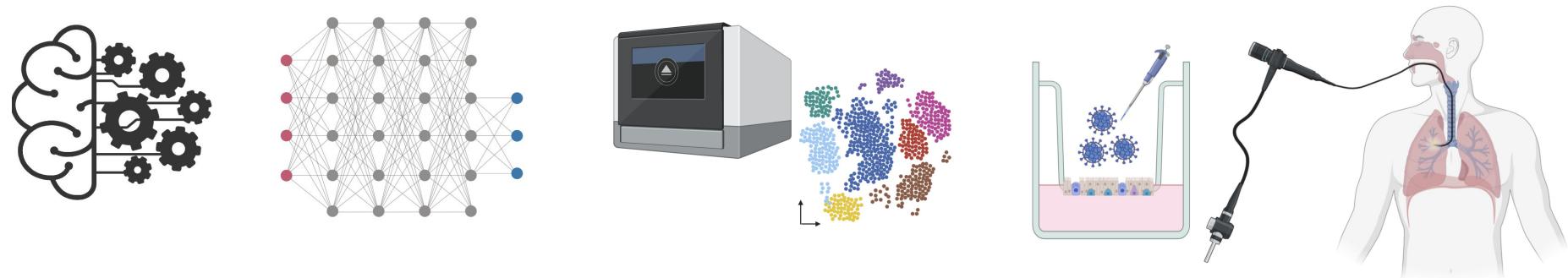
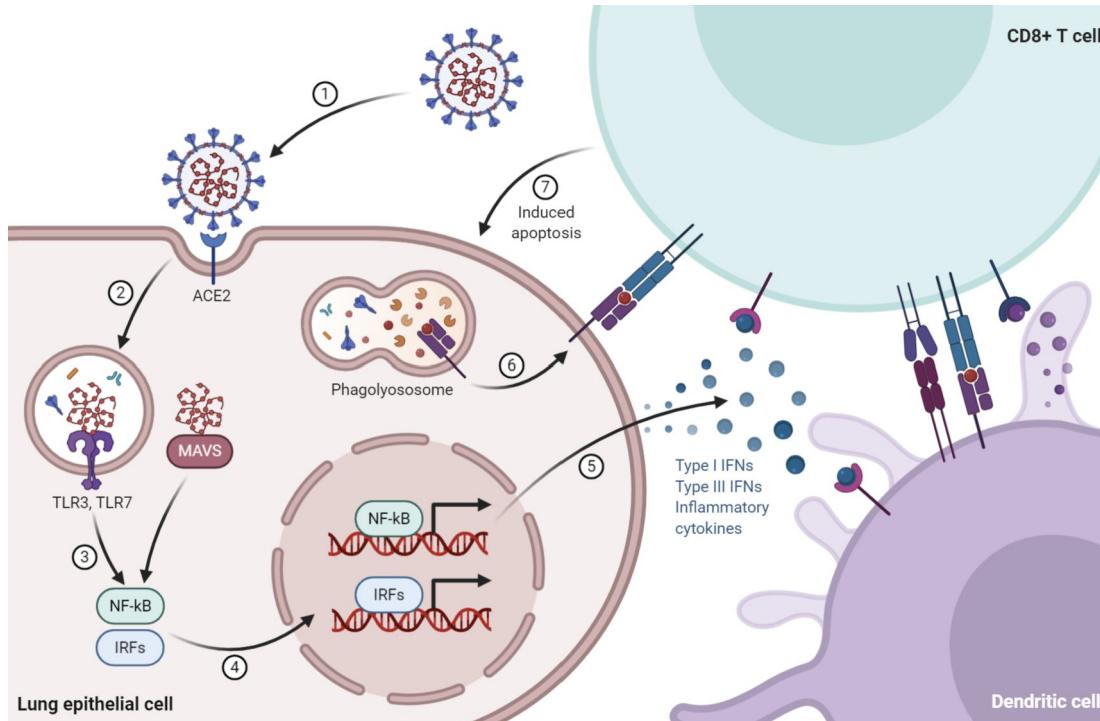


Identifying transcripts associated with severe COVID-19 and SARS-CoV-2 cell tropism using scRNA-seq



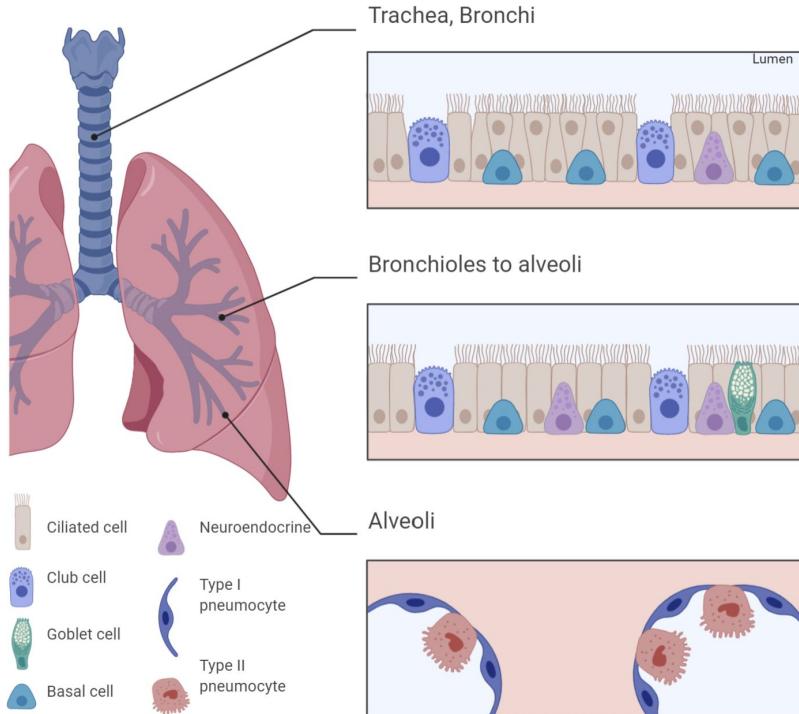
Neal G. Ravindra, Postdoctoral Fellow in the van Dijk Lab
Yale CVRC RIP talk, 30 September 2020

SARS-CoV-2 pathogenesis

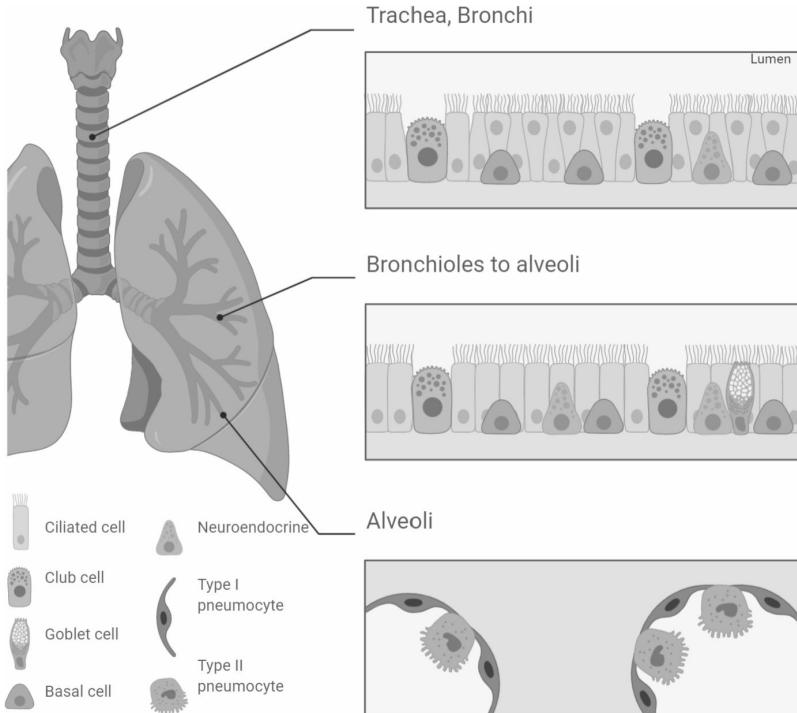


- Cell tropism
- Transcriptomic variability
 - Bystander v. infected
- Evolution of response

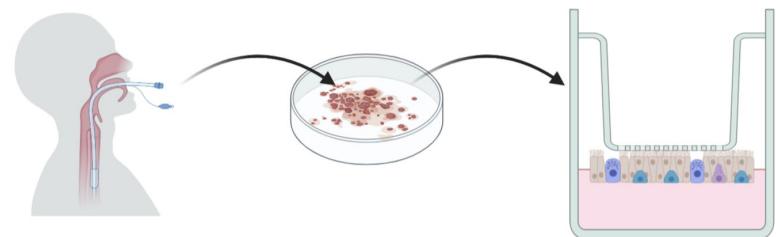
Human bronchial epithelium cells (HBEC)



Human bronchial epithelium cells (HBEC)

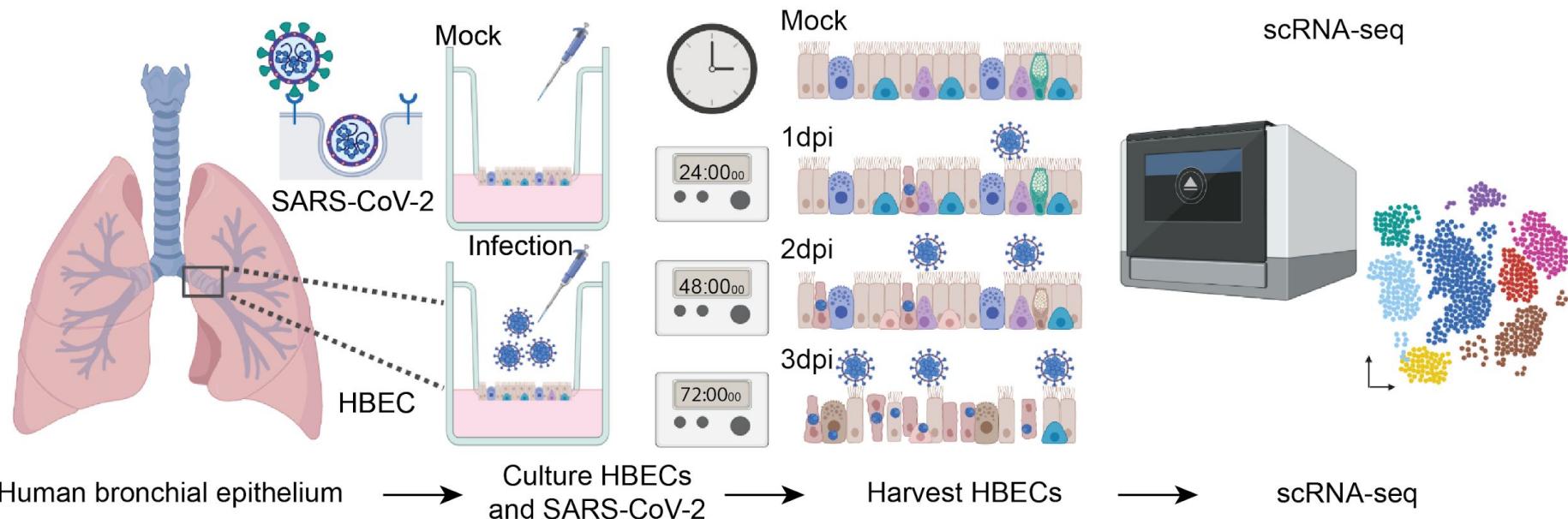


Air-liquid interface cultures (organoids)

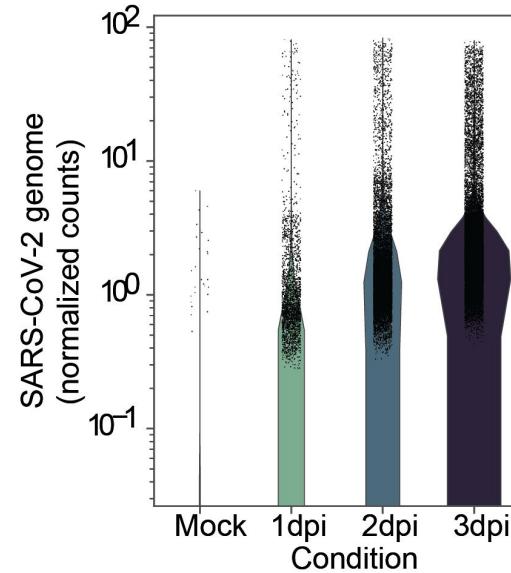
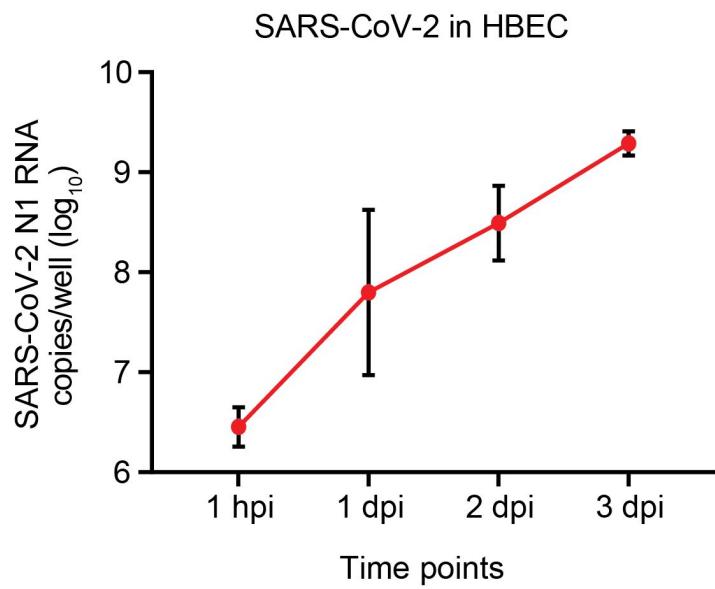


- High-quality scRNA-seq data
- Controlled time course

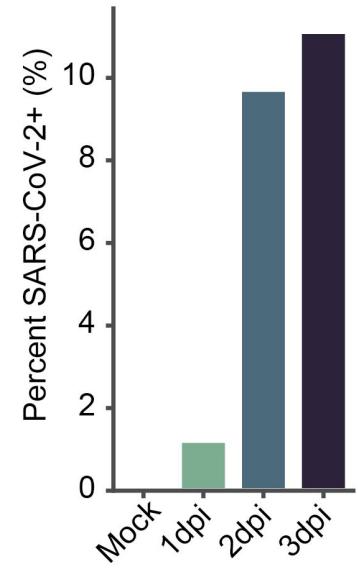
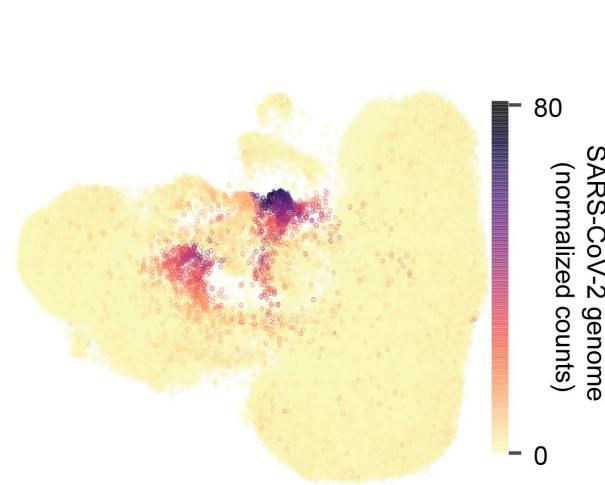
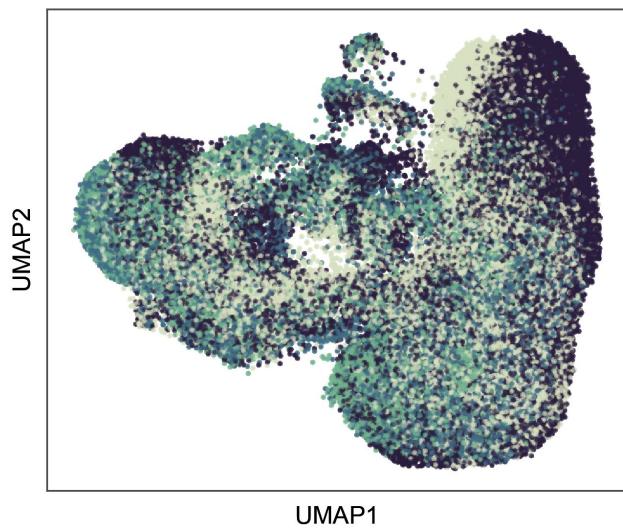
Longitudinal analysis of SARS-CoV-2 infection



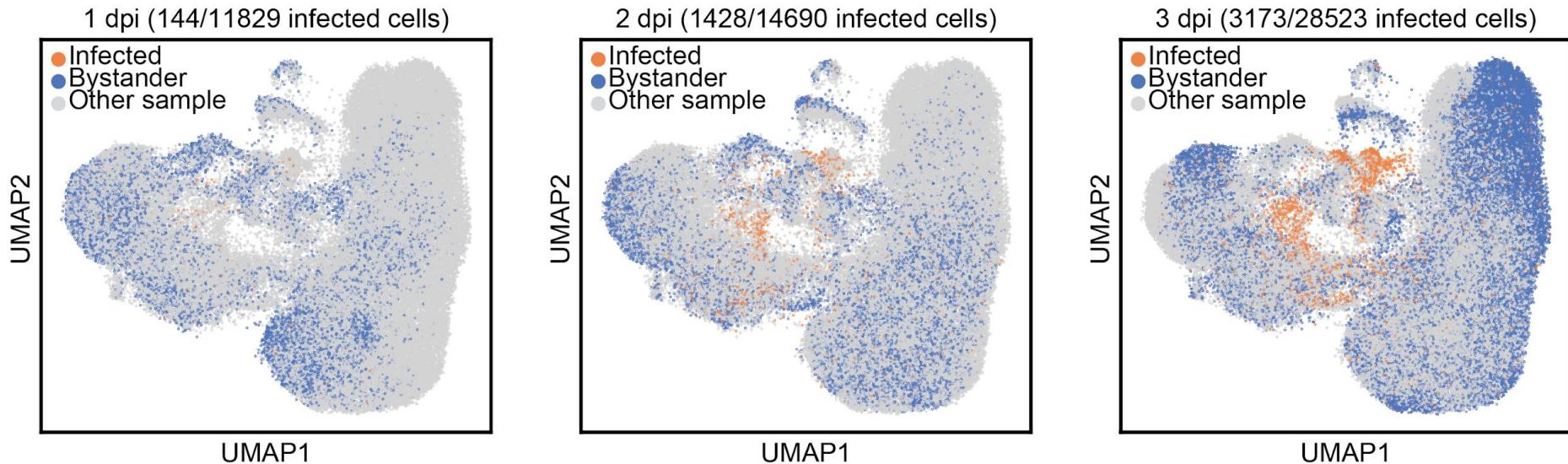
Viral transcripts are detectable by scRNA-seq



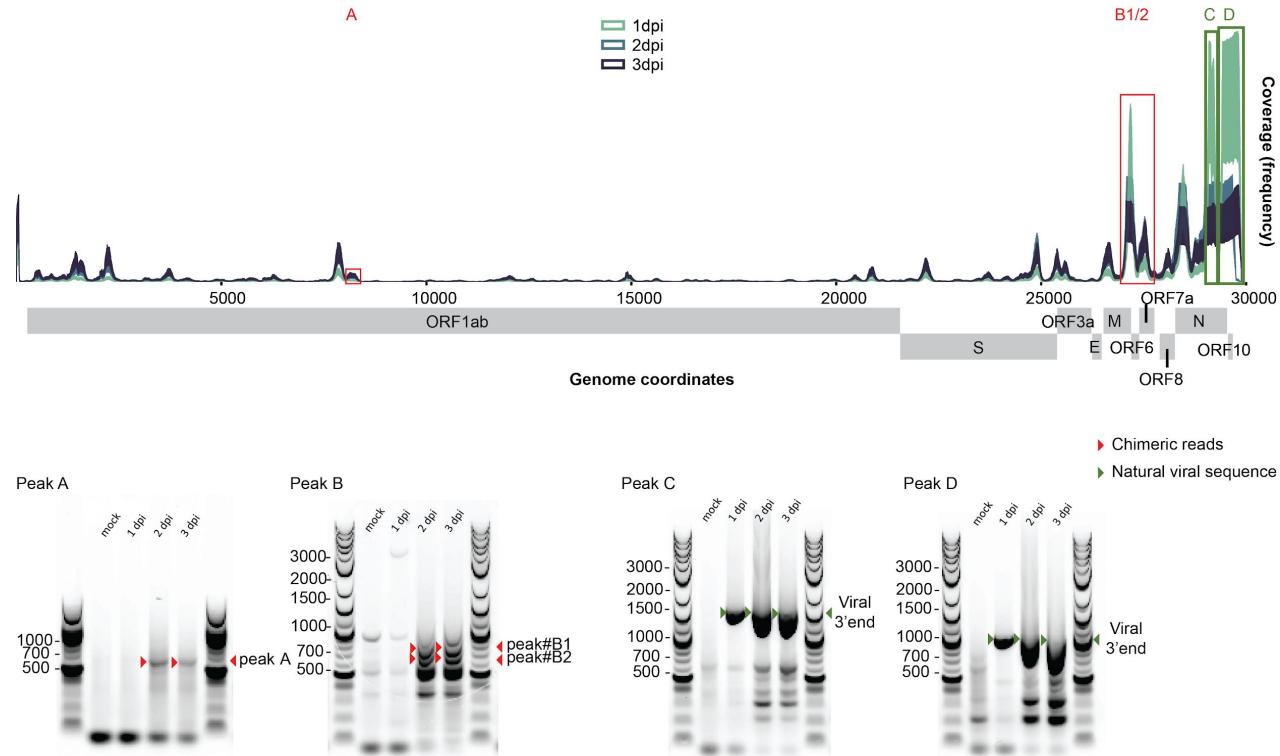
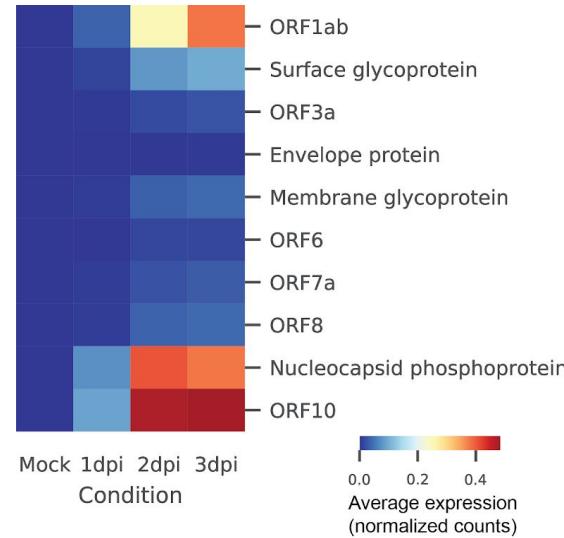
SARS-CoV-2's genome counts to label infected cells



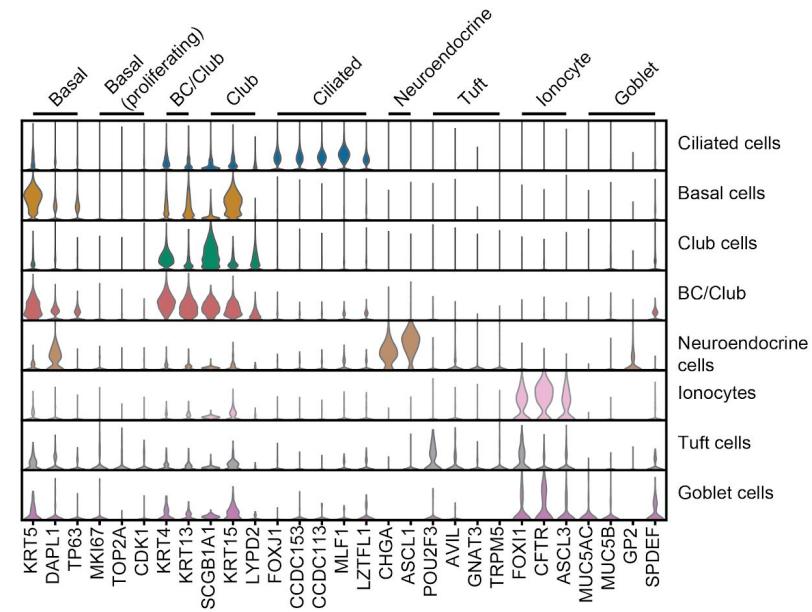
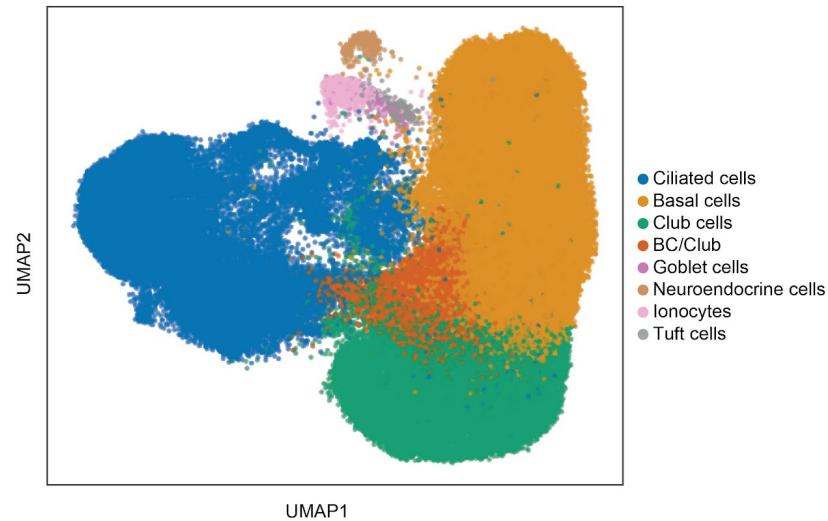
Studying evolution of response to SARS-CoV-2



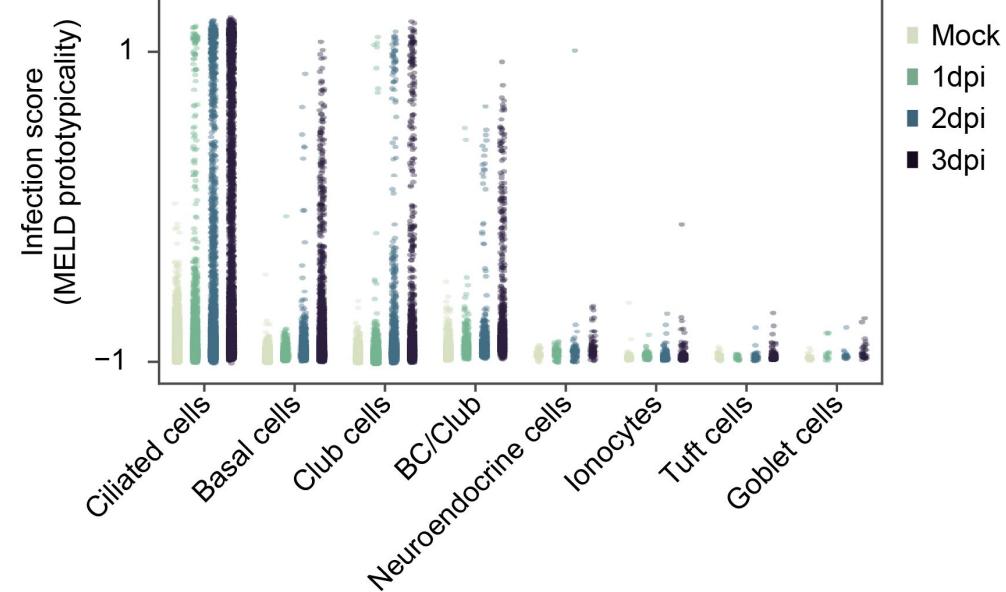
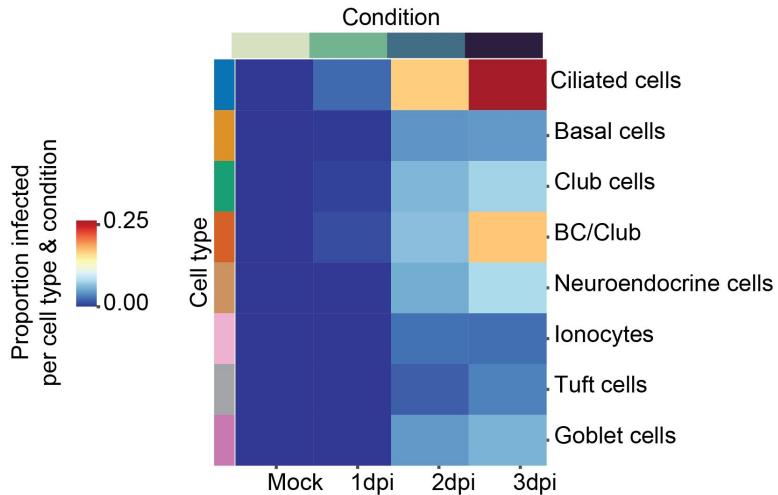
Sub-genomic viral transcripts: productive infection?



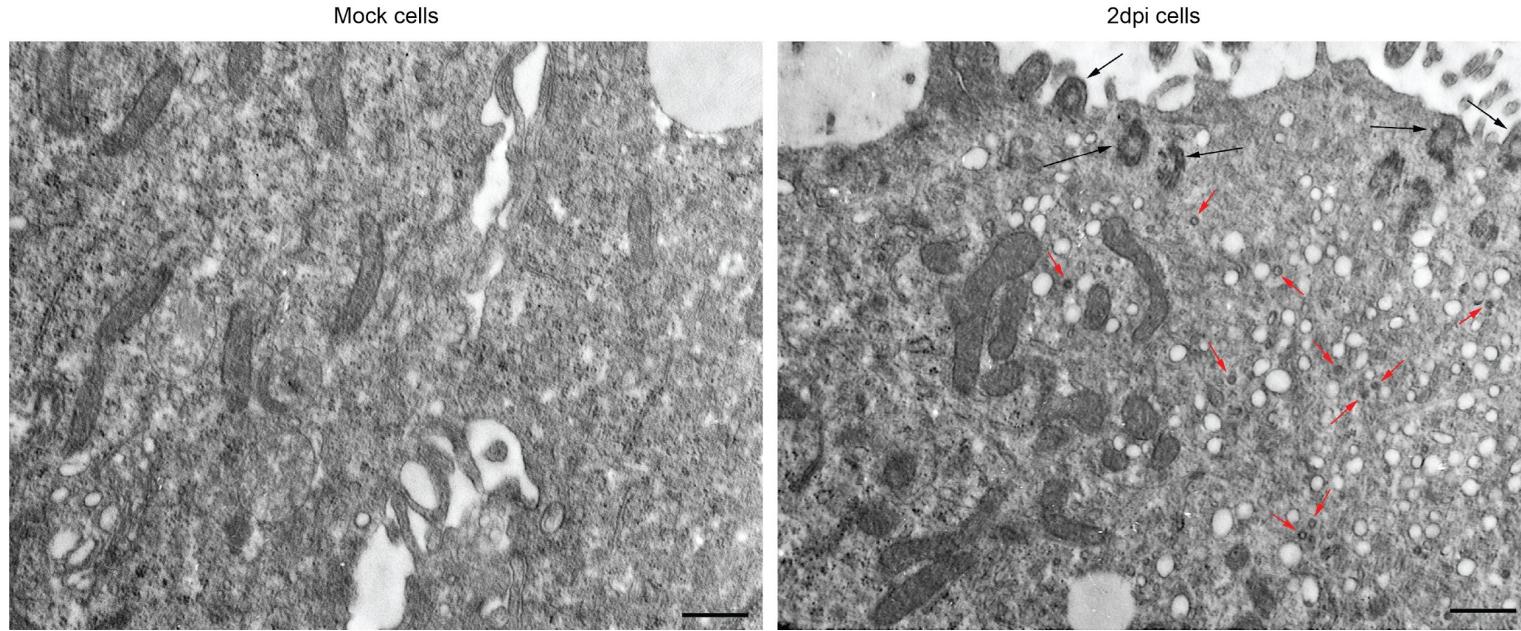
Ciliated, basal, and club cells are abundant



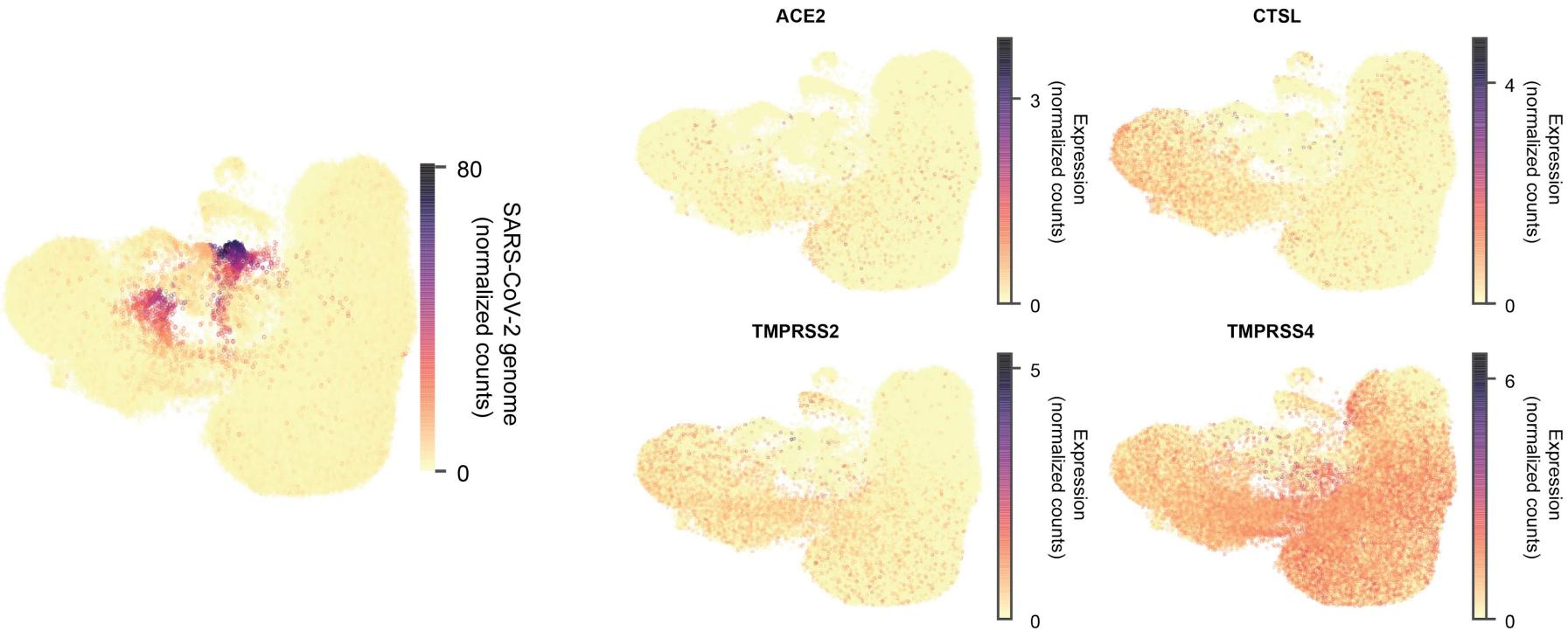
SARS-CoV-2 cell tropism: ciliated cells



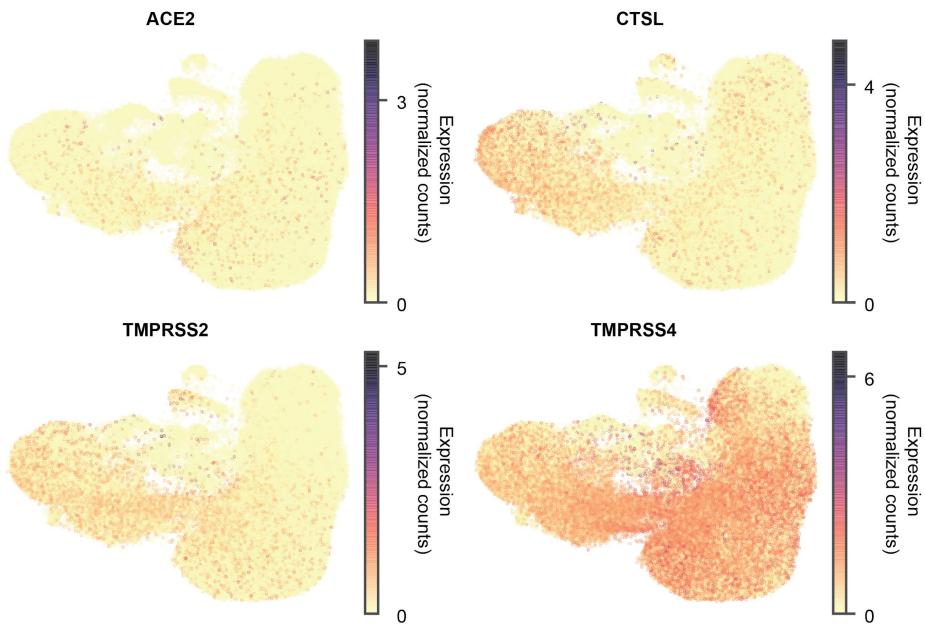
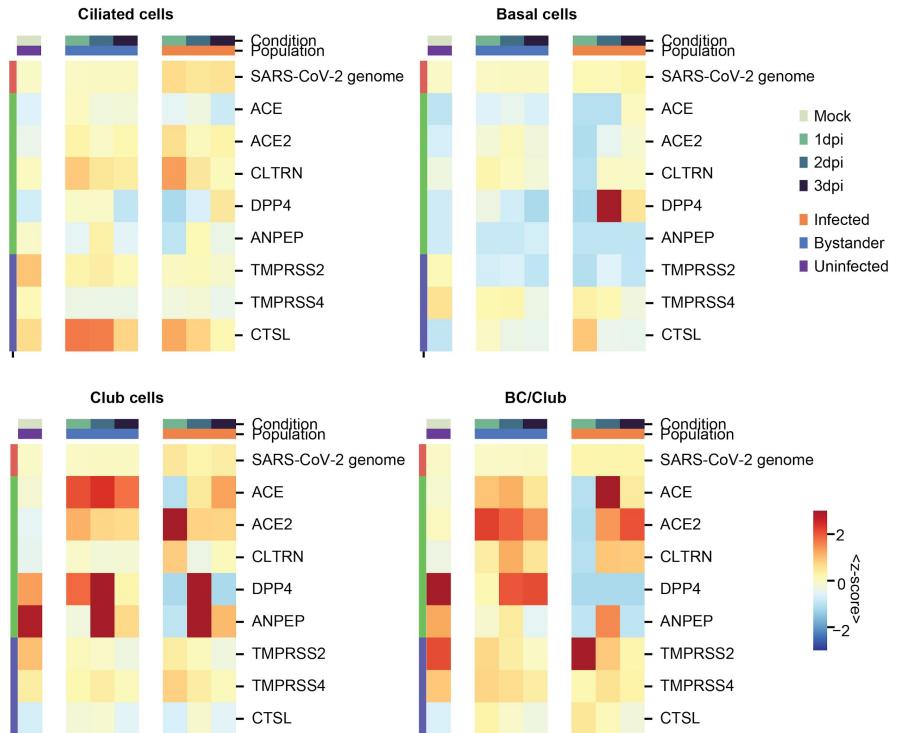
EM validation of SARS-CoV-2 ciliated cell tropism



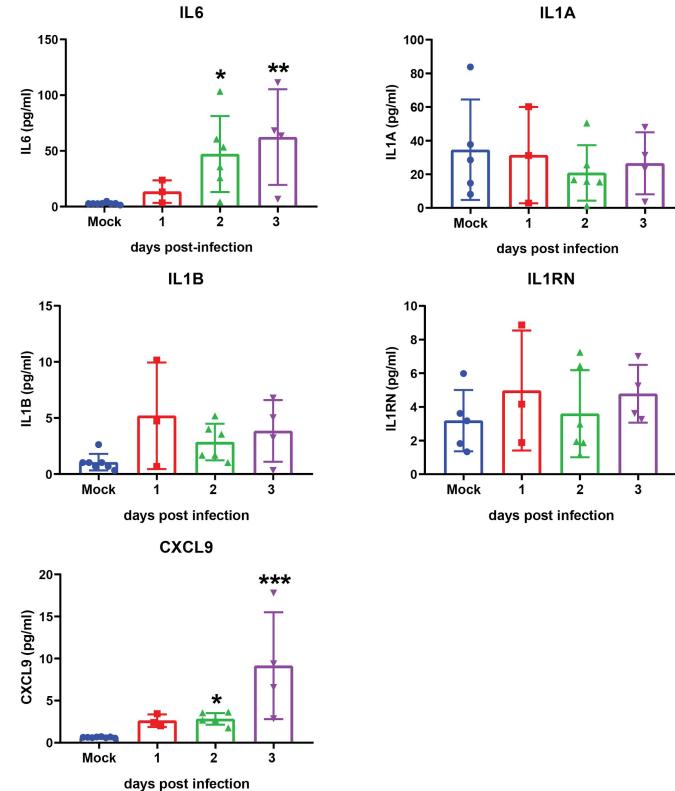
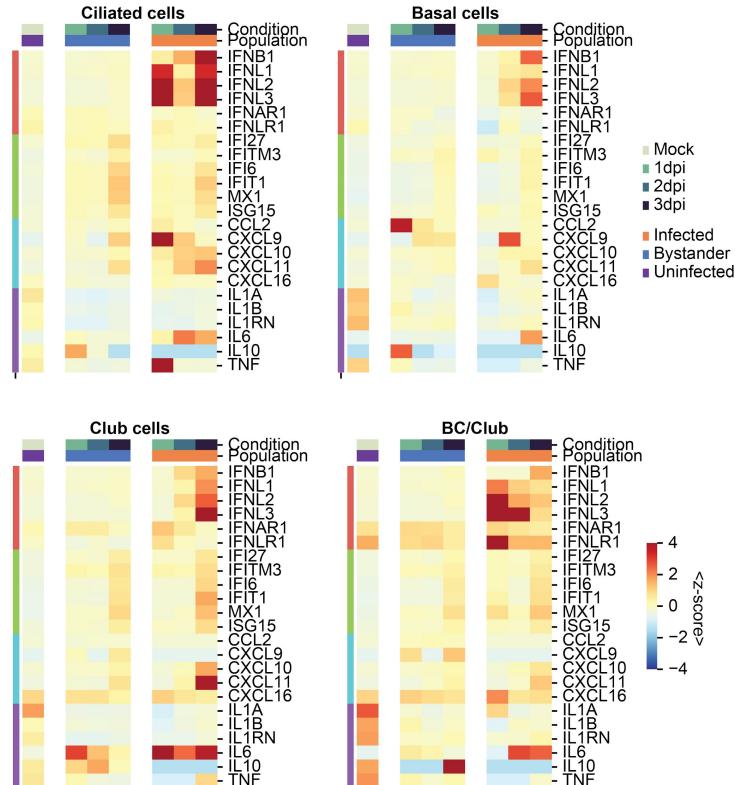
Limited signal from known viral entry factors



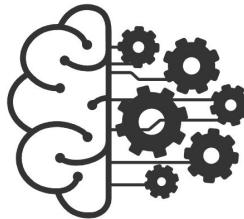
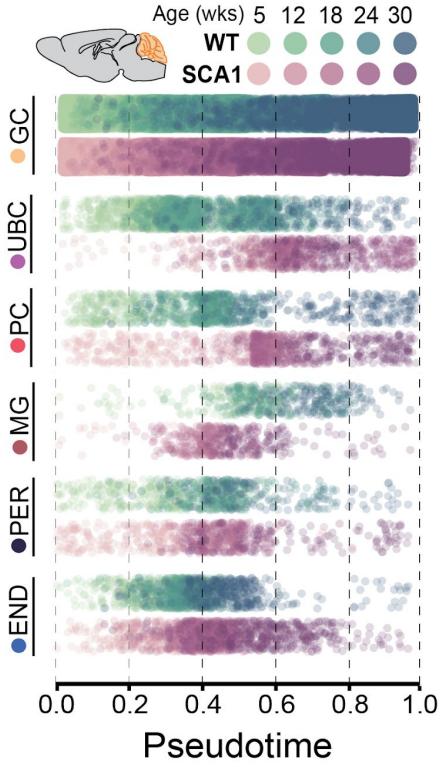
Known entry determinants and infection status



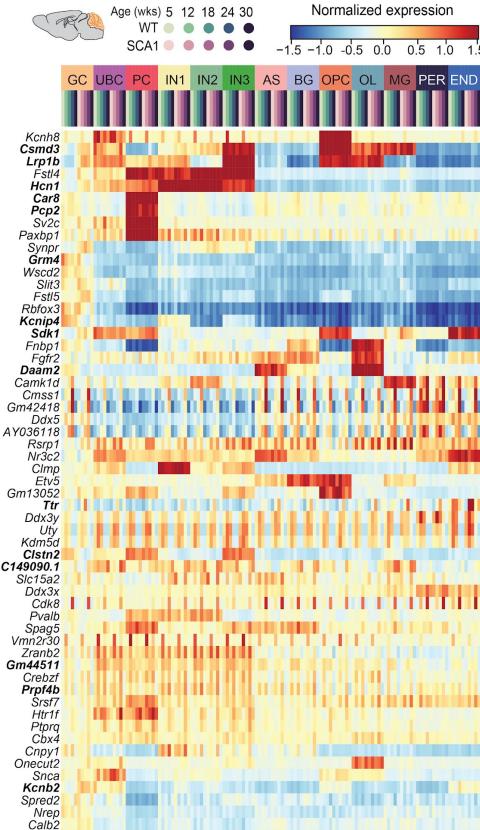
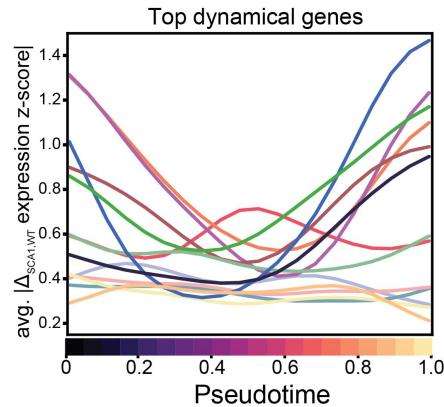
SARS-CoV-2 induces innate immune response



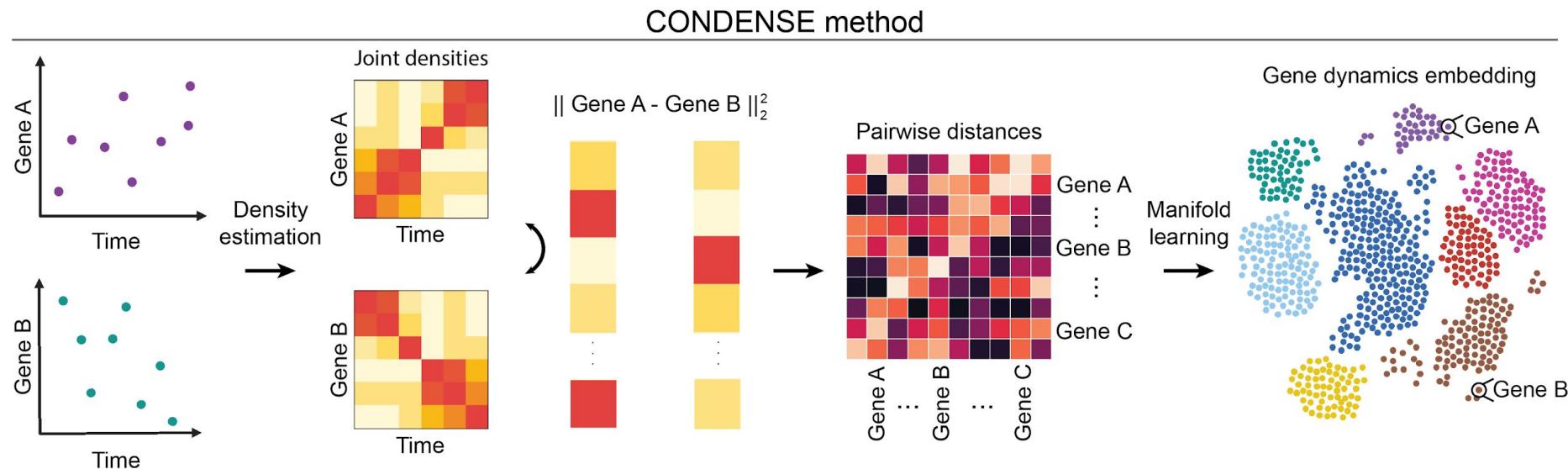
MELD pseudotime & ML to ID dynamical genes



Gradient boosting + MAGIC

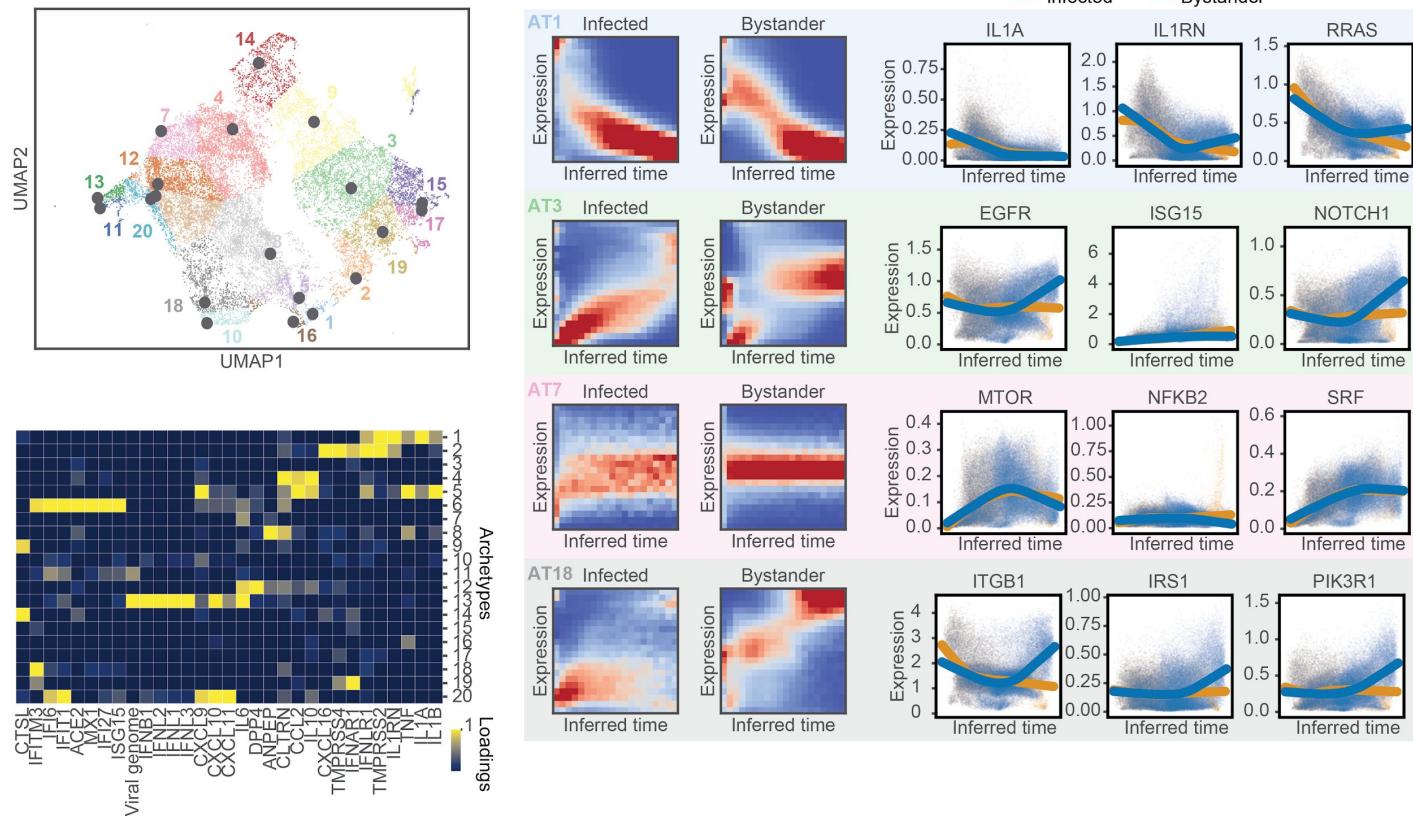


New method to account for independent timepoints

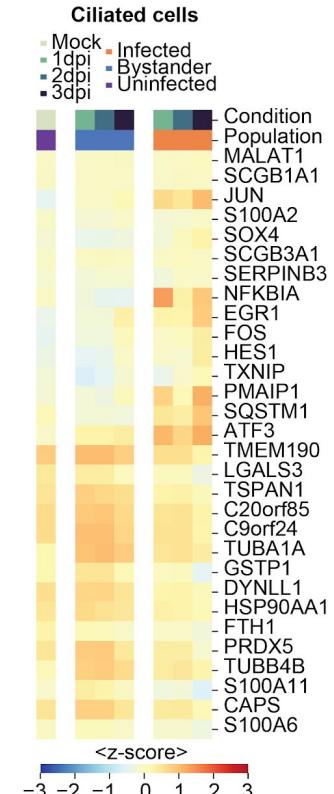
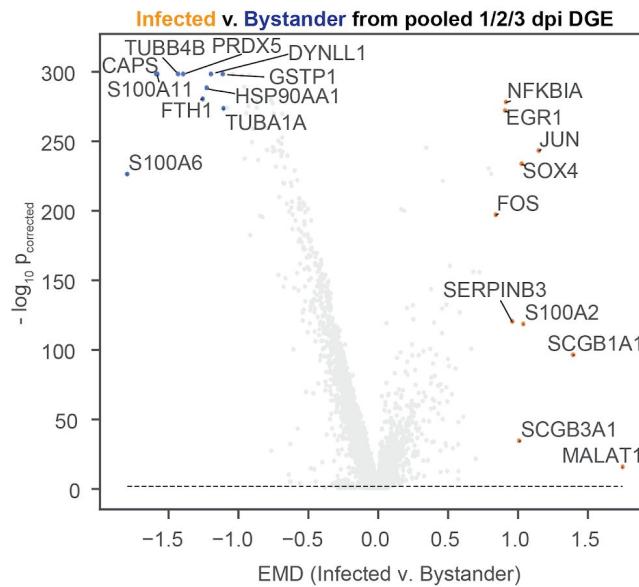
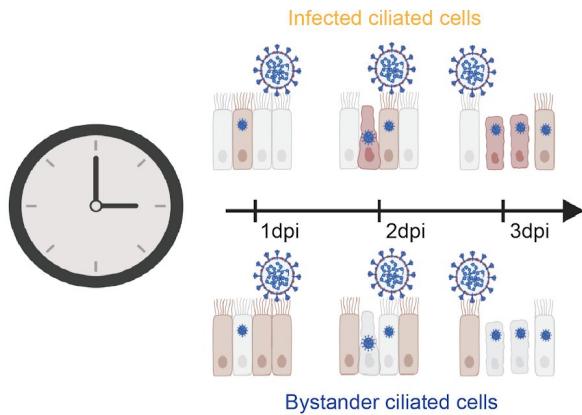


CONDENSE = CONditional DENSity Embedding

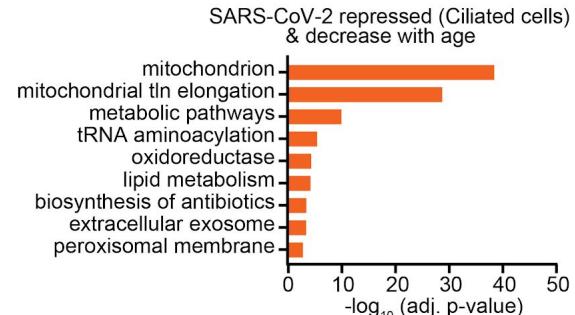
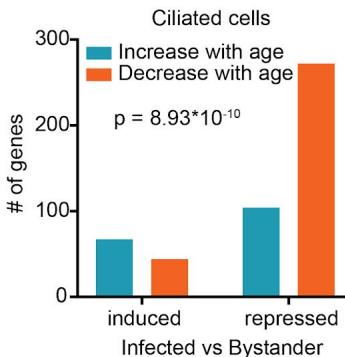
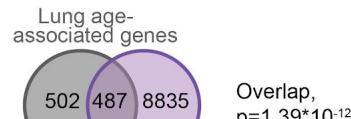
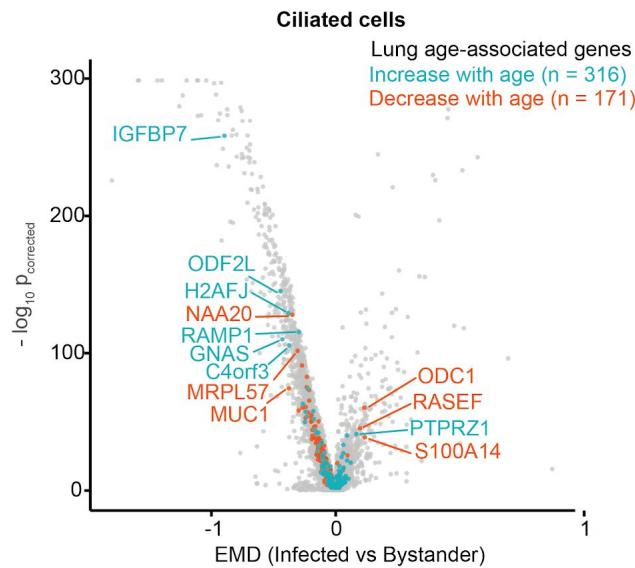
CONDENSE contrasts variation in gene dynamics



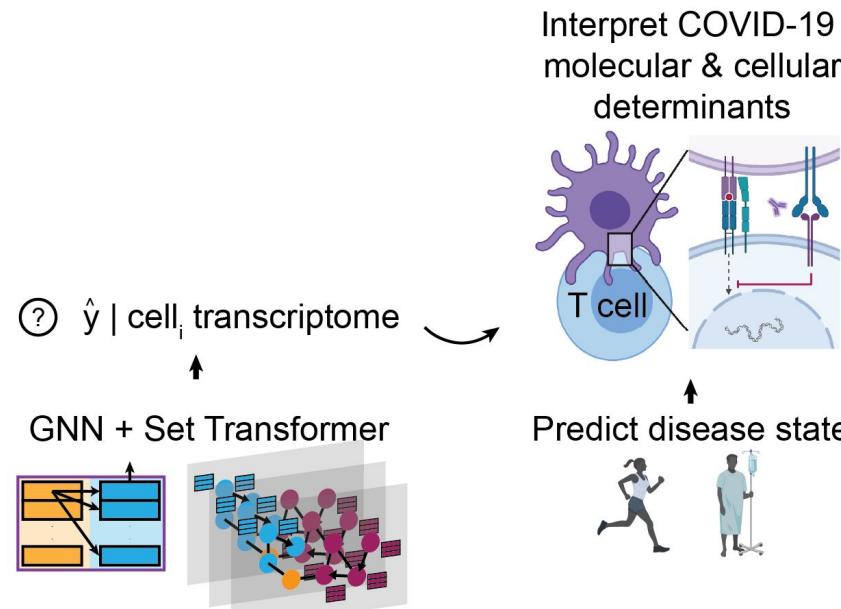
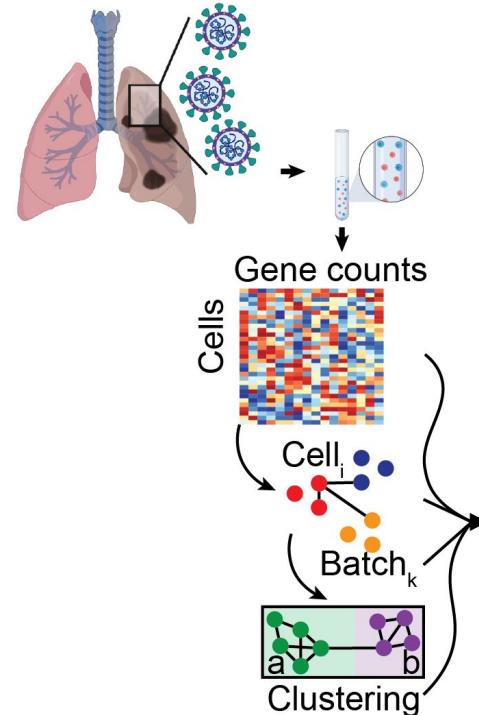
SARS-CoV-2 uniquely downregulates cilium function



SARS-CoV-2 trends in lung-age associated genes



DL to find transcriptomic determinants of metadata



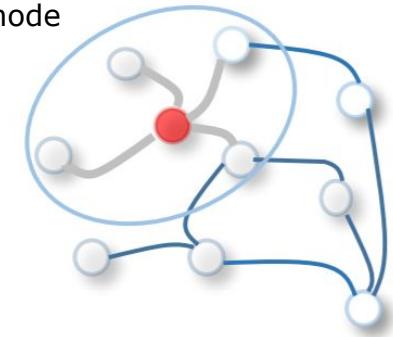
Graph neural networks

Used for tasks such as node classification, link prediction, and graph classification

Have been applied to EHR data, predicting protein-protein interactions, molecular structure & reactivity

No prior work prior predicting disease states from single-cell data

Graph convolution to get hidden representation of red node



From Wu et al. survey,
2019

Graph attention networks

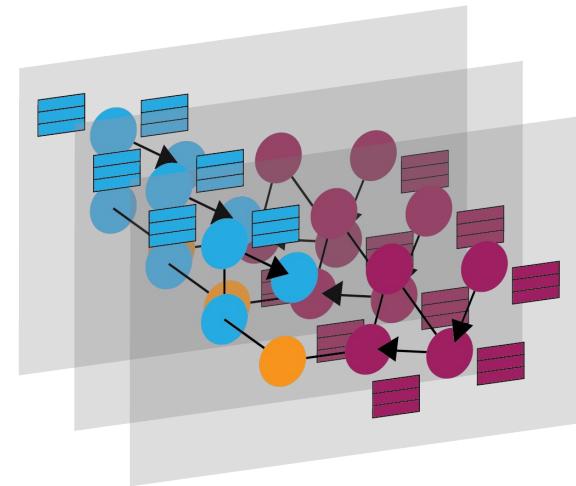
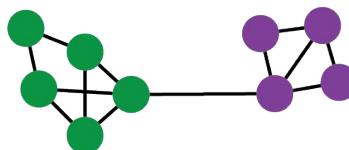
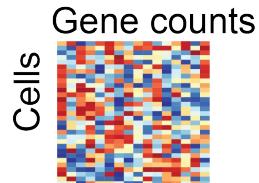
$$h = \{h_1, h_2, \dots, h_N\}$$

GAT layer
$$h'_i = \left\|_{l=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^l \mathbb{W}^l h_j \right) \right\|$$

$$h' = \{h'_1, h'_2, \dots, h'_N\}$$

Compute node self-attention to get multi-head edge attention coefficients over first-order neighborhood

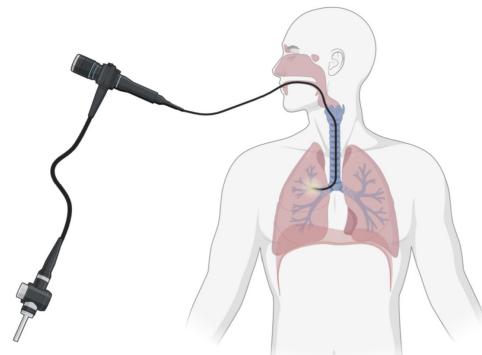
$$\alpha_{ij} = \text{softmax}_j(a(\mathbb{W}h_i, \mathbb{W}h_j))$$



Message passing
GNNs

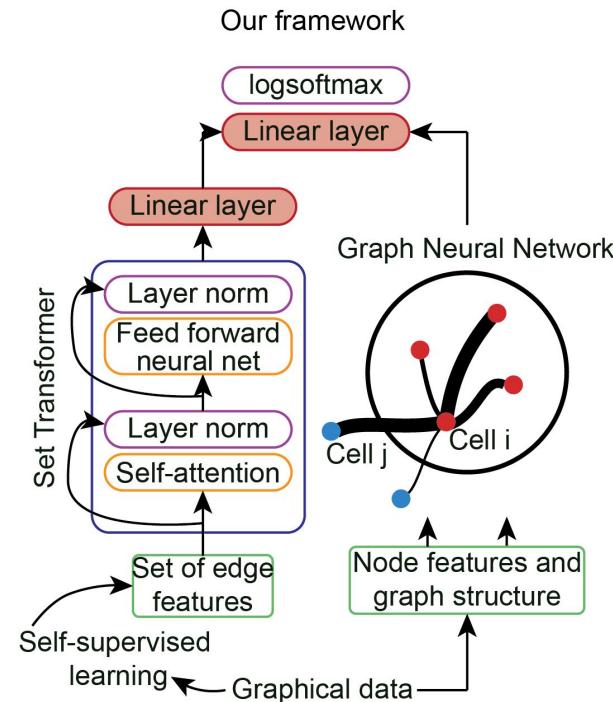
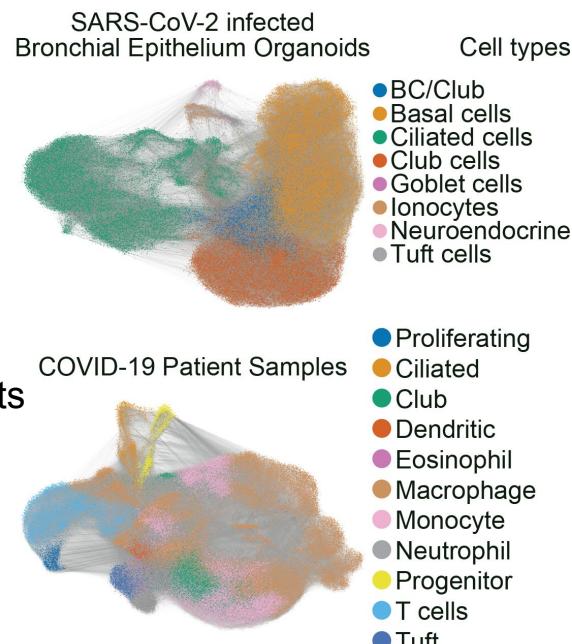
Kipf & Welling, *ICLR* 2017
Veličković et al., *ICLR* 2018

Graph construction & a new DL framework



12 BALFs of healthy/moderate/severe pts
(Liao et al., *Nat. Medicine* 2020)

Datasets	SARS-CoV-2 infected organoids	COVID-19 patients
# Nodes	54353/11646/11648	63486/13604/13605
# Node features	24714	25626
# Edges	1041226/230429/228630	2746280/703217/707529
# Edge features	18	18
# Classes	7	3



Ravindra NG*, Sehanobish A*, ... van Dijk, D. *ACM CHIL 2020*

Sehanobish, A*, Ravindra NG*, van Dijk, D. *ICML GRL+ workshop, 2020*

Sehanobish A*, Ravindra NG*, van Dijk, D. (*under review*; preprint, arXiv:2007.04777)

Improvement in accuracy with our framework

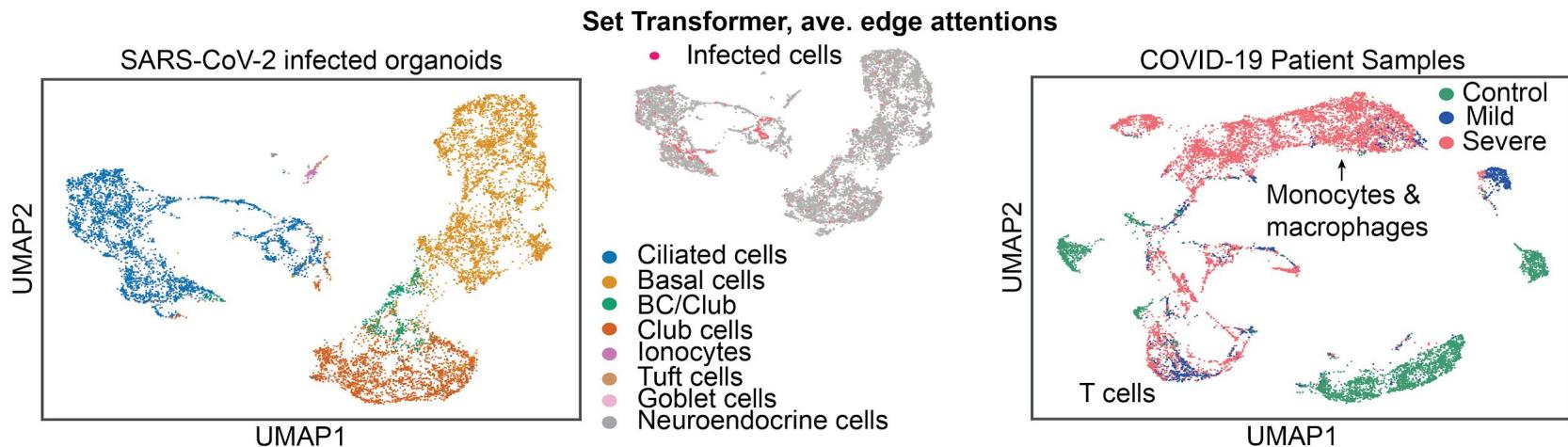
vs other ML approaches (MS dataset)

Task	Model	Accuracy
Inductive	Random	51.8
	MLP	56.7
	Random Forest	58.5
	Graph Convolutional Network	72.1
	Graph Attention Network(our)	92.3 ± .7
Transductive	Graph Convolutional Network	82.91
	Graph Attention Network(our)	86 ± .3

vs other geometric DL approaches

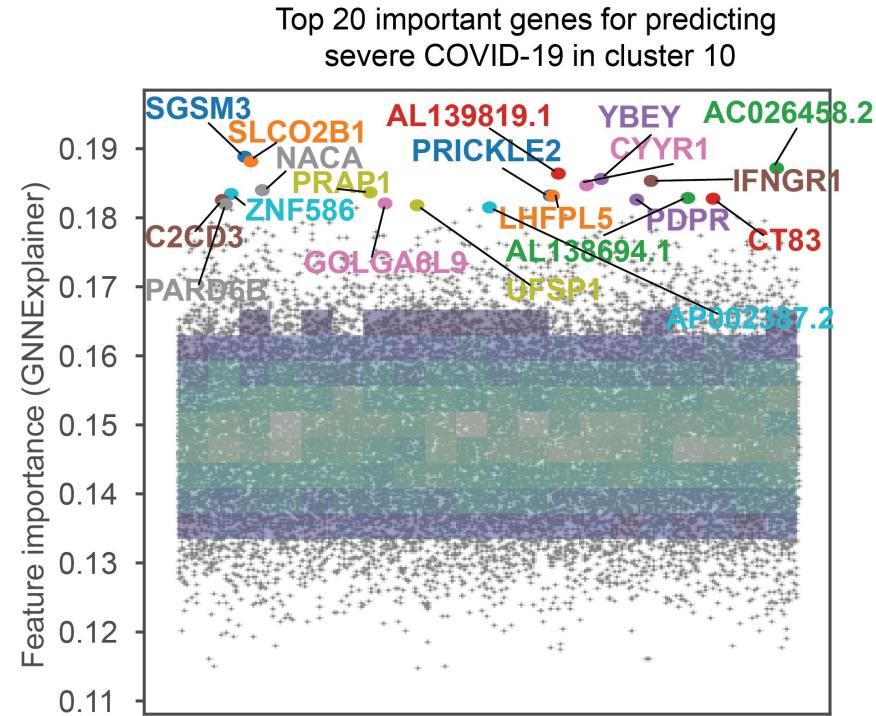
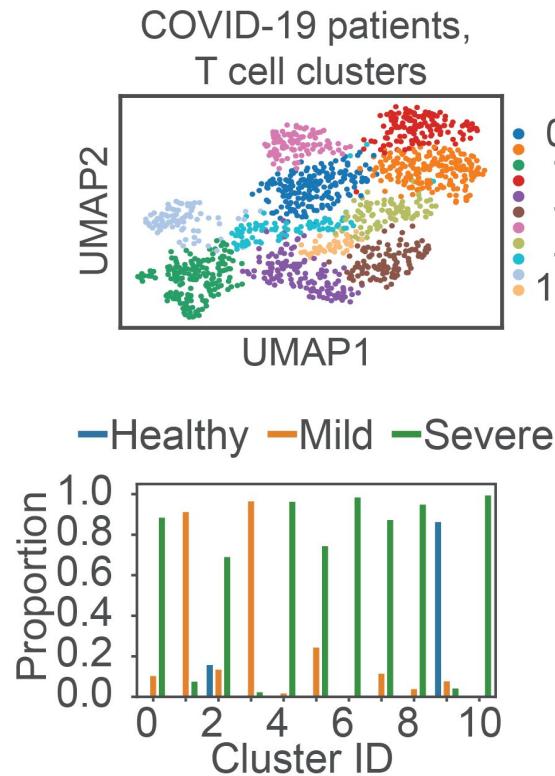
Models	SARS-CoV-2 infected organoids	COVID-19 patients
GCN	65.43 (65.21-65.65)	89.26 (89.06-89.47)
GCN + DeepSet	79.75 (78.75-80.75)	87.2 (87.02-87.38)
GCN + Set2Set	71.65 (69.89-73.42)	88.34 (87.89-88.79)
GCN + Set Transformer	81.61 (79.34-83.87)	92.84 (91.95-93.74)
GAT	73.10 (70.93-75.27)	92.25 (91.27-93.24)
GAT + DeepSet	79.45 (77.98-80.92)	75.99 (74.8-77.68)
GAT + Set2Set	82.95 (81.75-84.15)	92.87 (92.62-93.12)
GAT + Set Transformer (Ours)	89.8 (88.89-91.71)	95.12 (94.02-96.22)
GIN + EdgeConv ¹	63.36 (62.53-64.19)	89.56 (88.54-90.58)
EdgeConditionedConvolution ¹	46.15 (34.72-57.59)	88.63 (86.07-91.20)

New approach to manifold learning with attention



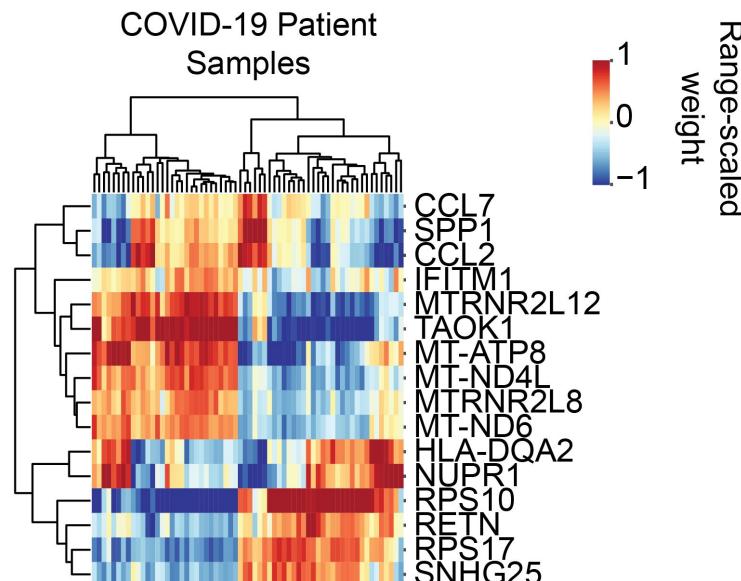
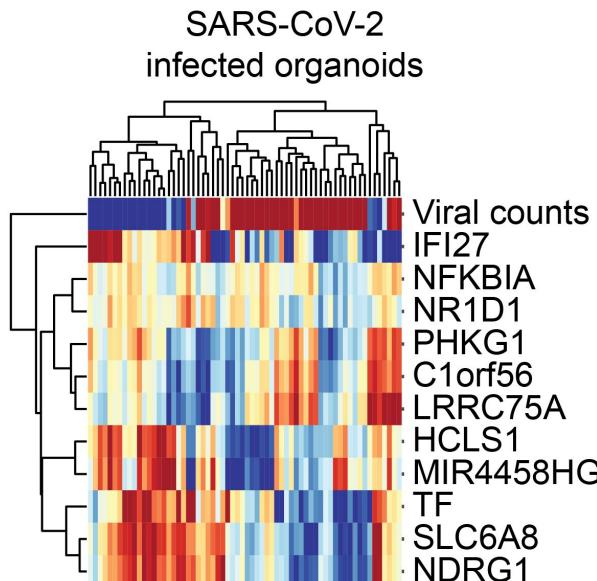
- May lead to identification of new “phenotypic” cells or unique disease state clusters

Determinants of COVID-19 severity in T cells



Trancripts learned by model for global discrimination

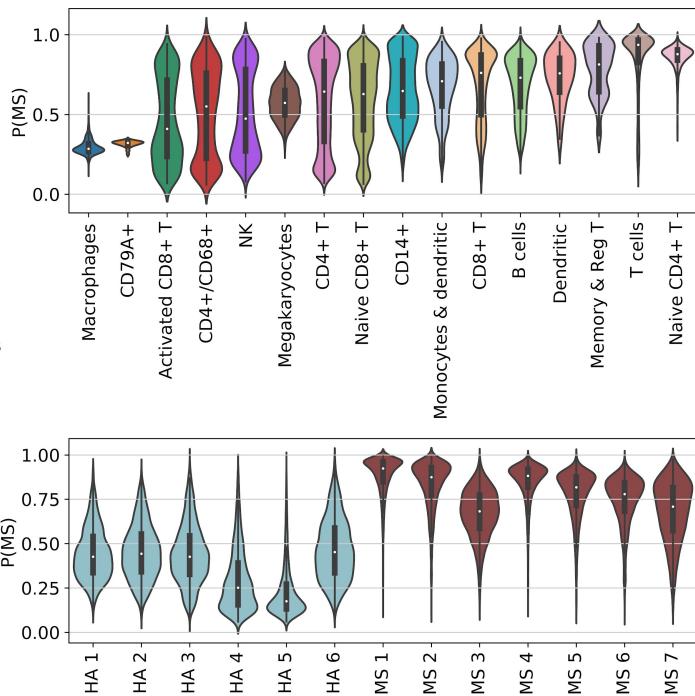
Per k attention head, $g_i^k = \max_j(|w_{ij}|)$



Contribution of cell types and cells to disease state

Predicted probabilities highlights important cell types for predicting disease state

Variance of cells' probability across patients may indicate hidden disease states



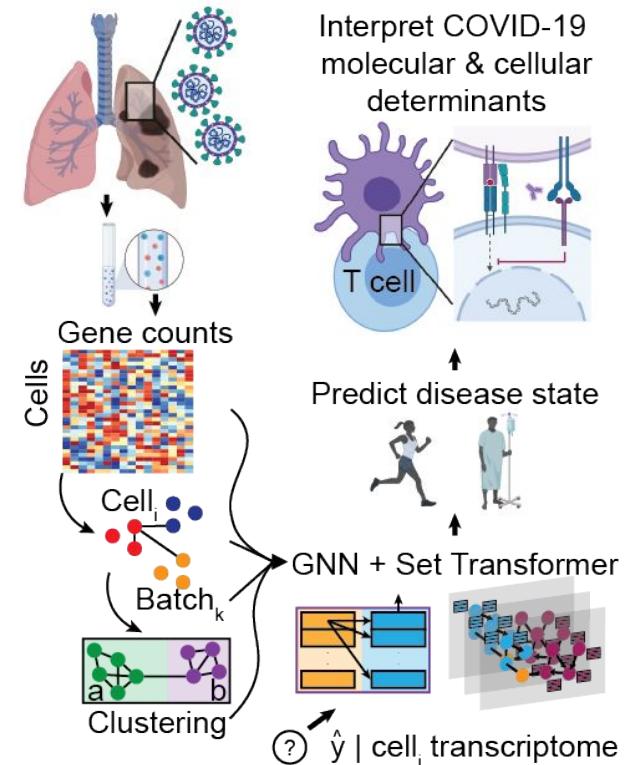
Finding genes associated with infection & severity

First longitudinal single-cell transcriptomic analysis of SARS-CoV-2 infected respiratory epithelium using an organoid model

Ciliated cells are the major target cell of SARS-CoV-2 at infection onset

SARS-CoV-2 induces IFN in infected cells and ISG in both infected and bystander cells

Innate immune response is rapid; increased expression of cell-death and aging genes may suggest cytotoxic anti-viral response



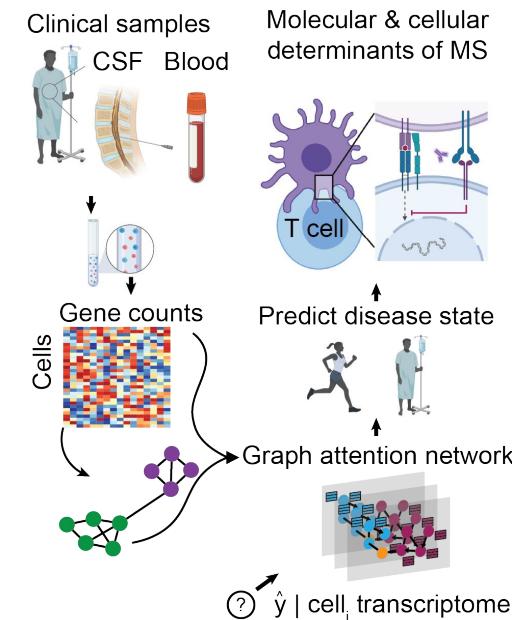
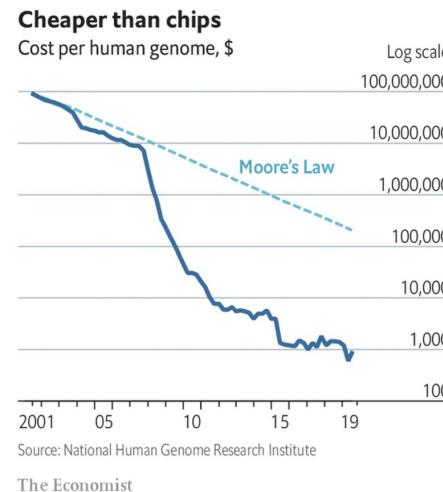
DL for insights into pathophysiology

XAI: interpretability of models to identify disease state determinants

Transfer learning (CSF v. blood) given availability of PBMC omcs data

Requiring fewer and fewer cells or features for diagnosis, prognosis, or treatment personalization

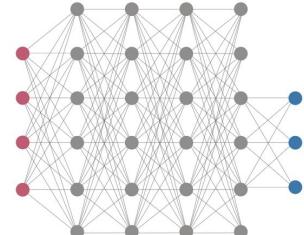
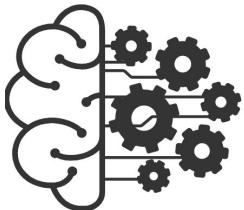
Comparison of ID'ed transcripts with DGE approaches



Acknowledgments

van Dijk Lab

Arijit Sehanobish	Shivam Saboo
Victor Gasque	Rishabh Gupta
Jason Bishai	Mingze Dong
Antonio Fonesca	Juanru Guo
Aagam Shah	David van Dijk



Collaborators

Craig B. Wilen, Yale	Stephanie C. Eisenbarth, Yale
Mia Madel Alfajore, Yale	Anna M. Pyle, Yale
Janghoo Lim, Yale	Tamas L. Horvath, Yale
Leon Tejwani, Yale	Ellen F. Foxman, Yale
Akiko Iwasaki, Yale	Richard W. Pierce, Yale
Adrian Haimovich, Yale	
Andrew Taylor, Yale	

NIH T32HL007950

Yale



Code available at
github.com/vandijklab

Add Covid Severity Index paper?

Part of American College of Emergency Physicians official COVID-19 workflow;
featured on MDCalc

Haimovich A*, Ravindra NG*, ... van Dijk D, Taylor RD. *Annal. Emer. Med.* 2020