

# Tissue-specific chromatin accessibility and architecture signature at single-cell level resolution

Amin Noorani<sup>1</sup>, Ayush T Raman<sup>2</sup>, Elias Orouji<sup>1</sup>

<sup>1</sup> Epigenome Lab, Princess Margaret Genomics Centre (PMGC), University Health Network, Toronto

<sup>2</sup> Epigenomics Program, Broad Institute of MIT and Harvard

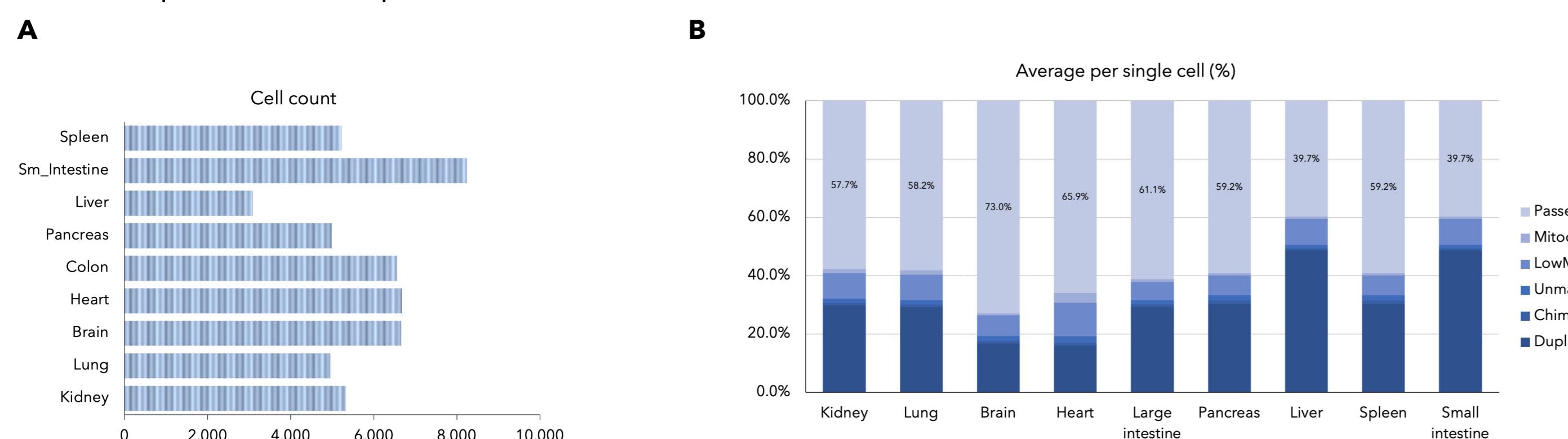
## Background

Most of the molecular profiling at single cell level are performed using single cell RNA sequencing (scRNA-seq). Mapping chromatin accessibility along with 3D chromatin structure can define chromatin regulatory landscape which governs transcription in each cell type.

In this study, we aimed to characterize accessible chromatin regions and further interrogate chromatin interactions within these accessible regions at single cell level. Using scATAC-seq and HiC we identified tissue-specific chromatin accessible regions and interaction of these regions in various mouse organs. Our analysis revealed uniquely structured open chromatin regions in each specific tissue type.

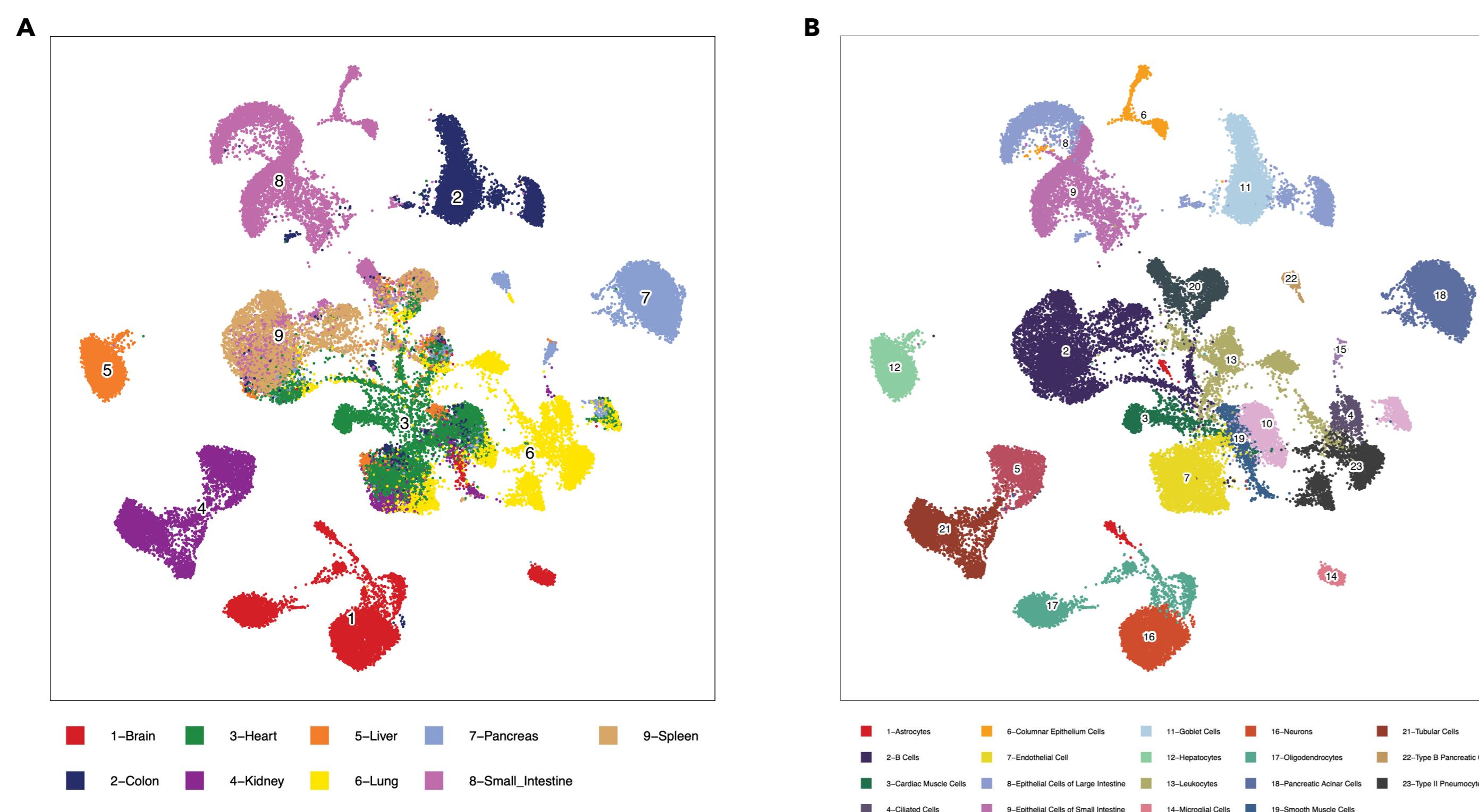
## Methods and Results

We isolated nuclei from 9 distinct flash-frozen tissues of 8 weeks-old female mice. Nuclei were processed using Triton-based lysis buffer. We applied scATAC-seq to profile genome-wide chromatin accessibility, and with ~1.72 billion read pairs we identified 51,704 single nuclei derived from these mouse tissues using 10x Genomics protocol. To ensure the quality of processed samples, QC metrics were applied. Samples were assessed for cell numbers, nucleosome pattern, read depth and proportion of reads overlapping peaks. The total number of cells profiled per tissue ranged from 3,090 for liver to 6,673 for heart. **Figure 1A** shows the number of nuclei profiled for each tissue type. Nucleosome pattern were observed in all samples. Median of the proportion of reads overlapping peaks were 0.42, which is in the optimal range. **Figure 1B** shows percentage of mean reads per cell that pass all the filters for each tissue.



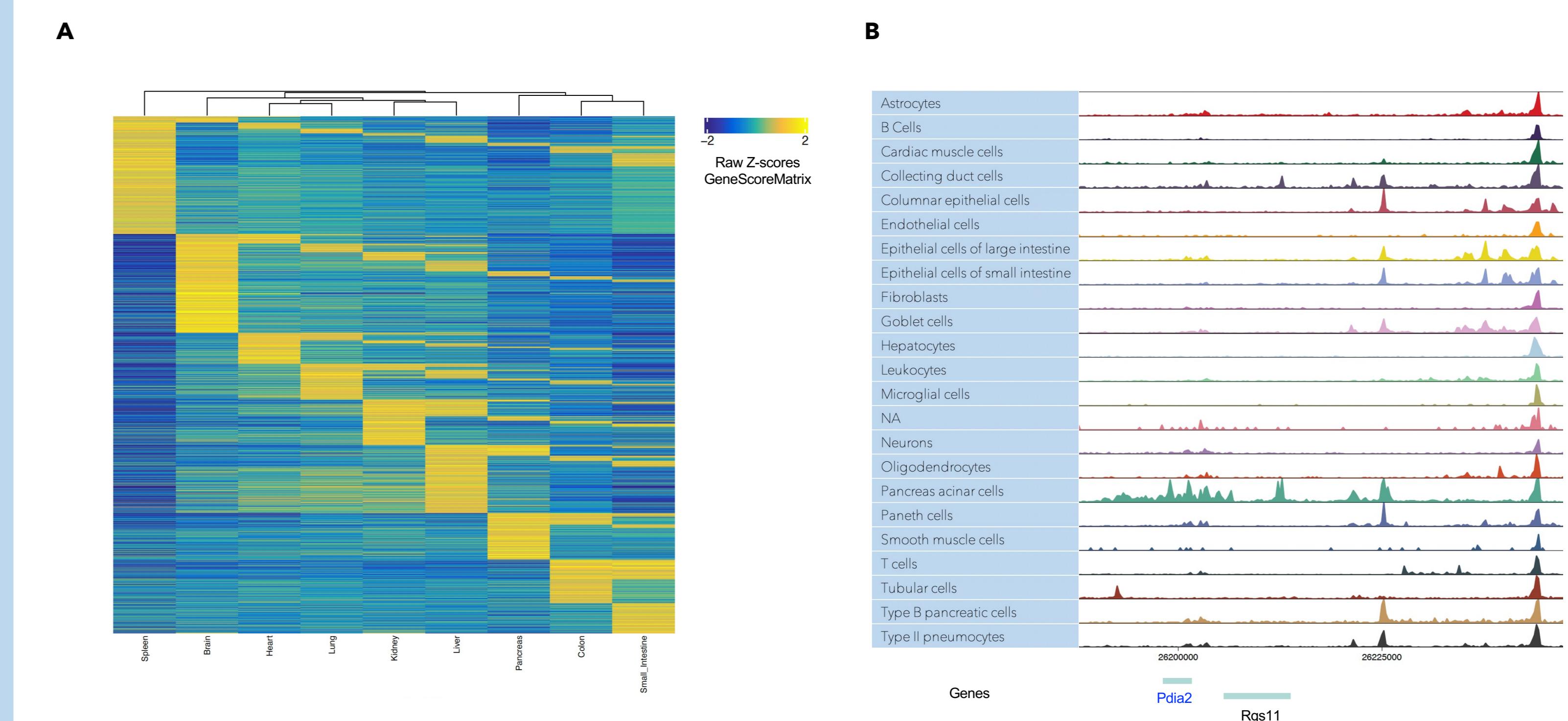
**Figure 1.** (A) Number of profiled cells for each tissue. (B) Percentage of reads passed all filters.

To identify clusters of cells with similar chromatin accessibility patterns, we first generated a comprehensive list of accessible sites, then scored all cells based on the presence or absence of these regions and subjected cells to UMAP embedding. Next, we identified 23 major clusters of cells. **Figure 2A** demonstrates cell clusters based on their tissue of origin and **Figure 2B** shows identified annotated cell types using gene activity scores.



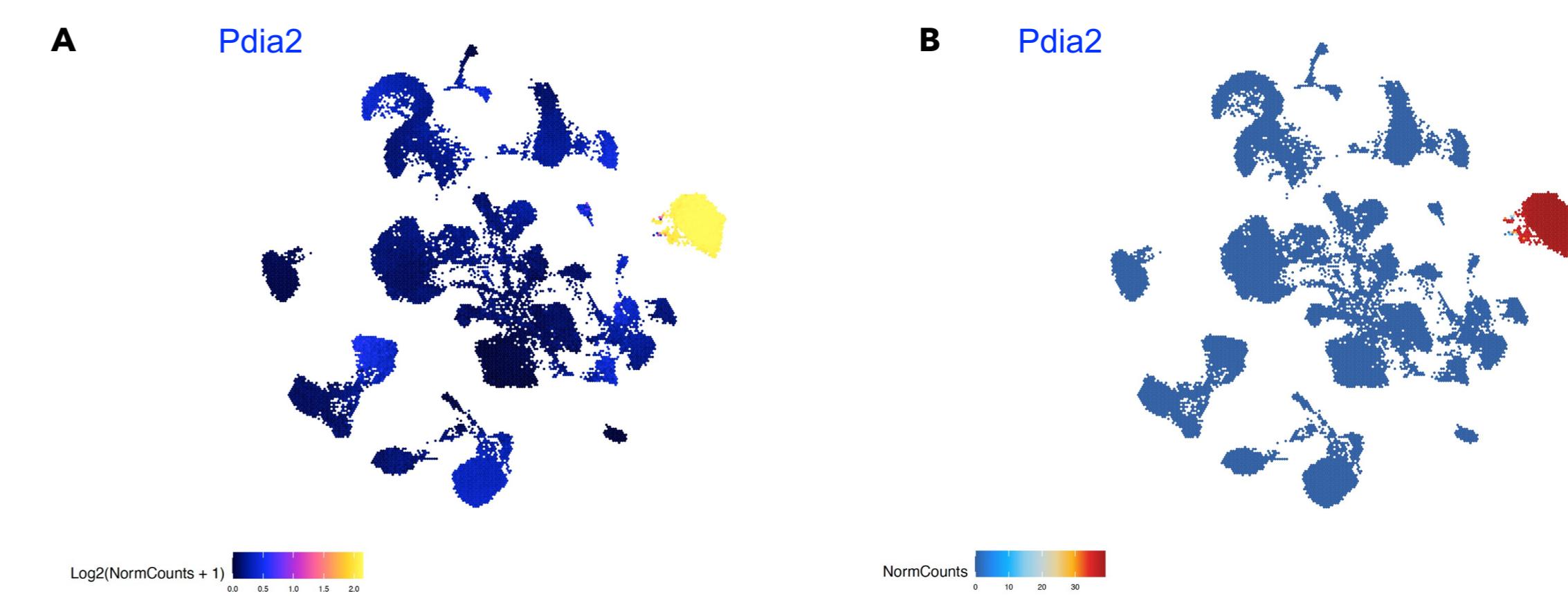
**Figure 2.** (A) Cell cluster based on the organ of origin. (B) Annotated cell types using scATAC-seq data.

Next, we identified open chromatin regions across all tissues, that are unique to each particular organ. **Figure 3A** shows a heatmap of open regions across various tissue types. We used this approach to select accessible chromatin regions that are unique to each organ, and in the next step identified open regions that are cell-type specific. We also identified a number of open chromatin regions that are shared across all tested tissue types which are considered as "essential open chromatin regions". **Figure 3B** demonstrates a representative open chromatin region that is specifically accessible in pancreas tissue.



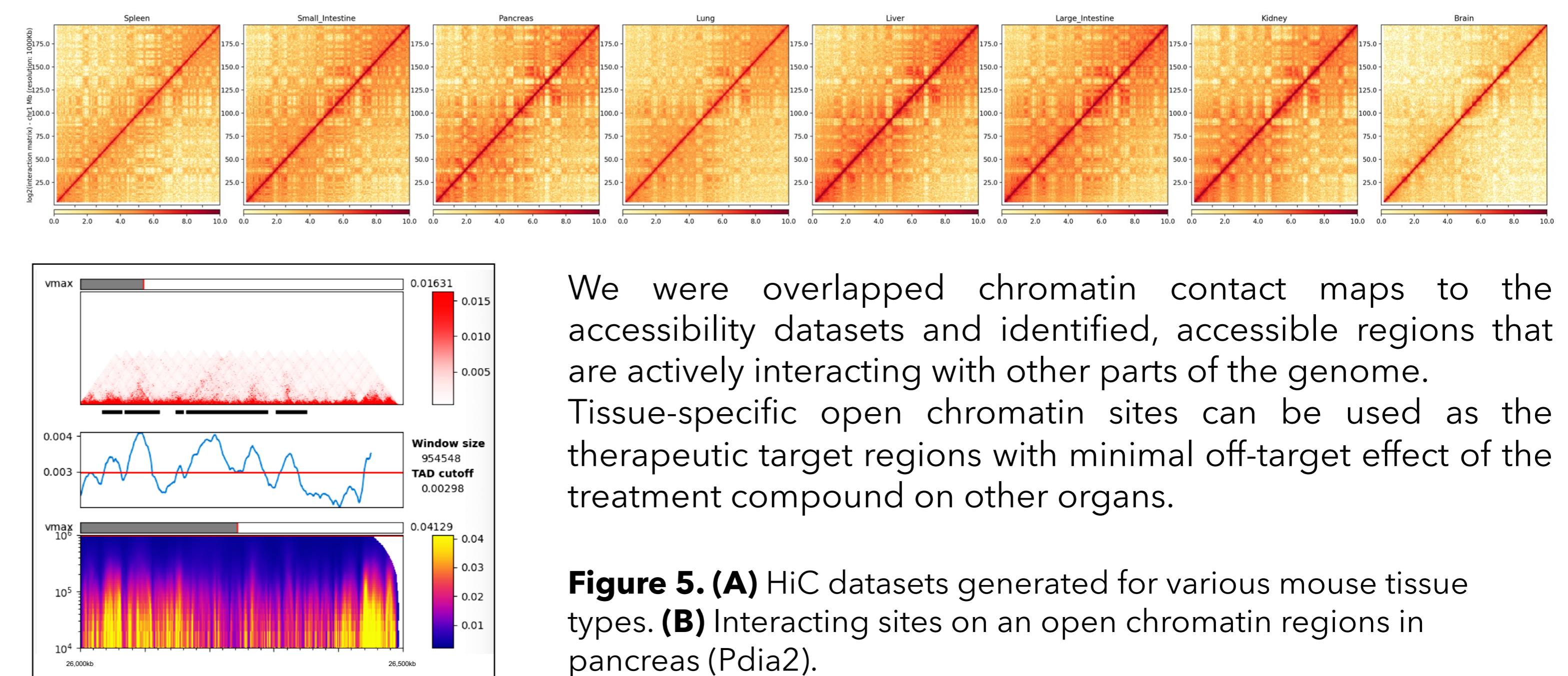
**Figure 3.** (A) Chromatin accessibility signature in various tissues (B) Representative open region (Pdia2) unique to pancreas tissue.

Next, we used scRNA-seq datasets available from TM mouse atlas. ArchR is used to integrate these datasets. **Figure 4** shows integrated map of scATAC and scRNA for Pdia2 region previously identified as a pancreas-specific accessible region.



**Figure 4.** A representative open region (Pdia2) in mouse (A) using scATAC data (B) using integrated map of scATAC with scRNA-seq.

We further generated chromatin conformation capture (HiC) datasets of these mouse tissues. Generated data were processed using HiC-Pro and then enhanced using DeepHiC.



We overlapped chromatin contact maps to the accessibility datasets and identified, accessible regions that are actively interacting with other parts of the genome. Tissue-specific open chromatin sites can be used as the therapeutic target regions with minimal off-target effect of the treatment compound on other organs.

**Figure 5.** (A) HiC datasets generated for various mouse tissue types. (B) Interacting sites on an open chromatin region in pancreas (Pdia2).

## Conclusion

Utilizing chromatin accessibility along with 3D genome contact maps across various tissue types results in (a) identification of interacting regions in accessible chromatin unique to each tissue, (b) identification of essential accessible sites across all tissues and (c) elucidates regulatory 3D map accessible chromatin specific to each organ.

## References

- Cusanovich DA et al. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell*, 2018
- Granja JM, Corces MR et al., ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics* (2021)
- Kruse K, Hug CB, Hernández-Rodríguez B, Vaquerizas JM. TADtool: visual parameter identification for TAD-calling algorithms. *Bioinformatics*. 2016 Oct 15;32(20):3190-3192.