CC BY 4.0

DeepSpaceDB: a spatial transcriptomics atlas for interactive in-depth analysis of tissues and tissue microenvironments

Vladyslav Honcharuk[1], Afeefa Zainab[1], Keiko Takemoto[1], Diego Diez[2], Shinpei Kawaoka[1,3], Alexis Vandenbon[1,4,5,*]

E-mail: alexisvdb@infront.kyoto-u.ac.jp

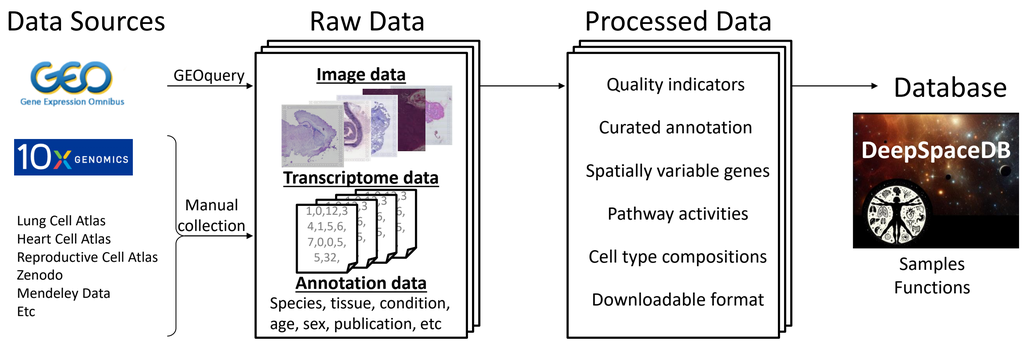
[1] Institute for Life and Medical Sciences, Kyoto University; [2] Immunology Frontier Research Center, Osaka University; [3] Institute of Development, Aging and Cancer, Tohoku University; [4] Institute for Liberal Arts and Sciences, Kyoto University; [5] Graduate School of Biostudies, Kyoto University

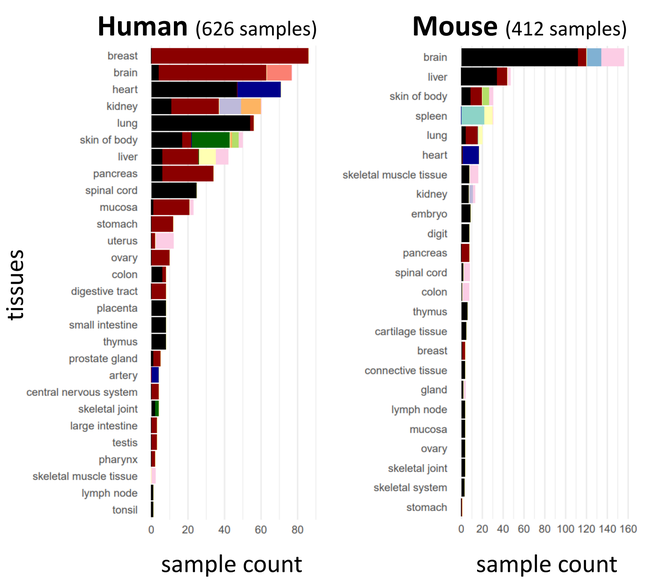
Summary

Spatial transcriptomics enables researchers to study the spatial organization of cells and gene expression patterns within tissues. However, this technology requires considerable financial resources and bioinformatics experience. Here, we present DeepSpaceDB, a spatial transcriptomics atlas that enables in-depth exploration of public spatial data. The roughly 2.2 million spots of more than 1,000 Visium samples were clustered by similarity of their gene expression patterns into manually annotated clusters, facilitating the interpretation of structures within tissues. Furthermore, DeepSpaceDB allows users to compare the gene expression of interactively selected sets of spots. Finally, the database provides spatially variable genes and biological pathways, as well as predicted cell type compositions of all samples. DeepSpaceDB is available at www.DeepSpaceDB.com.

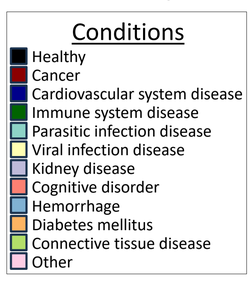
Overview of the data

We collected spatial transcriptomics samples of the 10x Genomics Visium platform from multiple sources. Data was processed using a consistent pipeline, and was made available in DeepSpaceDB along with various analysis functions.

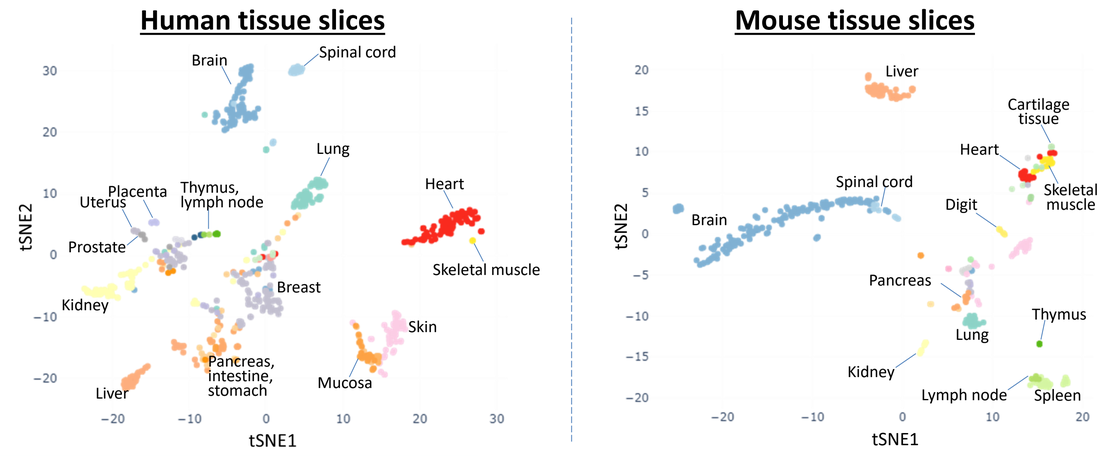




At present, our collection includes 626 human and 412 mouse samples, covering a variety of tissues and conditions. Human data is especially rich in cancer-related samples, while mouse data includes a large number of brain samples.

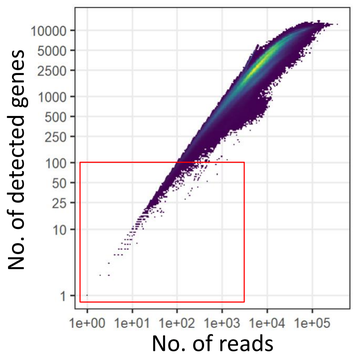


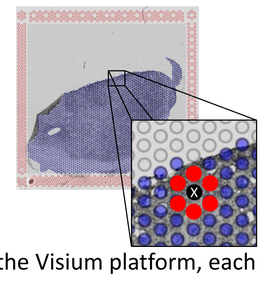
To evaluate the biological validity of the collection of samples, we converted each tissue slice into a "pseudo-bulk" sample, and visualized these in a 2D embedding. We confirmed that samples from similar tissues group together, even if they were generated by different studies.



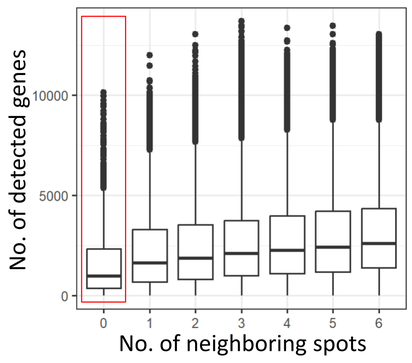
Quality control

There is considerable variety in the 1,038 samples and the 2.2 million spots that they contain. A subset of low-quality spots contains less than 100 detected genes. Especially isolated spots (with no neighboring spots) tend to have a lower quality.



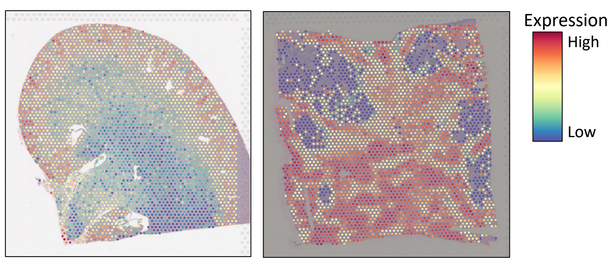


In the Visium platform, each spot has at most 6 neighboring spots.



Spatially variable features

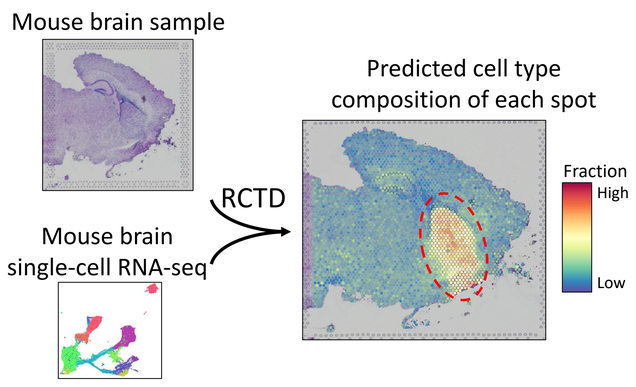
We used our method *singleCellHaystack* (Vandenbon & Diez, 2020 and 2023) to predict spatially variable genes and biological processes in each sample.



As an example, *Cyp2e1* shows clear spatial patterns of expression in kidney (left) and liver (right) samples. In the liver sample shown here, the top part includes colon cancer spots, which don't express *Cyp2e1*. At the bottom, a zonated expression pattern can be seen.

Cell type compositions

Each Visium spot covers roughly 10 cells. We are using RCTD (Cable *et al.,* 2022) and Cell2location (Kleshchevnikov *et al.*, 2022) to predicting the cell type composition of each spot. To improve the accuracy of these predictions, we are collecting and constructing suitable single-cell RNA-seq reference datasets.



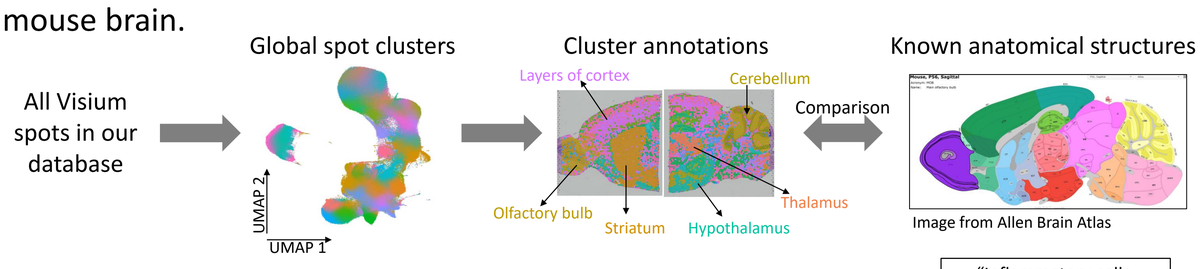
For example, to predict the cell type composition of a mouse brain sample, we use a mouse brain single-cell RNA-seq datasets with cell type annotation data as a reference. Shown here is the predicted fraction of medium spiny neuron cells at each spot.

Depending on the source tissue and condition, we use a different reference.

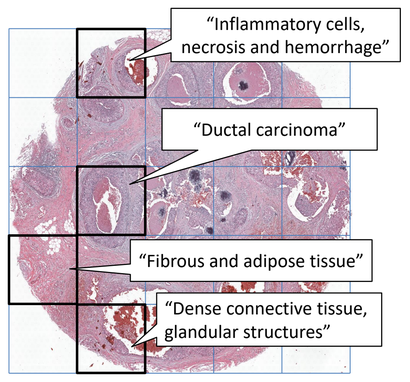
Tissue structure annotation

One way to assist the exploration of spatial data is by annotating different structures within tissues. We are exploring different approaches to doing this.

Approach 1: We clustered the spots of all Visium samples in our database by similarity of gene expression. By comparing the spot clusters with know anatomical features, we assigned annotations to spot clusters. Below we show an example of mouse brain.

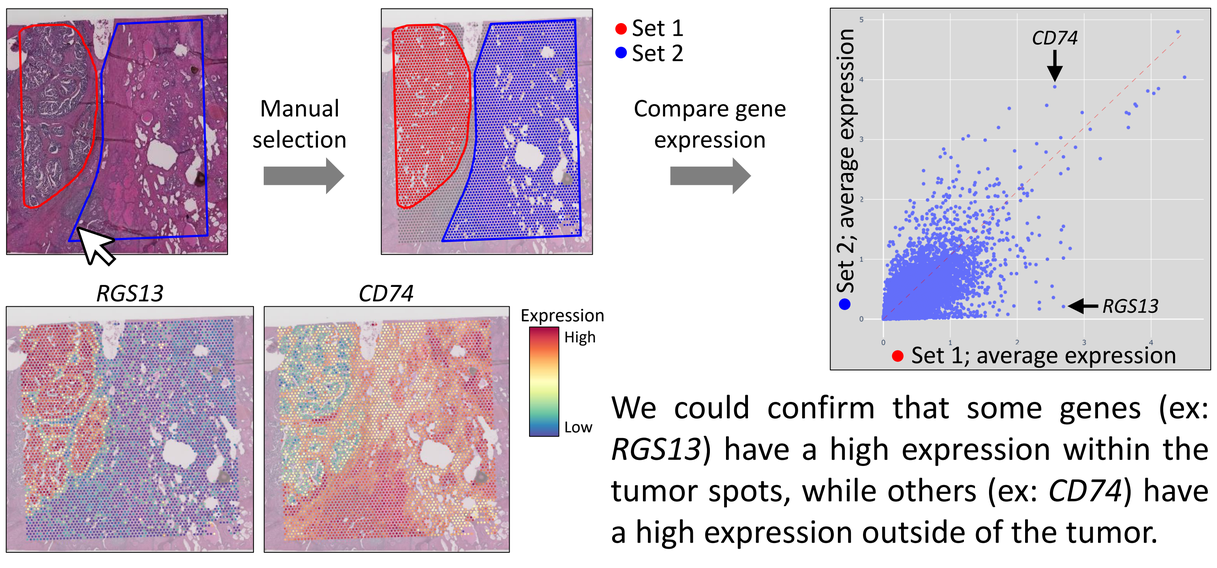


Approach 2: We annotated the Visium image data using a large language model (LLM). Do to this, we cut each image into a 5-by-5 grid and asked an LLM to summarize any anatomical and pathological features. The figures on the right shown an example with a few indicated descriptions. In the future, these will be added to the database.



Interactive exploration

We implemented a tool where users can interactively select multiple sets of spots within a tissue, and compare their gene expression. Below is an example of a prostate gland with prostate cancer. We selected a part which includes cancer spots (set 1) and a part which does not (set 2), and compared them.



We could confirm that some genes (ex: *RGS13*) have a high expression within the tumor spots, while others (ex: *CD74*) have a high expression outside of the tumor.