

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

1 Sensitivity Analysis to model parameters

We wanted to evaluate the sensitivity of CRAK-Velo for hyperparameters like the number of topics (T) and the size of the window taken around each gene. CRAK-Velo robustness was tested on the three datasets by varying T and window size for each run.

At the level of the number of topics we run cisTopic with different T (T = 30 and T = 50) to obtain ϕ for each region. The correlation between ϕ of the region in the two runs were visualized via histograms (see Figure S2 in supplementary materials). In addition the histograms of the correlations velocity field of each cell and transcription rates c_g for each gene obtained from the different runs of CRAK-Velo with different T are also shown in Figure S1.

Regarding the window size, the histograms of correlation of velocity field for each cell and transcription rates c_g for each gene obtained from the different runs of CRAK-Velo with different window sizes (10kb and 20kb) are shown in Figure S2.

We observe that, while the HSPC and mouse brain data sets appear remarkably robust under these perturbations, the human cerebral cortex behaves differently: while the inferred flows remain consistent, the specific region-level inferences change under both perturbations. This is likely due to the lower coverage in this data set: the set of genes passing or pre-selection is lower than in the other data sets (around 1000), and only about 50% of genes have at least one region in the 10Kb window upstream/ downstream (compared to 78-87% for the other two data sets).

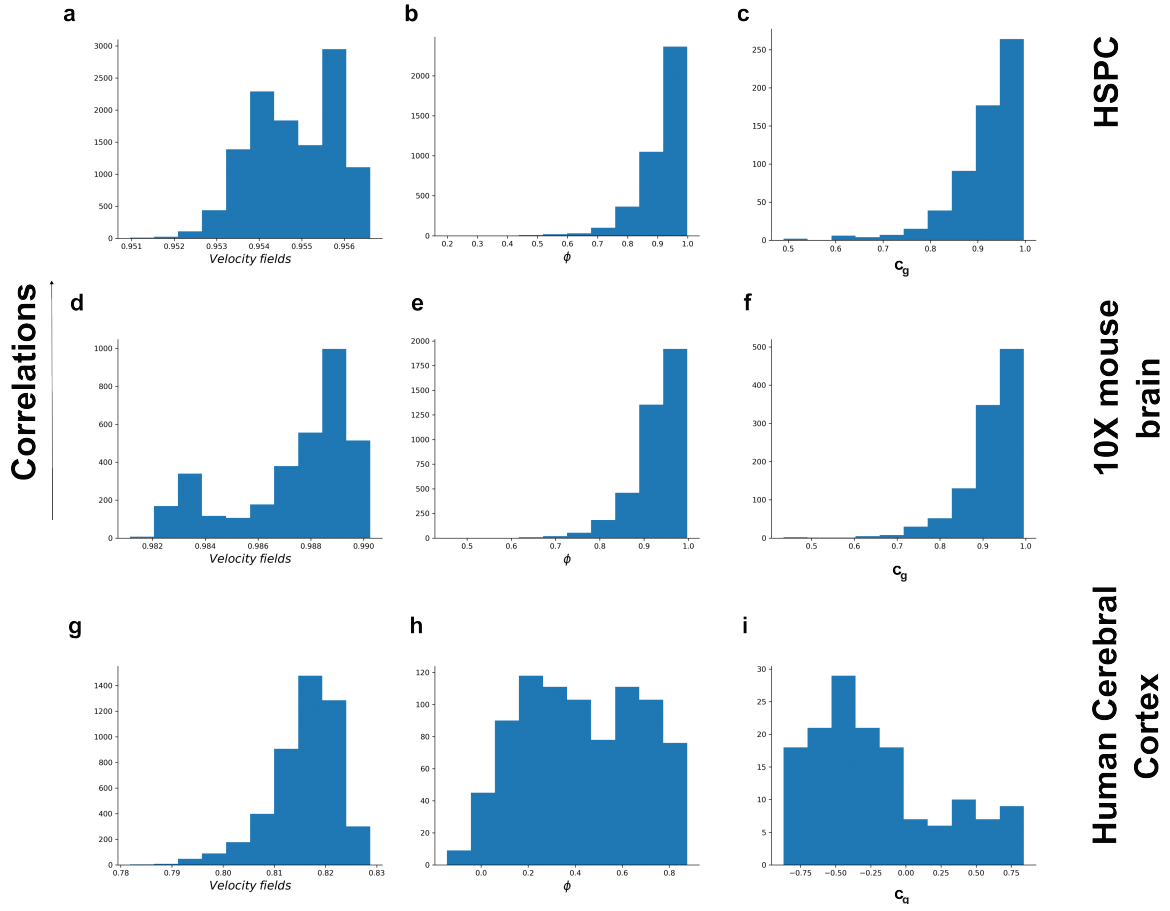


Fig. S1: Sensitivity analysis for the number of topics. CRAK-Velo is run on three datasets with different numbers of topics ($T = 20$ and $T = 50$). Histograms of correlations between the velocity fields of cells (a, d, g), ϕ region accessibility probability (b, e, h), and c_g transcription rate of genes (c, f, i). Each row represents a dataset.

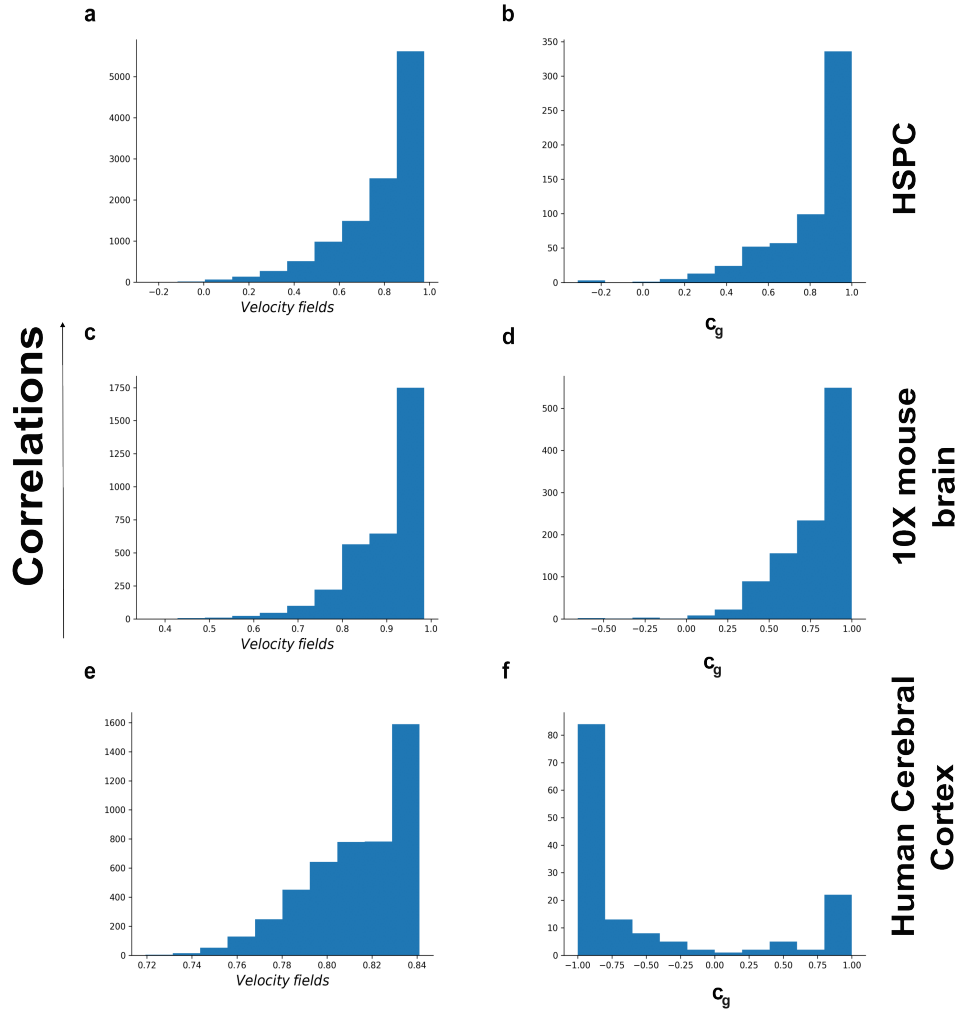


Fig. S2: Sensitivity analysis for the size of window w . CRAK-Velo is run on three datasets with different window sizes ($w = 10\text{kb}$ and $w = 20\text{kb}$). Histograms of correlations between the velocity fields of cells (a, c, e), and c_g transcription rate of genes (b, d, f). Each row represents a dataset.