
Supplementary information

Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram

In the format provided by the
authors and unedited

Deep learning and alignment of spatially-resolved single cell transcriptomes with Tangram – Supplementary Material

Tommaso Biancalani^{*,1,9,#}, Gabriele Scalia^{*,1,10}, Lorenzo Buffoni², Raghav Avasthi^{1,3}, Ziqing Lu^{1,3}, Aman Sanger¹, Neriman Tokcan¹, Charles R. Vanderburg¹, Åsa Segerstolpe¹, Meng Zhang^{4,6}, Inbal Avraham-Davidi¹, Sanja Vickovic¹, Mor Nitzan^{1,5,11}, Sai Ma^{1,7,8}, Ayshwarya Subramanian¹, Michal Lipinski^{1,8}, Jason Buenrostro^{1,8}, Nik Bear Brown³, Duccio Fanelli², Xiaowei Zhuang^{4,6}, Evan Z. Macosko¹ and Aviv Regev^{1,6,7,9,#}

¹Broad Institute of MIT and Harvard, Cambridge, MA 02142,

²Department of Physics and Astrophysics, University of Florence, Florence, Italy

³Northeastern University, Boston, MA 02115

⁴Department of Chemistry and Chemical Biology, Department of Physics, Harvard University

⁵School of Engineering and Applied Sciences, Harvard University

⁶Howard Hughes Medical Institute

⁷Department of Biology, MIT, Cambridge, MA 02140

⁸Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, 02138 USA.

⁹Current address: Genentech, 1 DNA Way, South San Francisco, CA, 94080

¹⁰Current address: Roche, Monza, Italy

¹¹Current address: School of Computer Science and Engineering, Racah Institute of Physics, Faculty of Medicine;
The Hebrew University, Jerusalem, Israel

* These authors contributed equally

To whom correspondence should be addressed: tommaso.biancalani@gmail.com (TB) aviv.regev.sc@gmail.com (AR)

Tangram helps detect cell type patterns conserved cross species

Here, we tested how Tangram performs when the input scRNA-seq and spatial data do not match as closely, as in the special case when each is derived from a different species (**Extended Data Fig. 3**), which we tested in both the brain and the kidney (chosen by data availability).

For cross species brain mapping, we mapped pre-annotated human brain MOp snRNA-seq data¹, collected post mortem (**Methods**) onto mouse MERFISH MOp, and compared it to the corresponding within-species mapping (mouse MOp snRNA-seq with mouse MERFISH MOp; **Fig. 2**), finding high concordance in the mapping patterns for most cell types (**Extended Data Fig. 3a,b, Methods**). Specifically, we compared the probability maps for the cell types shared between the human and mouse snRNA-seq datasets, namely, all cell types except *Meis* and *Peri*, which were not present in the human snRNA-seq data (**Extended Data Fig. 3a**). Excitatory neurons (glutamatergic) formed the expected cortical patterns, with the sole exception of L6b neurons being distributed across layer 6, and inhibitory neurons (GABAergic) were also in excellent agreement, with *Lamp5* cells being slightly less localized in the cross-species case. The least similar patterns were in non-neuronal *mPVM* and *VLMC* cell types, also reflected quantitatively by the cosine similarity metric (**Extended Data Fig. 3b**). Overall, the mappings at cell type level were highly concordant, consistent with the high similarity of cell profiles for mouse and human cortical brain cells².

When we compared the predicted spatial gene expression of individual genes (focusing on gene orthologs of those we validated against images from the Allen atlas (**Figs. 2e,f**)), the similarity, while present, was lower than that for cell type maps (**Extended Data Fig. 3c**). Interestingly, the

extent of agreement does not relate to specific cell types. For instance, some marker genes of cortical layers display similar patterns (*SEMA3E* and *ADAMTS3*), whereas other cortical markers are not localized in the cross-species case (*HTR4*, *KCNH5* and *CRISPLD1*). Differences in the extent of agreement are also observed between pairs of genes with granular patterns (*ERBB4* vs *ESRGG*) and pairs of paralogs (*ADAMTS2* and *ADAMTS3*).

Tangram also successfully mapped scRNA-seq profiles from human kidney³ onto a Visium coronal slide for mouse kidney⁴ (**Extended Data Fig. 3d, Methods**). To this end, we computed the top 100 marker genes for each pre-annotated cell type in the scRNA-seq data (**Methods**), intersected them with their orthologs measured in Visium, and performed a probabilistic mapping using the shared orthologs as a training set (**Extended Data Fig. 3e**). The projected cell type maps (**Extended Data Fig. 3e**) correctly capture several structures, including the connecting tubule, the collecting duct (the center part of the ROI where *Principal cells* concentrate), and the vasa recta as we move from the cortex to the medulla. We also observe co-localization of podocytes with *Glomerular EC*, and co-localization of *Vasa Recta EC* with *TAL of Loop of Henle* cells in the medulla, as we expected from kidney anatomy. Tangram did not assign high probability to *Pelvic epithelium* cells and *Transitional urothelium* cells, as we expected, since these pelvic regions are not captured in the coronal slice. Some immune cell types ($CD4^+$ and $CD8^+$ *T* cells, *MNP* cells) did not map as well and seemed to overfit specific voxels, suggesting that the resulting patterns are not biologically sound. These may reflect lower conservation of marker genes in immune cells, which can be addressed in the future by using scRNA-seq data from both species to optimize marker selection.

References

1. Human M1 10x. <https://portal.brain-map.org/atlas-and-data/rnaseq/human-m1-10x>.
2. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
3. Stewart, B. J. *et al.* Spatiotemporal immune zonation of the human kidney. *Science* **365**, 1461–1466 (2019).
4. 10X Genomics, Mouse Kidney Dataset. https://support.10xgenomics.com/spatial-gene-expression/datasets/1.1.0/V1_Mouse_Kidney.