Histogram of Spearman correlations of all non-significant peak–gene correlations ($P > 0.05$)
(G) Aggregate scATAC-seq track of LSECs at the *Igf1* locus and its upstream region. Loops denote significantly correlated pCREs with *Igf1* and are coloured by their respective Spearman correlation. Shaded grey area denotes potentially LSEC-specific *cis*-regulatory element regulating *Igf1* expression.
(H) Gene Ontology enrichment analysis of genes whose associated pCREs are associated with five or more genes.