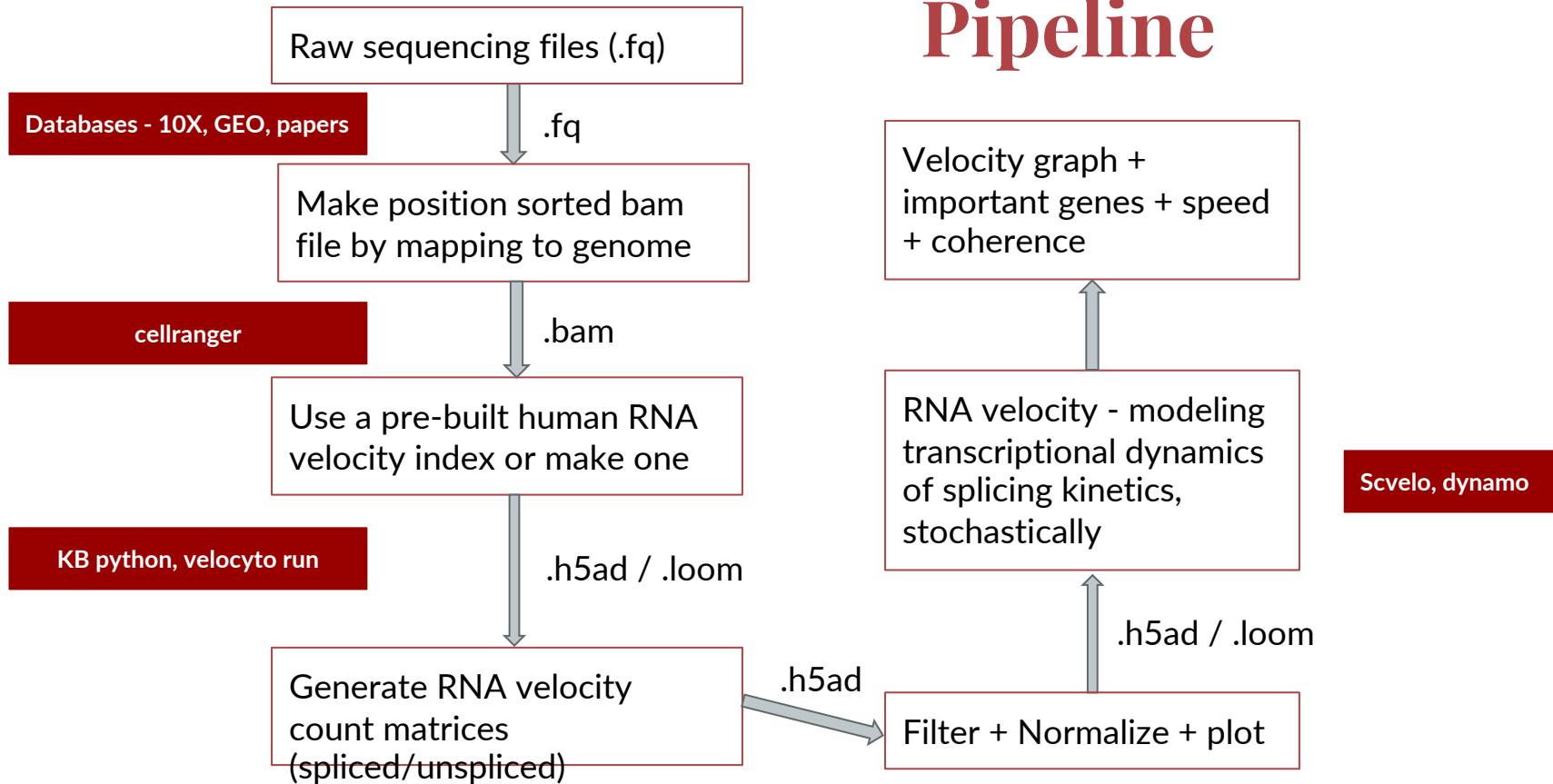


# **RNA Velocity in Single cells - An analysis of Publicly available datasets**

**Pallavi Surana**

**2022/04/03**

# Pipeline



1. [https://www.kallistobus.tools/tutorials/kb\\_velocity/python/kb\\_velocity/](https://www.kallistobus.tools/tutorials/kb_velocity/python/kb_velocity/)
2. <https://www.10xgenomics.com/resources/analysis-guides/trajectory-analysis-using-10x-Genomics-single-cell-gene-expression-data>

# Datasets analyzed

## Mouse Dentate gyrus -

- Dentate gyrus (DG) is part of the hippocampus involved in learning, episodic memory formation and spatial coding.
- The experiment from the developing DG comprises two time points (P12 and P35) measured using droplet-based scRNA-seq (10x Genomics Chromium). [1]

## Human neural stem cells -

- HNSC from the developing mammalian telencephalon (the most developed part of the forebrain).
- scRNA-seq to examine the cell cycle states of expanding human neural stem cells (hNSCs). The technology used was Illumina HiSeq 2000 [2]

## Human + Mouse neuronal splicing -

- Single-cell metabolically labeled new RNA tagging sequencing (scNT-Seq), a method for massively parallel analysis of newly-transcribed and pre-existing mRNAs from the same cell. [3]

1. <https://doi.org/10.1038/s41593-017-0056-2>

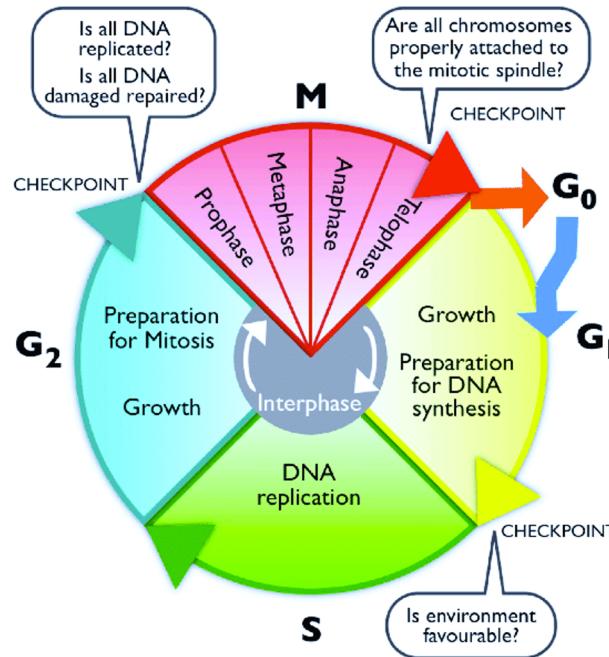
2. <https://doi.org/10.15252/msb.20209522> ; <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117004>

3. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141851>; <https://doi.org/10.3389/fmolb.2018.00012>

# Motivation and Aim

- To analyze the current and future cell states from scRNA seq datasets by mapping RNA velocity
- Compare and contrast the different cell cycle phases from the neural stem cells data with the other 2 brain datasets.
- Focused analysis of different regions of the brain.
- Mouse model is very accessible and widely used to study and understand more about the mammalian brain.

# Why cell cycle genes are important?



What happens in cancer?

- Cells are able to enter cell cycle without the presence of growth factors.
- They can evade growth suppressors and cell cycle checkpoints are usually evaded in cancer cells = indefinite growth.
- Hence studying cell cycle genes is important

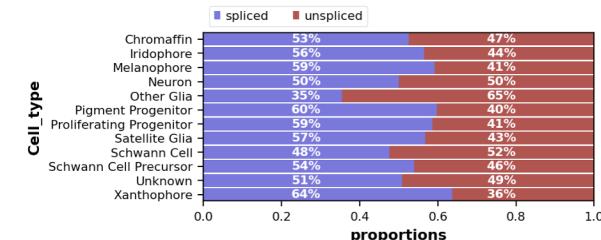
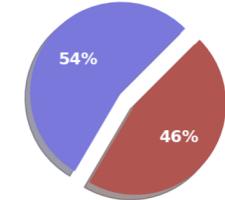
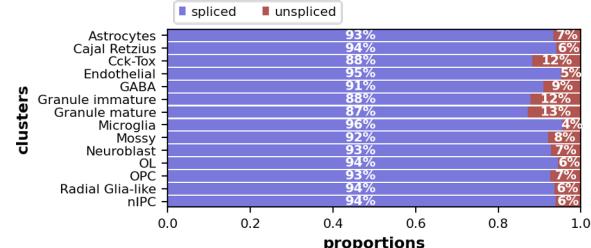
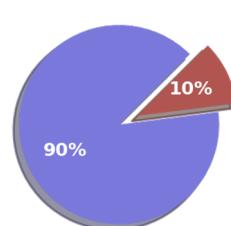
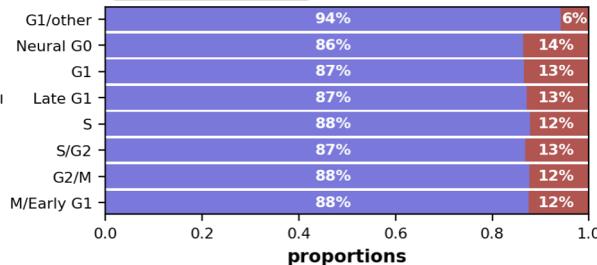
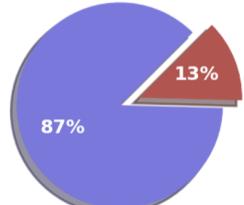
1. [https://www.researchgate.net/figure/The-cell-cycle-is-divided-into-four-phases-G-1-S-G-2-M-Resting-cells-are-in-a-G\\_fig1\\_264548273](https://www.researchgate.net/figure/The-cell-cycle-is-divided-into-four-phases-G-1-S-G-2-M-Resting-cells-are-in-a-G_fig1_264548273)

# Spliced and unspliced mRNA content

DG

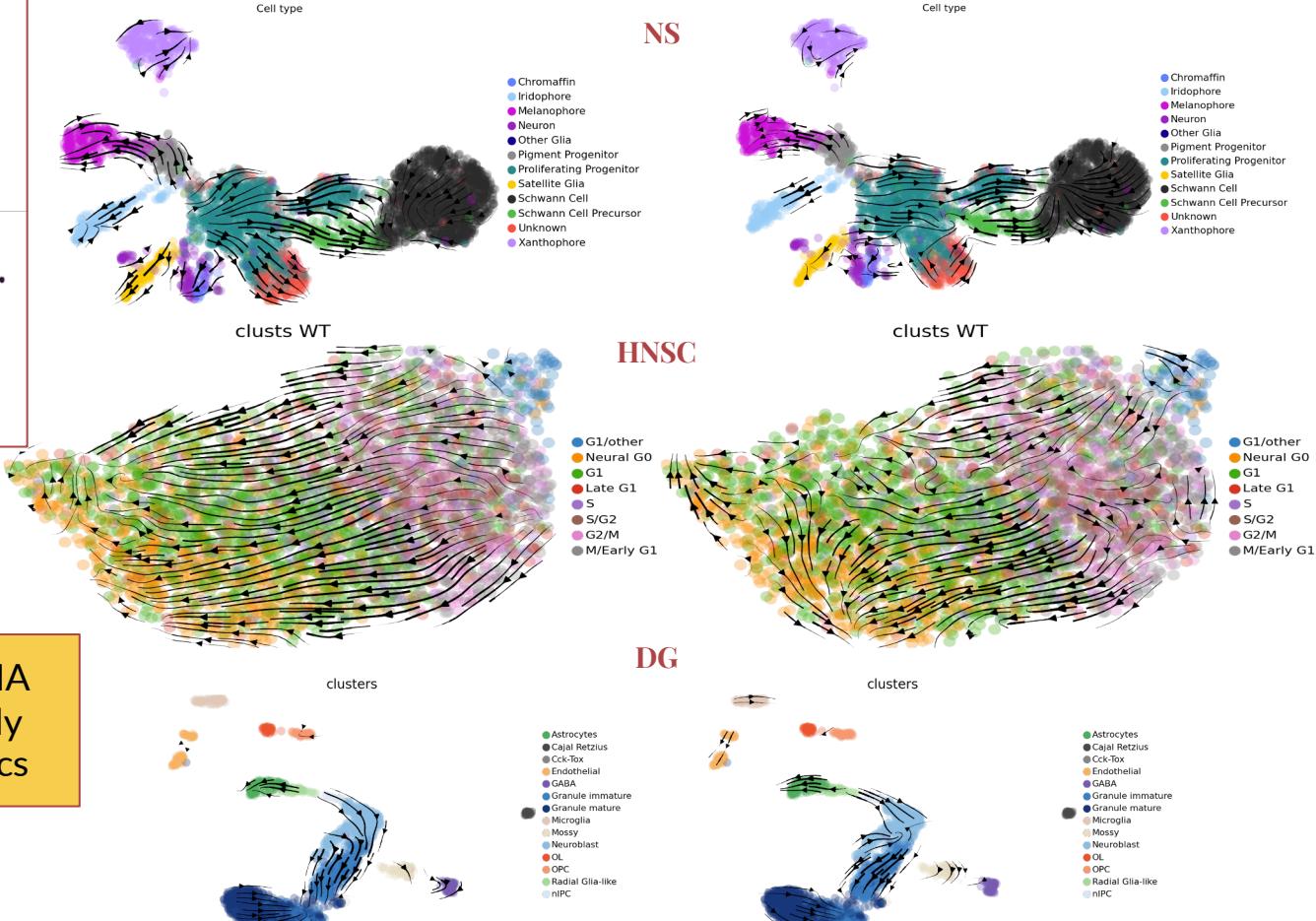
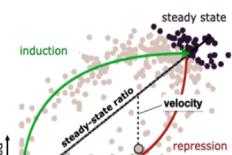
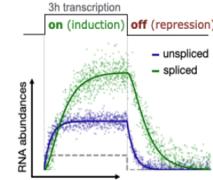
NS

HNSC

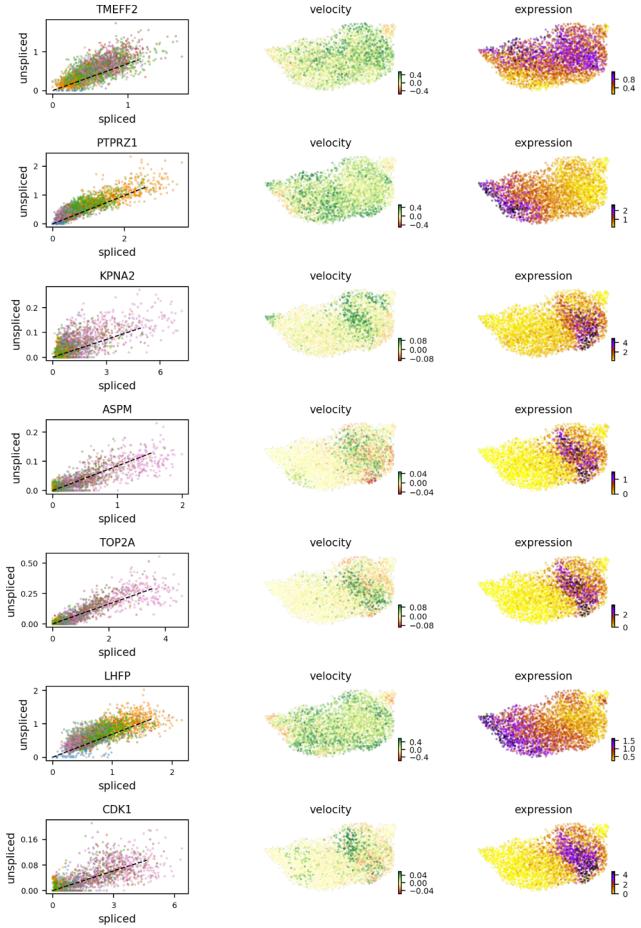


# Stochastic model + Dynamical model

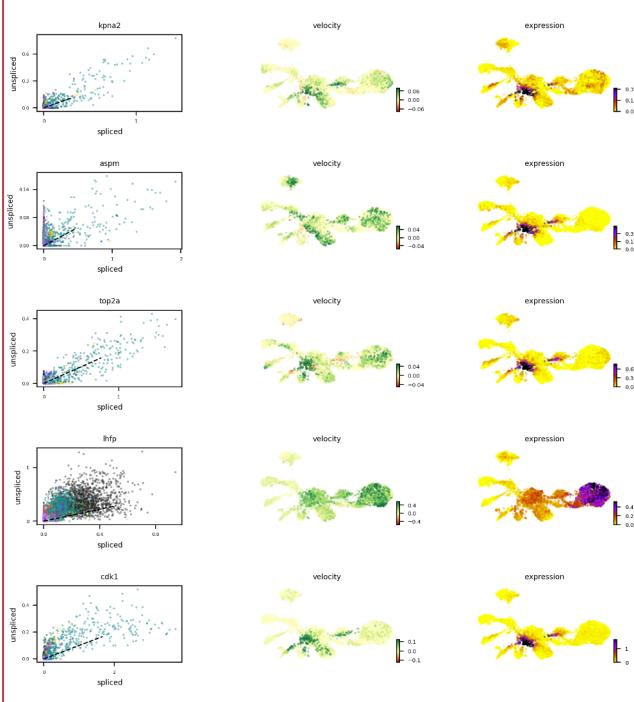
Transcriptional dynamics for a particular gene:



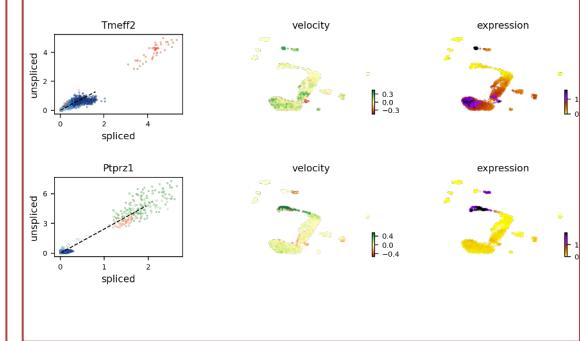
## HNSC



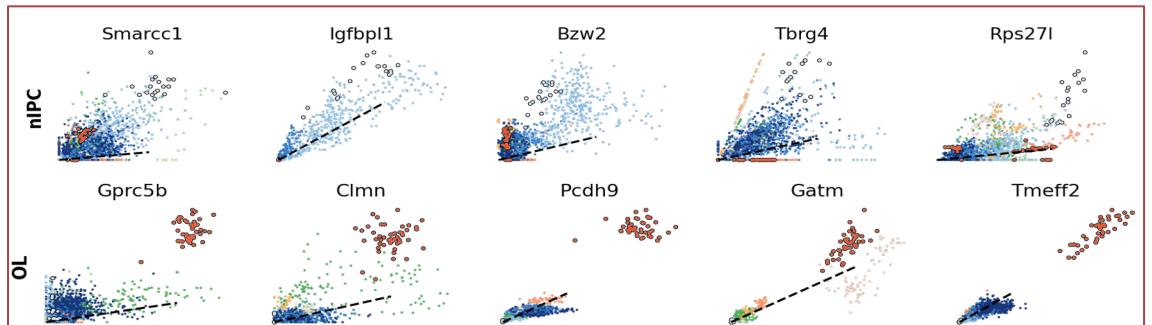
## NS



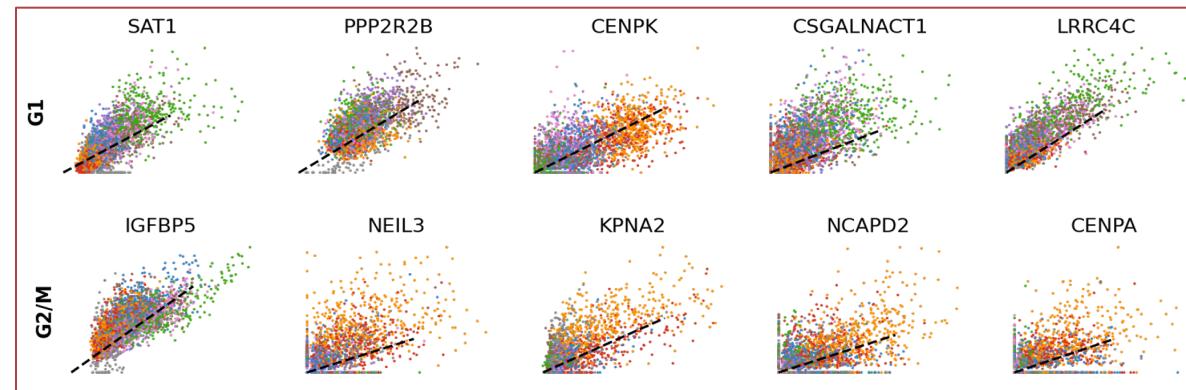
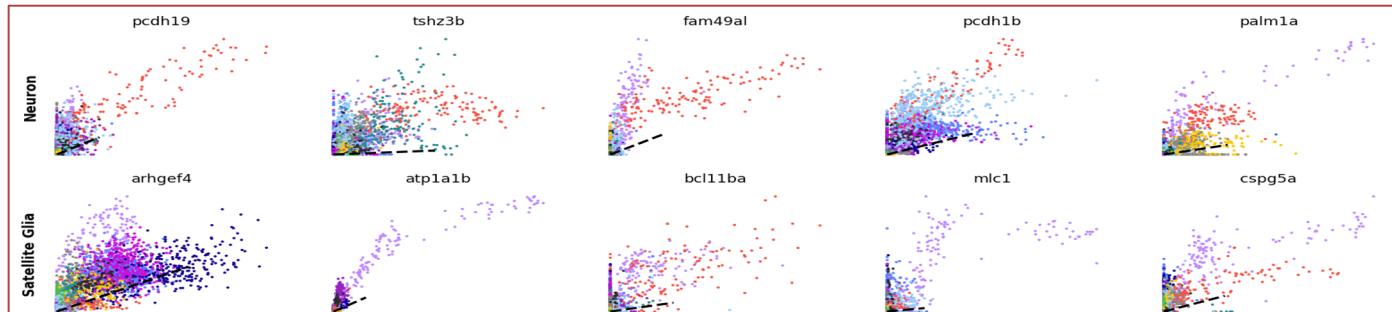
## DG



Phase portraits of some genes



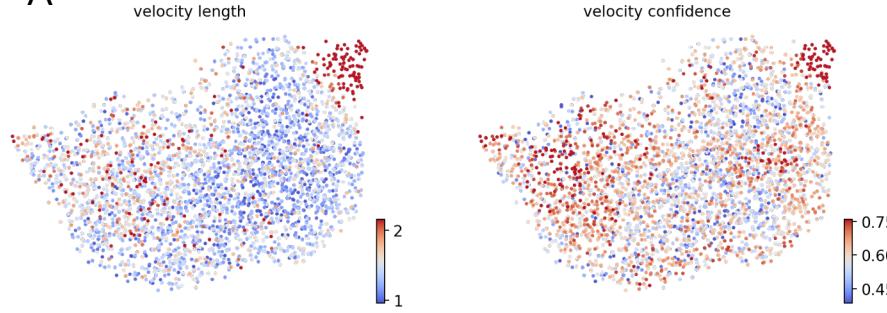
NS



Identify putative driver genes and regimes of regulatory changes.

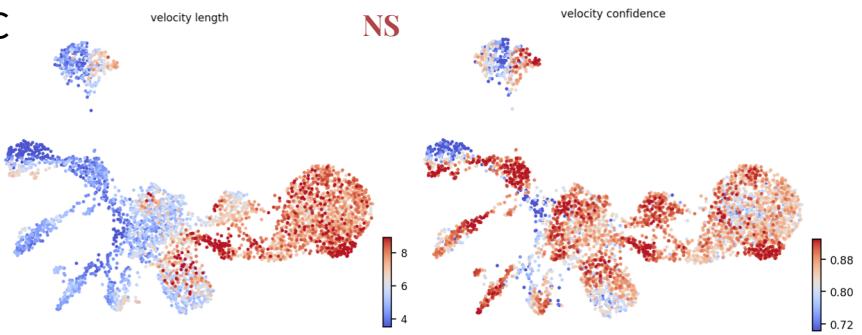
# Velocity length and confidence

A



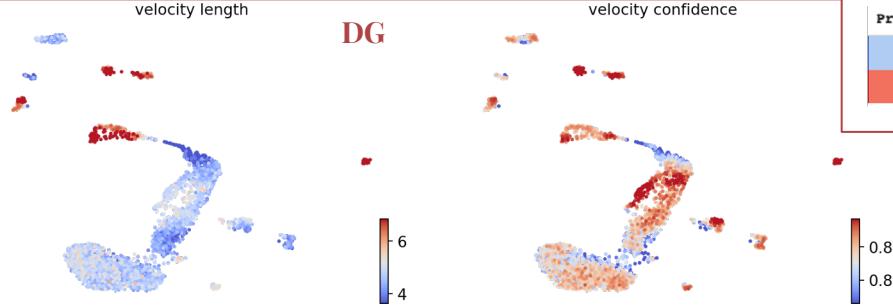
	clsts_WT	G1/other	Neural G0	G1	Late G1	S	S/G2	G2/M	M/Early G1
velocity_length	3.363947	1.507628	1.419545	1.451633	1.368037	1.304428	1.261703	1.392656	
velocity_confidence	0.764893	0.600738	0.592971	0.585594	0.578008	0.592013	0.539647	0.580121	

C



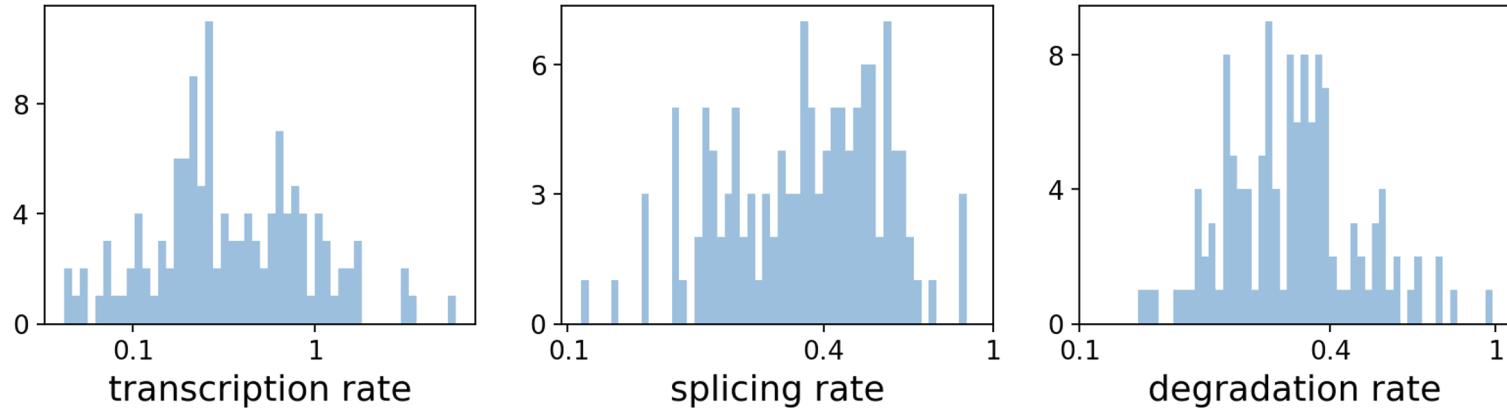
Cell_type	Chromaffin	Iridophore	Melanophore	Neuron	Other Glia	Pigment Progenitor
velocity_length	5.761765	4.529040	4.084706	4.905255	6.025000	4.135232
velocity_confidence	0.874039	0.870789	0.795781	0.809546	0.772622	0.885446
Proliferating Progenitor	Satellite Glia	Schwann Cell	Schwann Cell Precursor	Unknown	Xanthophore	
5.684814	4.399683	7.494294		8.193619	5.960882	4.847516
0.864172	0.874458	0.840751		0.879925	0.819929	0.781090

B



clusters	Astrocytes	Cajal Retzius	Cck-Tox	Endothelial	GABA	Granule immature	Granule mature	Microglia	Mossy	Neuroblast	OL	OPC	Radial Glia-like	nIPC
velocity_length	6.894834	6.836765	5.143333	5.697931	4.050820	4.333694	4.515224	4.281481	4.521067	4.260623	6.957600	6.482264	5.972549	3.084737
velocity_confidence	0.899229	0.983757	0.888597	0.867098	0.877451	0.851931	0.817038	0.861059	0.875730	0.870156	0.943305	0.923774	0.901141	0.749565

# Kinetic rate parameters - for hNSC

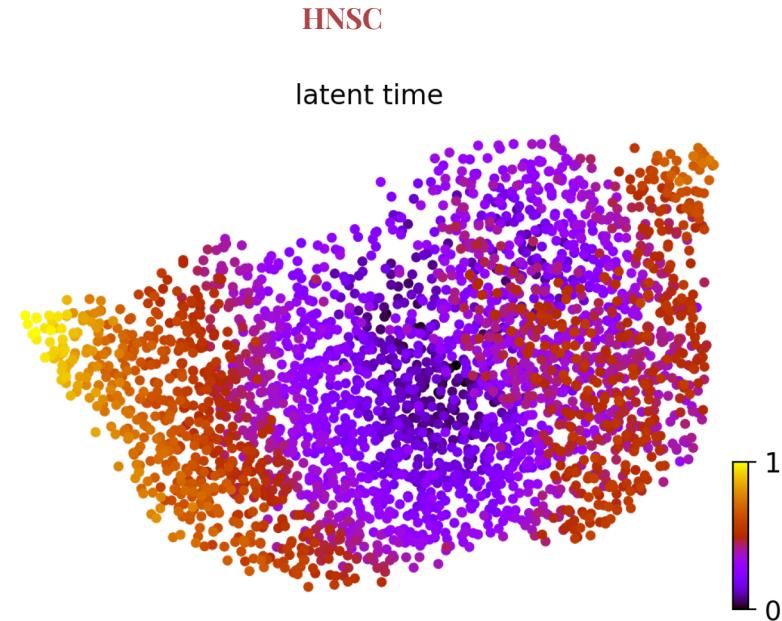
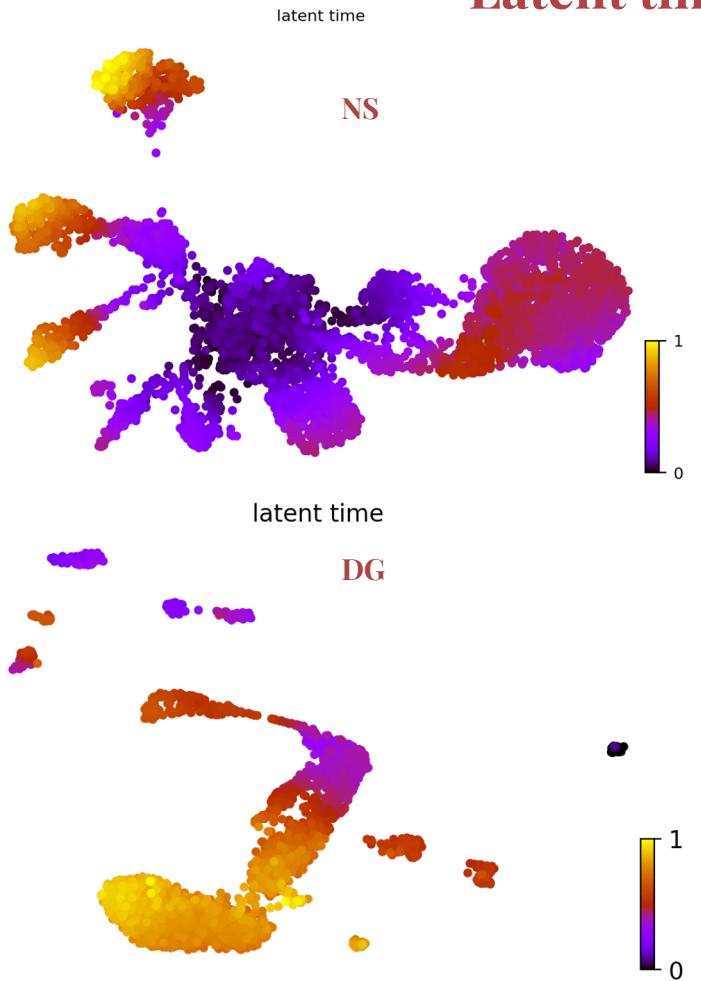


rates of transcription (fit\_alpha), splicing (fit\_beta), degradation (fit\_gamma)

Gene	fit_alpha	fit_beta	fit_gamma	fit_t_	fit_scaling	fit_std_u	fit_std_s	fit_likelihood	fit_u0	fit_s0	fit_pval_steady	fit_steady_u	fit_steady_s	fit_variance	fit_alignment_scaling	fit_r2
CLSPN	0.46	3.61	0.33	6.97	0.07	0.04	0.25	0.22	0.0	0.0	0.42	0.11	0.61	0.72	1.78	0.33
CDCA8	0.59	3.36	0.65	4.80	0.11	0.05	0.20	0.28	0.0	0.0	0.41	0.14	0.50	0.65	1.21	0.68
KIF2C	0.49	3.66	0.62	4.54	0.09	0.03	0.16	0.23	0.0	0.0	0.38	0.10	0.44	0.90	1.37	0.62
DEPDC1	0.90	8.20	0.58	4.21	0.04	0.03	0.28	0.27	0.0	0.0	0.43	0.08	0.68	0.67	1.28	0.65
NEXN	0.27	1.98	0.29	9.82	0.29	0.05	0.22	0.22	0.0	0.0	0.31	0.14	0.77	1.10	2.23	0.15

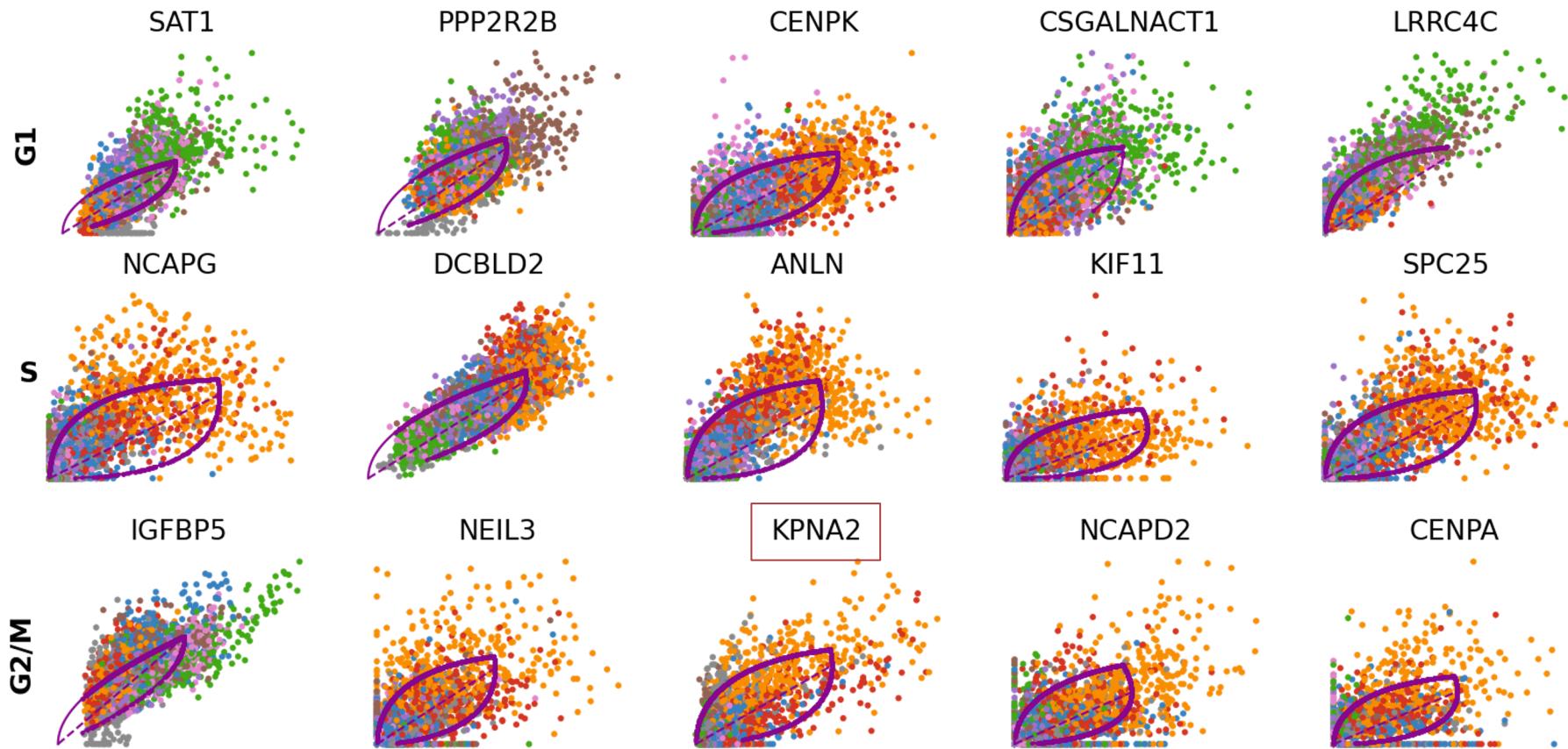
To estimate reaction rates of transcription, splicing and degradation

# Latent time from dynamical model

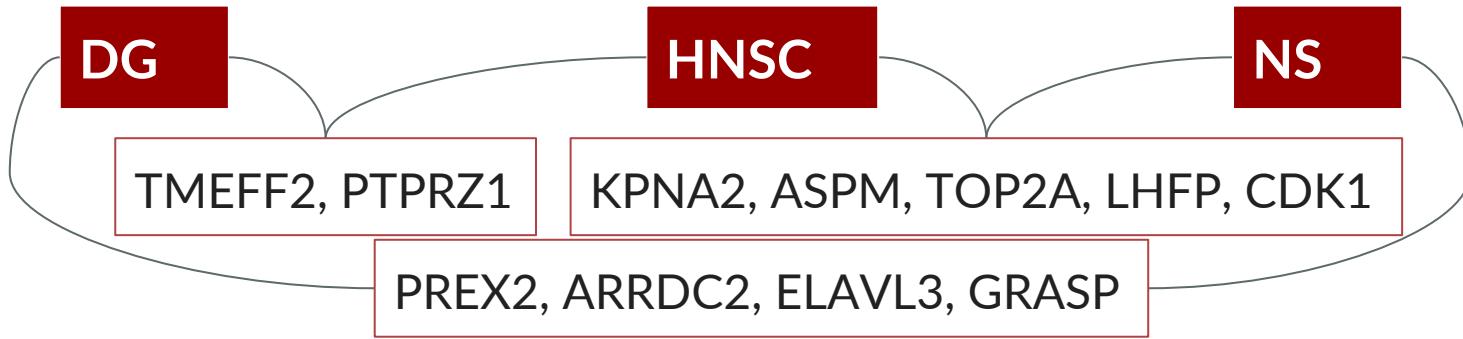


To infer a latent time to reconstruct the temporal sequence of transcriptomic events.

# Cluster specific Top likelihood genes for hNSC



# Overlapping genes



Cell cycles involved	Genes	Dataset	Cell types involved (top 10 only)
Early G1 dominant, S, G2	TMEFF2	Dentate gyrus	Oligodendrocytes
G1 dominant, Neural G0	PTPRZ1	Dentate gyrus	intermediate progenitor cells for neurons
G2/M transition dominant	KPNA2	Neuronal splicing data	neuron
G1, Neural G0 dominant	ASPM	Neuronal splicing data	none which are significant
S/G2 dominant	TOP2A	Neuronal splicing data	chromaffin
M/early G1 dominant	LHFP	Neuronal splicing data	Proliferating progenitor cells
M/early G1 dominant	CDK1	Neuronal splicing data	Satellite glia

# Overlapping genes pathways

Table of top 10 significant p-values and q-values for WikiPathway 2021 Human

	term	p-value	q-value	overlap_genes
	Retinoblastoma gene in cancer WP2446	0.000387	0.006754	[TOP2A, CDK1]
	Spinal Cord Injury WP2431	0.000711	0.006754	[PTPRZ1, CDK1]
	PPAR-alpha pathway WP2878	0.009066	0.033760	[CDK1]
	Gastric Cancer Network 1 WP2361	0.010107	0.033760	[TOP2A]
	Gastric Cancer Network 2 WP2363	0.010801	0.033760	[TOP2A]
	ATM Signaling Pathway WP2516	0.013918	0.033760	[CDK1]
	Integrated Cancer Pathway WP1971	0.015301	0.033760	[CDK1]
ATM Signaling Network in Development and Disease	WP3878	0.015646	0.033760	[CDK1]
Regulation of Microtubule Cytoskeleton	WP2038	0.015992	0.033760	[CDK1]
G1 to S cell cycle control	WP45	0.022189	0.038935	[CDK1]

Table of top 10 significant p-values and q-values for Elsevier Pathway Collection

	term	p-value	q-value	overlap_genes
	Proteins Involved in Glioma	0.000101	0.004551	[ASPM, PTPRZ1, CDK1]
	Proteins Involved in Glioblastoma	0.001606	0.015698	[ASPM, PTPRZ1]
	GPR3 in Oocyte Meiotic Arrest	0.002448	0.015698	[CDK1]
	Metaphase/Anaphase Phase Transition	0.002797	0.015698	[CDK1]
	HPV E4 Contribution to Life Cycle of Virus and Cell	0.002797	0.015698	[CDK1]
	HPV Entry into the Keratinocyte	0.003495	0.015698	[KPNA2]
	FOXM1 Signaling in Prostate Cancer	0.003844	0.015698	[CDK1]
	Proteins Involved in Muscular Dystrophy, Facioscapulohumeral	0.003844	0.015698	[KPNA2]
	SCF/FBXO7 Complex	0.004890	0.015698	[CDK1]
	Chromosome Condensation	0.005239	0.015698	[CDK1]

- **TOP2A (S/G2 pathway) and CDK1 (early G1) which is linked to retinoblastoma.**
- **CDK1 which is essential for G1/S and G2/M phase transitions of eukaryotic cell cycle. This gene is also involved in Lung, PDAC, breast and bladder cancers.**
- **Expression of TOP2A was higher in proliferative subtypes of breast cancers such as triple negative and HER2-enriched diseases than luminal type.**

# Takeaways

1. Computed RNA velocity for 3 brain datasets from human and mouse using the stochastic and dynamical model.
2. Velocity length used to find the maximum rate of differentiation
3. Found overlapping genes between the 3 datasets from the analysis by finding the top ranked genes. These genes have important roles in development and cell cycle processes.
4. Understood phase portrait of some overlapping genes

## Limitations:

1. The steady state model makes two central assumptions:
  - a. a common splicing rate across genes and
  - b. the presence of at least partial observation of the steady state expression levels in the sampled data.
2. Dynamical model extends velocity estimation to transient systems.