

CRAK-Velo: Chromatin accessibility kinetics integration improves RNA velocity estimation

1 Performance on additional data set

As an additional data set, we consider a human cerebral cortex 10x multi-ome data set [1]. Figure S1 shows results analogous to figures in the main manuscript. (a) Shows a UMAP visualization of the velocity fields (shown in streamlines) inferred by CRAK-Velo, MultiVelo, and UniTVelo. The cells are colored by cell types. (b) PAGA graph summarizing transitions between different cell types obtained from the velocity field embeddings. (c) Visualising cell type deconvolution performance for three genes in un/spliced phase space versus chromatin accessibility-unsplliced space obtained by CRAK-Velo or MultiVelo. The number below each subplot is the accuracy of a KNN cell-type classifier. (d) Histogram of the performance of the KNN cell type classifier across all genes for the phase space and accessibility-unsplliced space inferred by the CRAK-Velo and MultiVelo. (e) Log enrichment of binding at high-weighted regions (from CRAK-Velo) relative to what is expected by chance. (f) Dotplot showing the biological processes enriched within the list of low entropy genes. (g) Regulatory dynamics of the DLGAP2 gene. Left is the region kinetic plot with double y axes that show the average accessibility c_r^g for each region around DLGAP2 on the left and the average normalized unsplliced reads u on the right one. The bottom left subplot shows the number of intersections of the binding sites of TF with the different regions (relative distance d of the regions from the gene is shown). The bottom right subplot is an arrow plot showing where thickness is the regions contribution/weight w_r^g inferred by CRAK-Velo located at distance d from the gene.

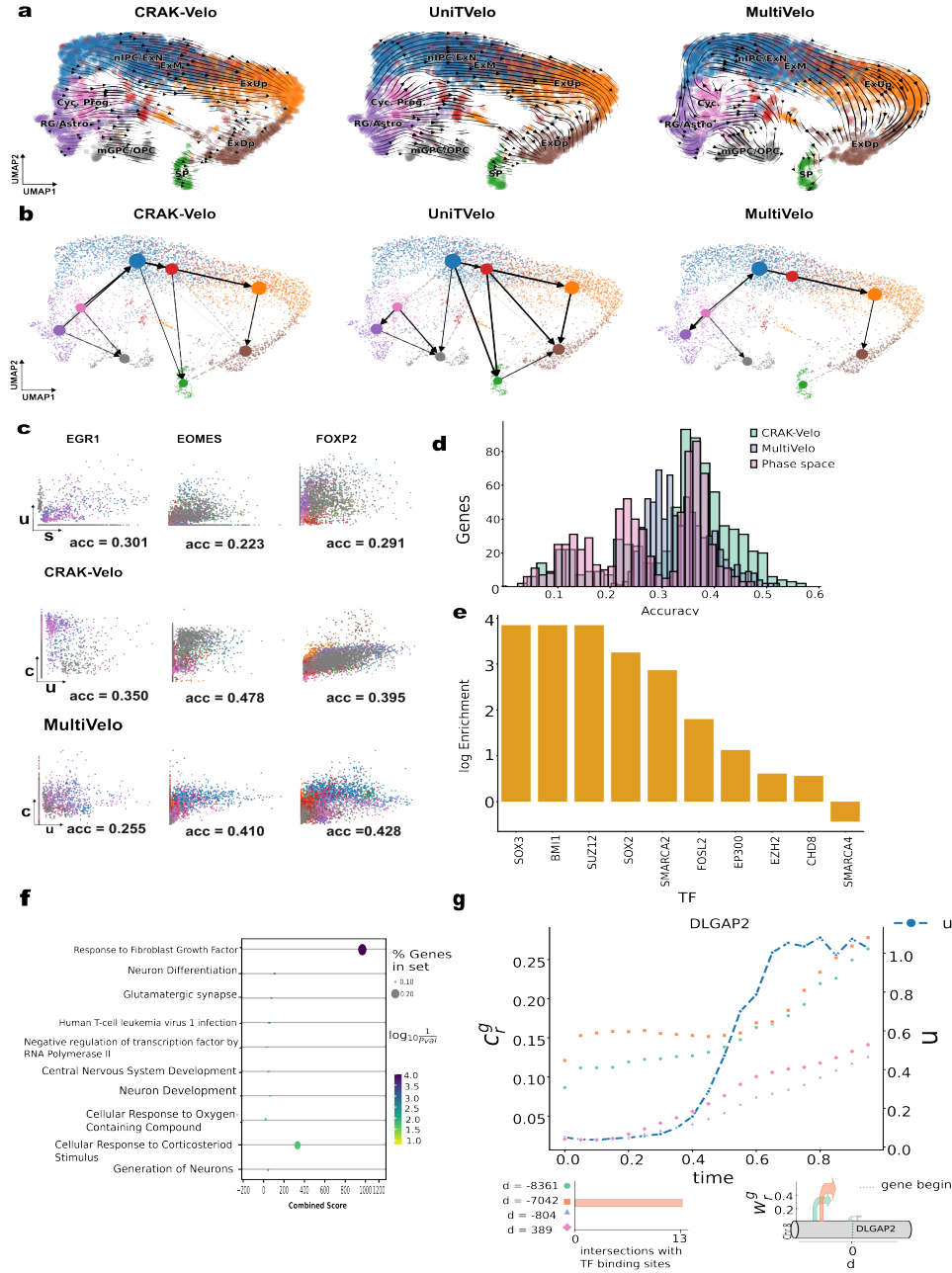


Fig. S1: See main supplementary text

2 Performance comparisons

We compare the performance (in terms of hours of compute time) for CRKA-Velo and MultiVelo in Table S1. We also insert the time needed by CisTopic (a pre-step for CRAK-Velo) for reference using standard settings

	CRAK-Velo	MultiVelo	CisTopic(Gibbs Sampler; T = 50)
HSPC	15	> 24	3
Mouse brain	2	4	0.5
HCC	6	24	2

Table S1: Runtime (in hours) of CRAK-Velo, MultiVelo, and CisTopic (T is the number of topics) across three datasets: HSPC, Mouse brain, and HCC. CRAK-Velo and Multivelo are run on CPU 86 Intel Xeon Platinum 8276/L 2.4GHz. While Gibbs sampler is run on NVIDIA A100-PCIE-40GB GPU.

References

- [1] Alexandro E Trevino, Fabian Müller, Jimena Andersen, Laksshman Sundaram, Arwa Kathiria, Anna Shcherbina, Kyle Farh, Howard Y Chang, Anca M Paşca, Anshul Kundaje, et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell*, 184(19):5053–5069, 2021.