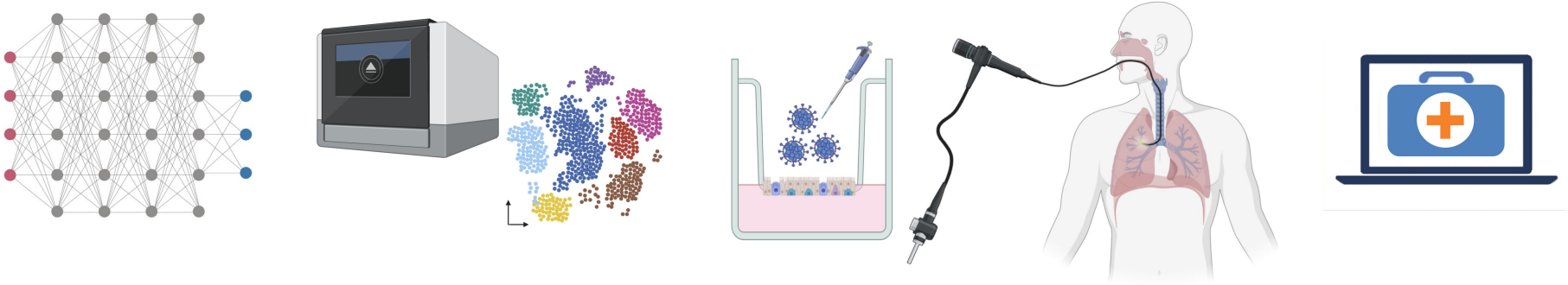
Leveraging geometric deep learning for knowledge discovery from single-cell data and explainable AI for clinical decision support tool development

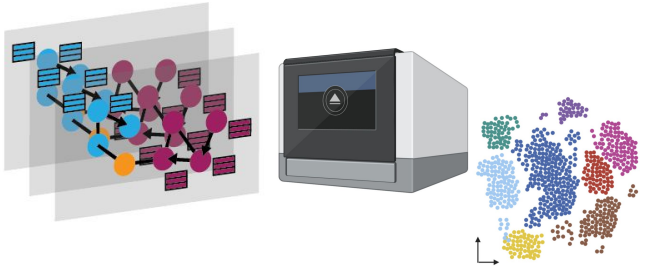


Neal G. Ravindra, Ph.D.
Machine Learning Postdoctoral Scholar at Stanford University
Dept. Of Biomedical Data Science, Anesthesiology, Pediatrics, & Stanford Artificial Intelligence Laboratory

Overview

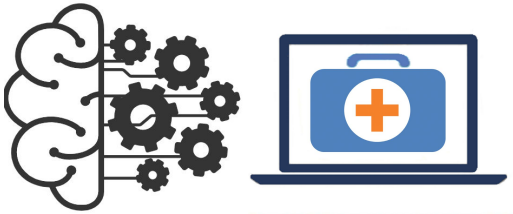
Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

- Dynamical genes from landmark time-points
- single-cell Graph Attention Networks (scGAT)



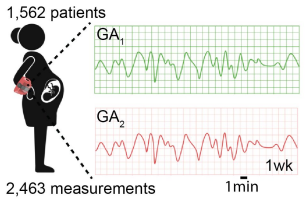
XAI to create clinically useful and parsimonious models

- qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department



Moving from applications of ML/AI to fundamental research

- actigraphy2GA: sleep and activity disruptions and their relation to preterm birth
- Permutation invariant networks to encode distributions
- sc2drug: perturbation modeling to align similar but disparate distributions

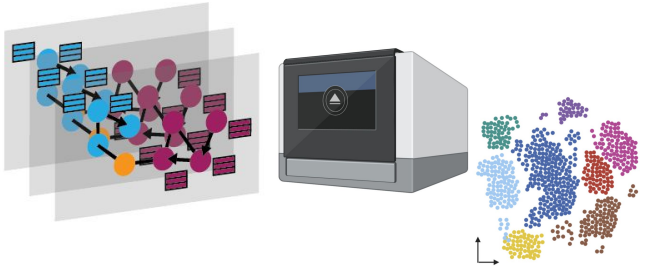


Overview

Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

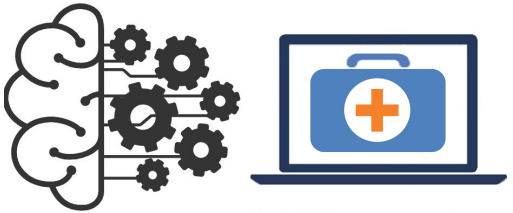
Dynamical genes from landmark time-points

single-cell Graph Attention Networks (scGAT)



XAI to create clinically useful and parsimonious models

qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department

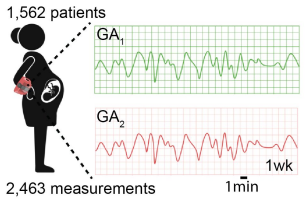


Moving from applications of ML/AI to fundamental research

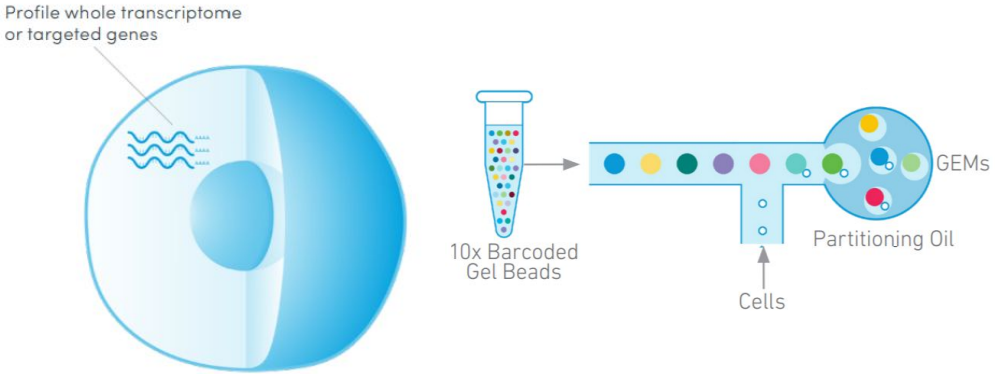
actigraphy2GA: sleep and activity disruptions and their relation to preterm birth

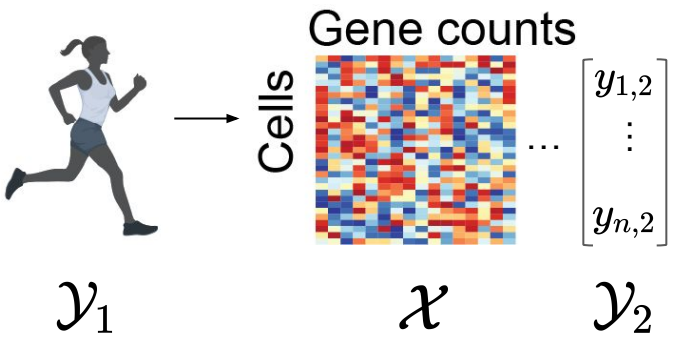
Permutation invariant networks to encode distributions

sc2drug: perturbation modeling to align similar but disparate distributions



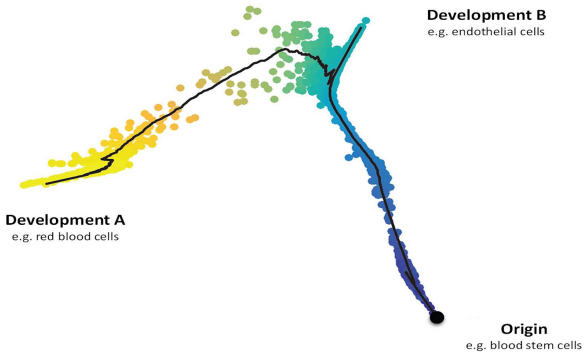
Modeling single-cell omics data

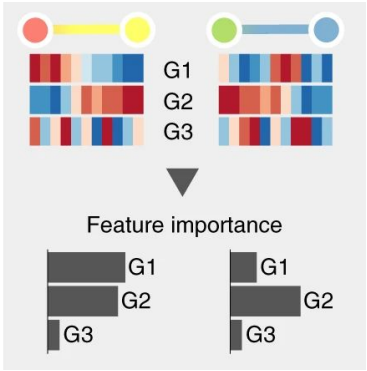




Images, https://www.10xgenomics.com/
Zheng et al. *Nat. Comm.*, 2017

Cell “pseudotime” for un-labeled transitions and ordering





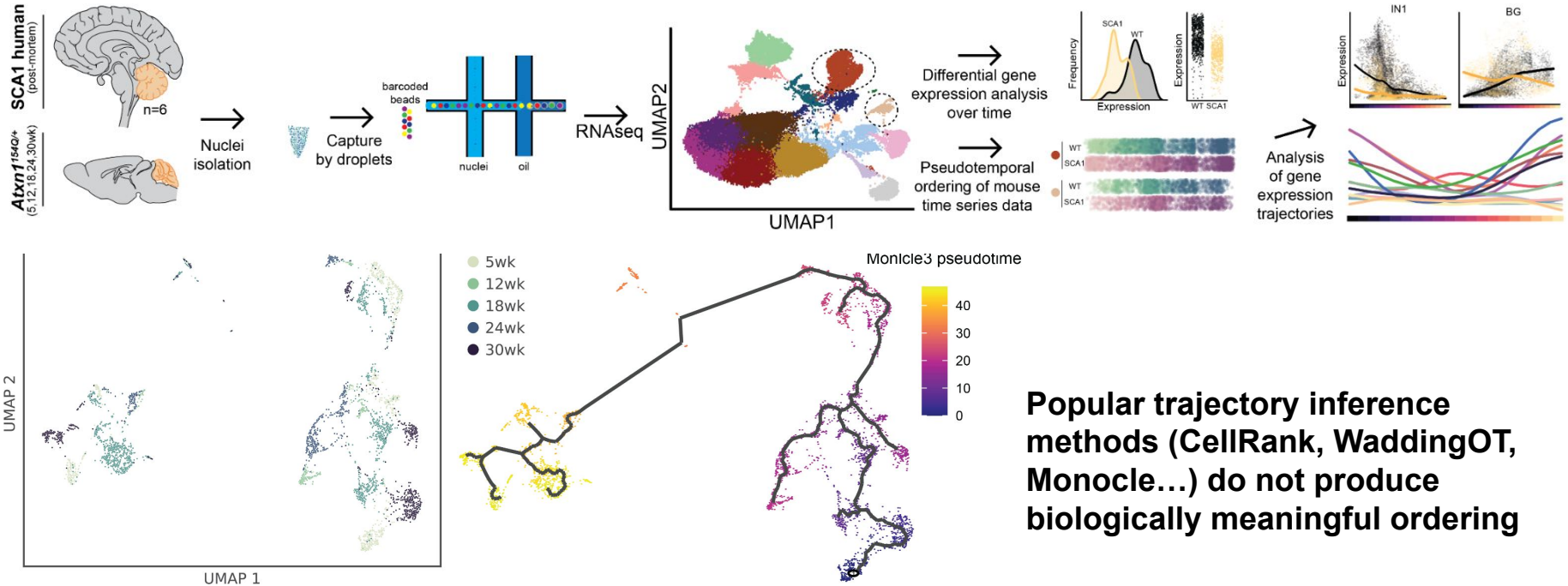
Semantically meaningful embeddings for “landmark timepoints”

Timepoints in controlled time-courses or model organisms at known discrete development stages

Manual selection of start/stop points for clusters with imperfect purity

Comparison of 45+ TI methods: Saelens et al. *Nat. Biotech.*, 2019

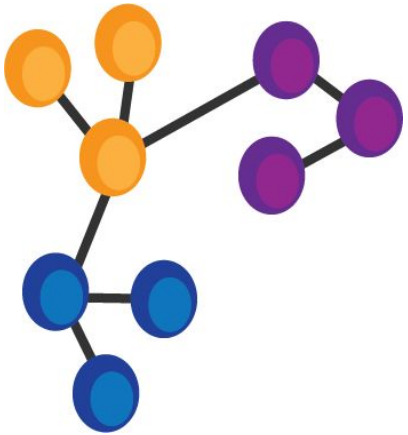
Longitudinal snRNA-seq analysis of neurodegeneration



Popular trajectory inference methods (CellRank, WaddingOT, Monocle…) do not produce biologically meaningful ordering

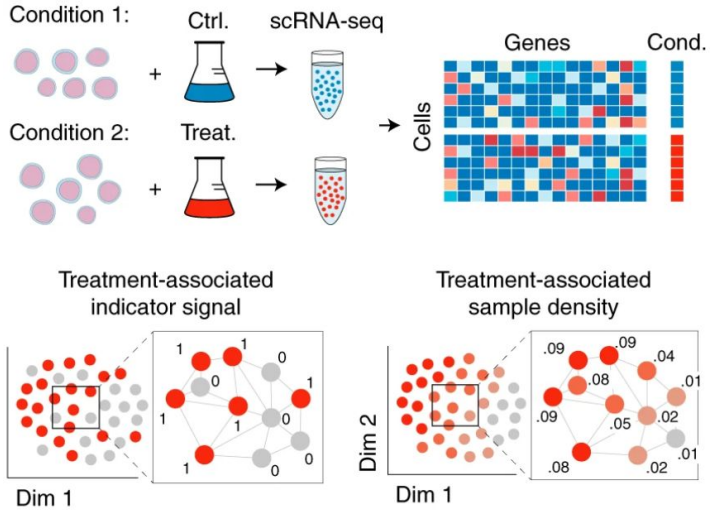
Tejwani L*, **Ravindra NG***, … van Dijk D, Lim J. *in revision* at *Cell*
Tejwani L*, **Ravindra NG***, … van Dijk D, Lim J. *in submission* at *Nature Neuroscience*

Label smoothing based on cell-cell similarity



Cell graph

cells are *nodes*,
edges have weights ~ distance in embedding

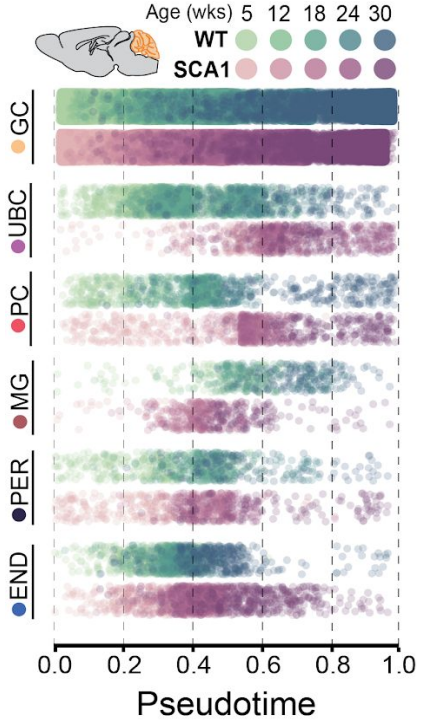


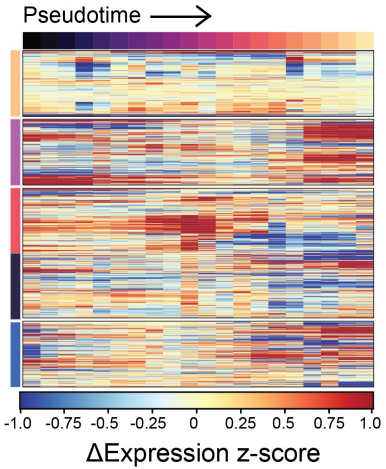
MELD: Burkhardt et al. *Nat. Biotech.*, 2021

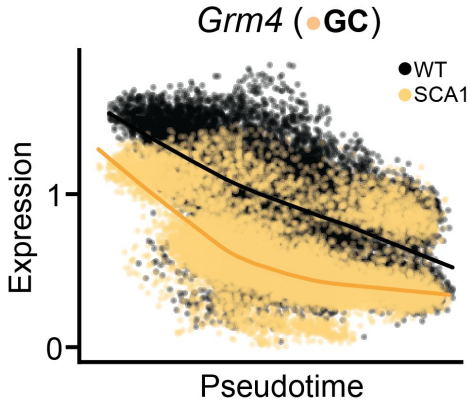
Borrow from graph signal processing to diffuse known landmark timepoint according to cell-cell similarity, which may have impure modularity w.r.t. time

Apply $\hat{f}(x,\gamma) = e^{-\gamma\mathcal{L}}$ to $(\mathbf{G}, \mathbf{Y})$

Label smoothing according to mouse developmental age

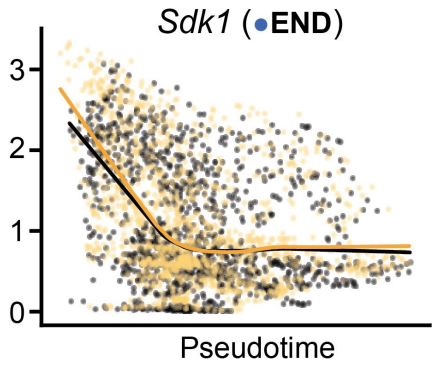


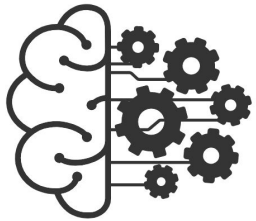




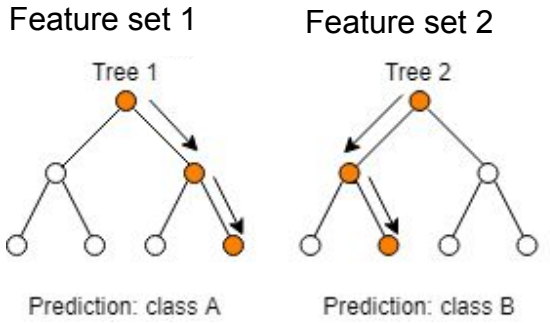
Pseudolabeling reduces confounding of discrete time-point groupings while respecting landmark order

Regression model interpretability per pseudotime branch





Gradient boosting + MAGIC



Importance ~ *gain* in accuracy when a feature is added to branch

Discovering "dynamical" genes

- Need an interpretable, non-linear model
- Fast because many features and sub-analyses by cell type

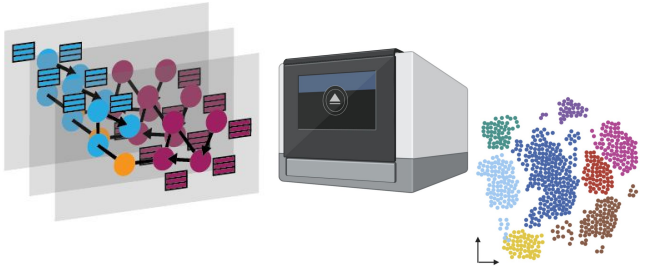
Images from Thomas Wood, *DeepAI*

Overview

Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

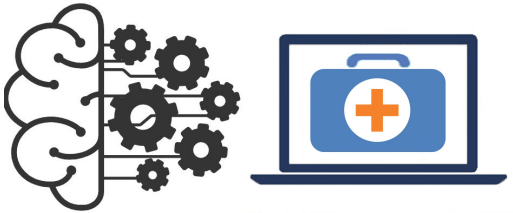
Dynamical genes from landmark time-points

single-cell Graph Attention Networks (scGAT)



XAI to create clinically useful and parsimonious models

qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department

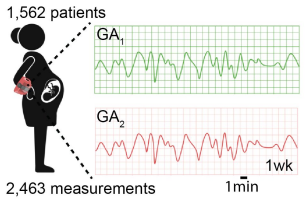


Moving from applications of ML/AI to fundamental research

actigraphy2GA: sleep and activity disruptions and their relation to preterm birth

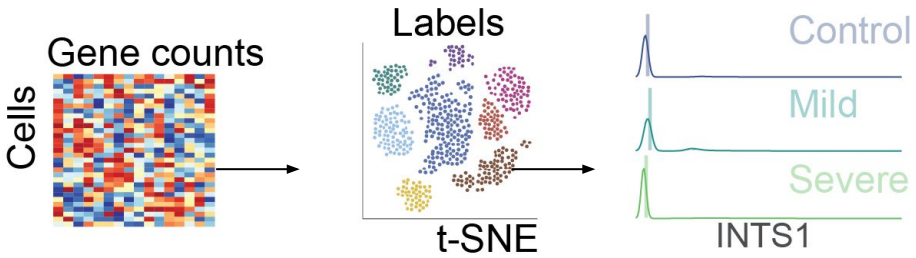
Permutation invariant networks to encode distributions

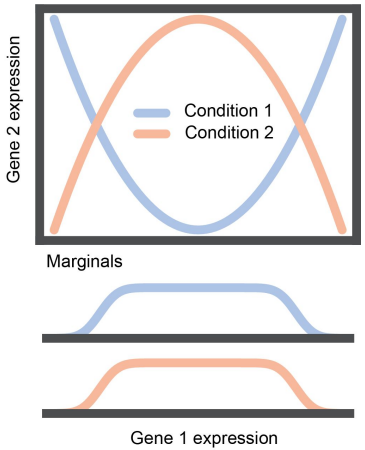
sc2drug: perturbation modeling to align similar but disparate distributions



Discovering molecular mechanisms from single-cell data

scRNA-seq pipelines to study genetics and disease



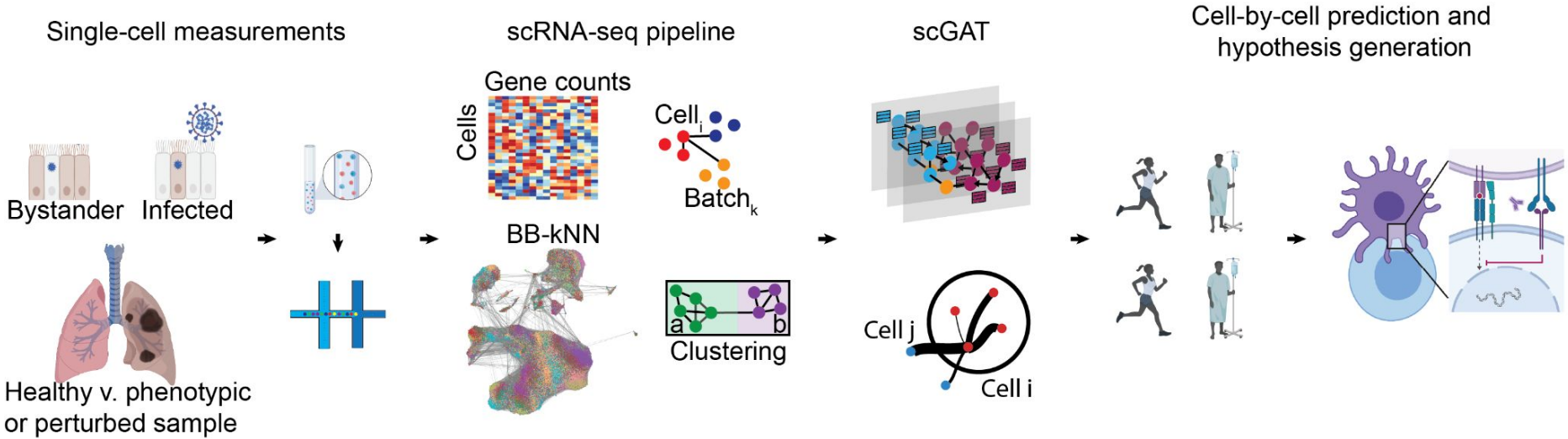


DGE is not necessarily *associated* with disease or cell state and “most” differentially expressed genes do not yield causal structure

Most DGE methods don’t allow for interactions between features

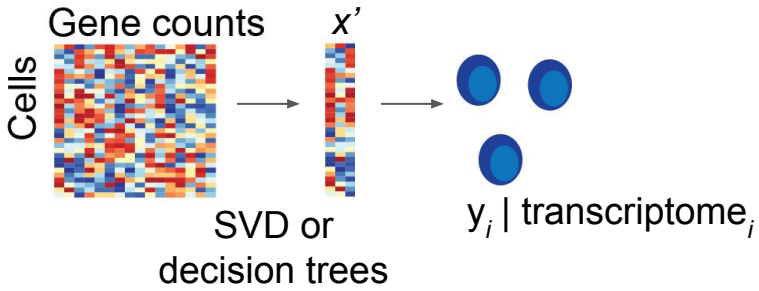
Pipeline error and missing trivial non-linear, causal covariance structures

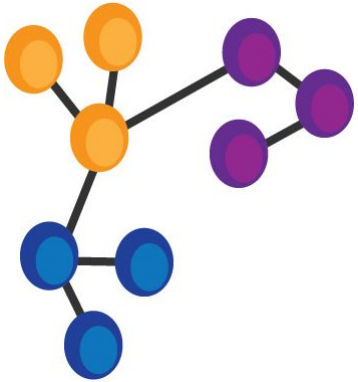
Geometric deep learning to represent single-cell data

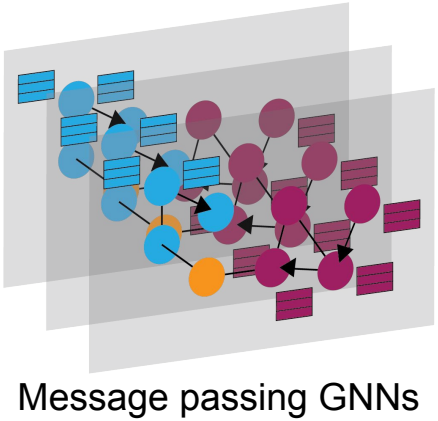


Ravindra NG*, Sehanobish A*, Pappalardo J, Hafler D, van Dijk D. *ACM CHIL*, 2020
Sehanobish A*, **Ravindra NG***, van Dijk D. *ICML'20 GRL+*
Sehanobish A*, **Ravindra NG***, van Dijk D. *AAAI'21*
Ravindra NG, Sehanobish A, Alfajaro MM, Wang B, Foxman EF, Wilen CB, van Dijk. (*in submission* at *Nature Methods*)
Ravindra NG, Sehanobish A, van Dijk D. (*in submission at ICML'22 workshop*)

Geometric deep learning and attention mechanisms

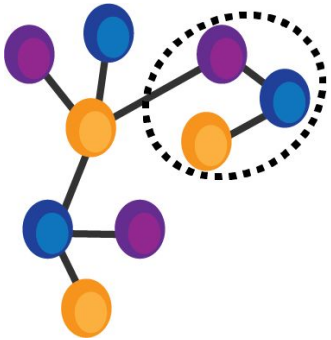






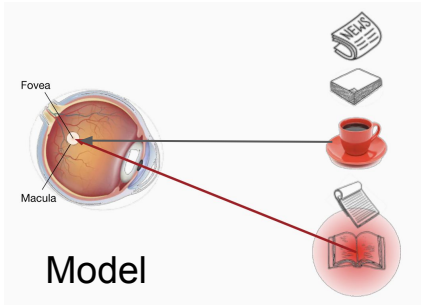
Single-cell data is heterogeneous, sparse, noisy, and # features ~ # cells so sub-sampling of features by classical ML is random--need inductive bias of self-*similarity*

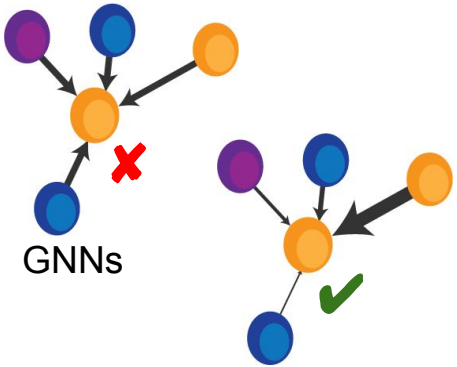
Geometric deep learning and attention mechanisms



Realistic cell graph

Attention





GNNs

GAT or Graph attention networks
(axial attention)

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

GAT

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

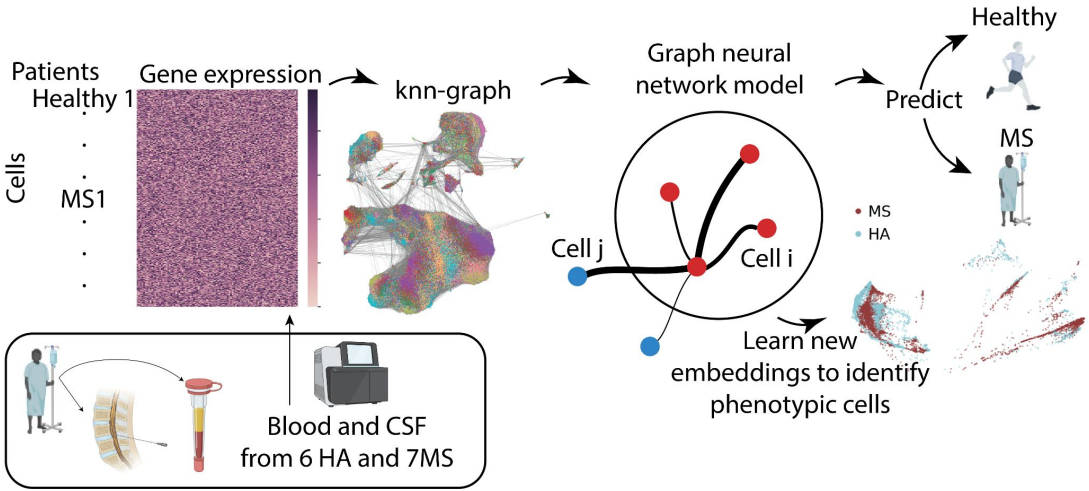
$$h' = \{h'_1, h'_2, \ldots, h'_N\}$$

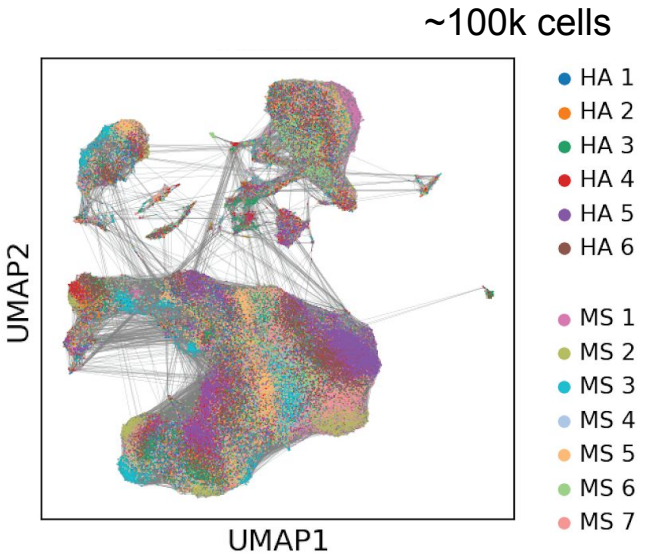
$$\alpha_{ij} = \mathrm{softmax}_j(a(\mathbb{W}h_i, \mathbb{W}h_j))$$

Most GNNs equally weigh messages across edges, which is problematic with *bad* input cell graphs (e.g., high degree of heterophily)

Vaswani et al. *NIPS'17*; Lin et al. *ICLR'17*; Velickovic et al. *ICLR'18*

GAT supervised learning to study disease mechanisms





Multiple sclerosis (MS) is a disease of the nervous system, can exhibit relapsing-remitting neuroinflammation

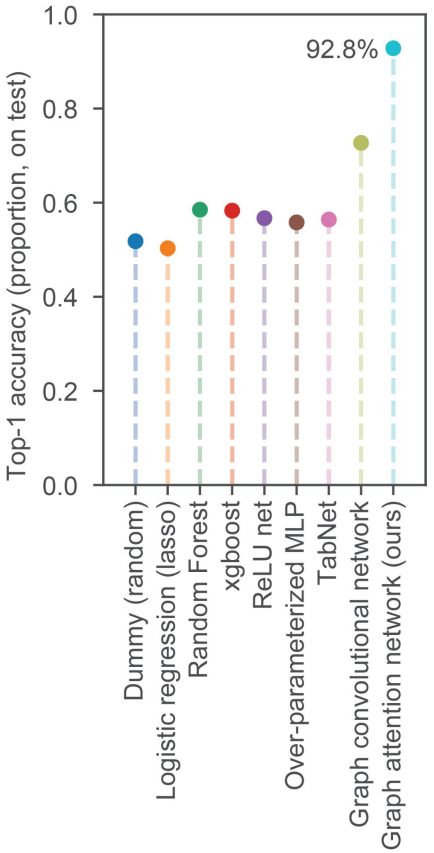
What factors are up-regulated in MS patients and what cell types/subsets are responsible

GATs learn to predict disease state from a transcriptome

Other common supervised learning approaches fail to learn how to predict each cell's disease state

GATs perform better relative to other popular GNN models

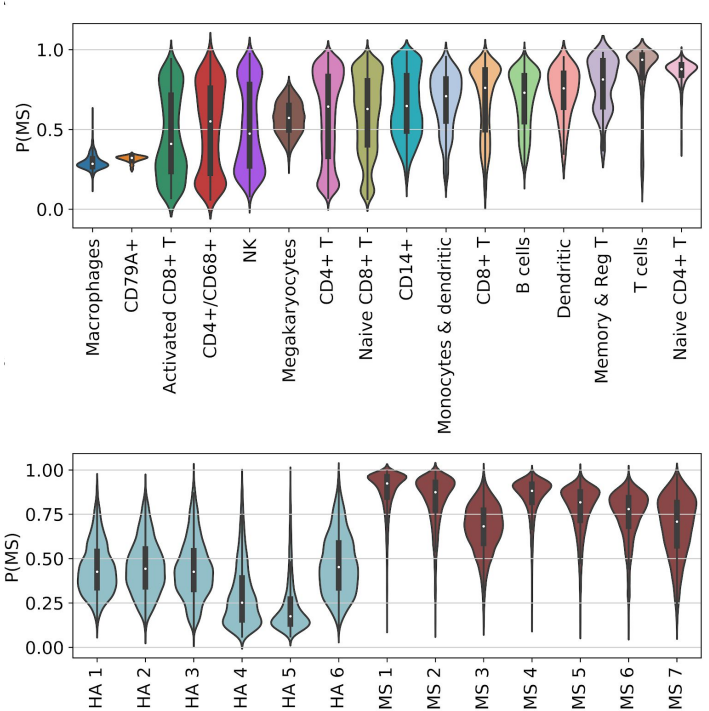
Task	Model	Accuracy
Inductive	Random	51.8
	MLP	56.7
	Random Forest	58.5
	Graph Convolutional Network	72.1
	Graph Attention Network(our)	**92.3** $\pm$ .7
Transductive	Graph Convolutional Network	82.91
	Graph Attention Network(our)	**86** $\pm$ .3



Simultaneous molecular and cellular interpretability

Aggregating predicted probabilities shows cell types important for predicting disease state

Variance of a patient's cells' probability of being in an MS state may indicate timing of flare-up



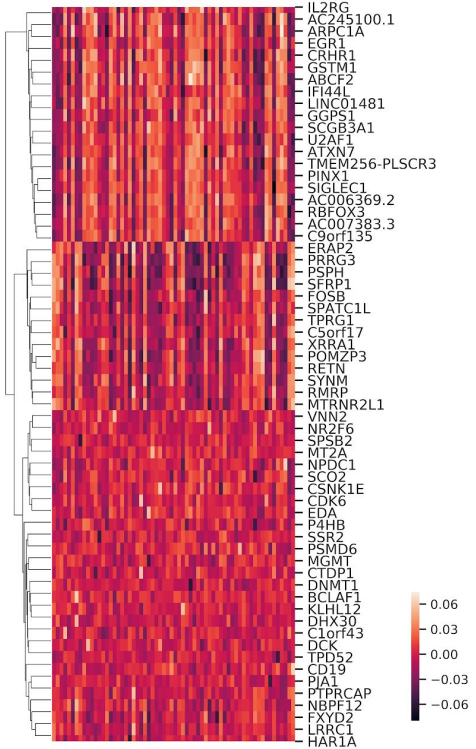
Finding genes important for predicting disease state

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

Interleukin-2 receptor subunit among top 10 predictive features per head

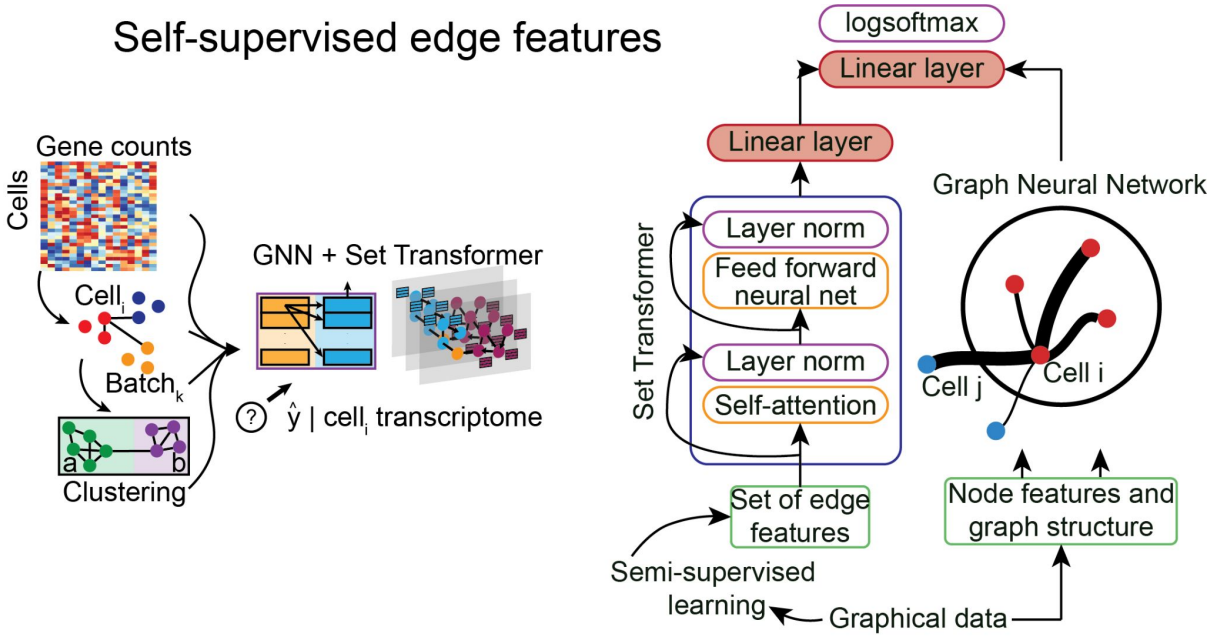
Marker for therapeutically targeted B cells (CD19) also among top features

Top predictive features regulate hormone secretion, nerve cell development, and lipid metabolism, suggesting relevant but novel hits

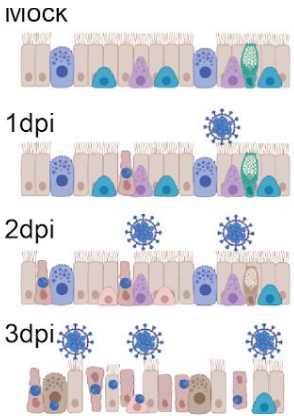


Integrating self-supervised learning and GATs

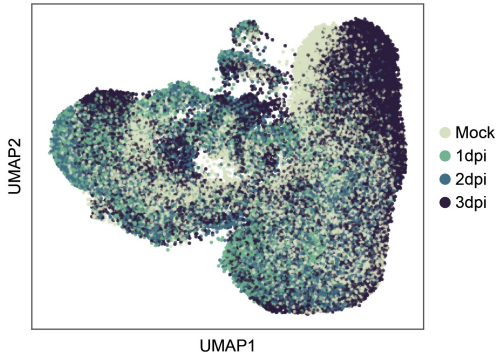
Goal to gain insight into SARS-CoV-2 infection and COVID-19 severity with additional controls for patient and sample source variability for robustness

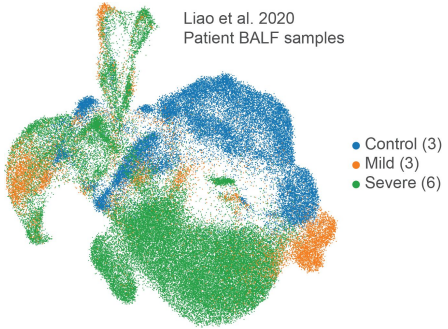


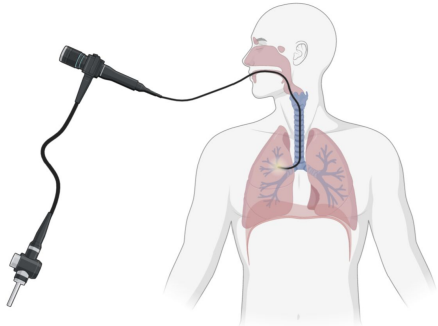
scGAT for HBEC & heterogeneous BALF samples



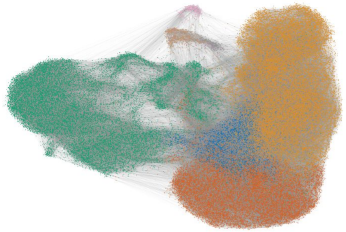
HBEC



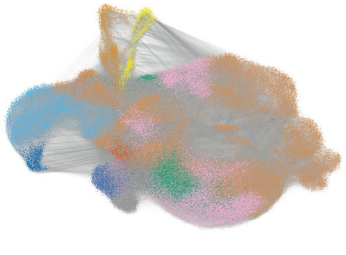




SARS-CoV-2 infected Bronchial Epithelium Organoids



COVID-19 Patient Samples



~100k cell datasets: HBEC to study molecular and cellular determinants of early infection, BALF samples to study aberrant cell types and genes responsible for severe/critical COVID-19

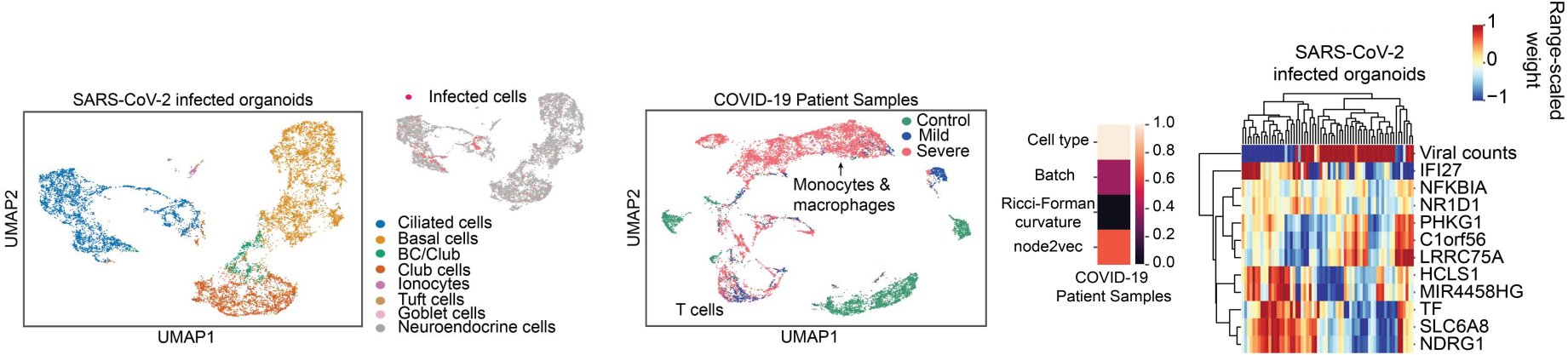
Our novel ingestion of *de novo* edge features using SSL

Models	SARS-CoV-2 infected organoids	COVID-19 patients
ClusterGCN	65.43 (65.21-65.65)	89.26 (89.06-89.47)
ClusterGCN + DeepSet	79.75 (78.75-80.75)	87.2 (87.02-87.38)
ClusterGCN + Set2Set	71.65 (69.89-73.42)	88.34 (87.89-88.79)
ClusterGCN + 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)

Our model significantly improves predictive performance over popular GNN methods, and controls for sample source and heterogeneity, giving high-confidence that it can be interpreted

Accuracy and 95% CIs for n=6 trials

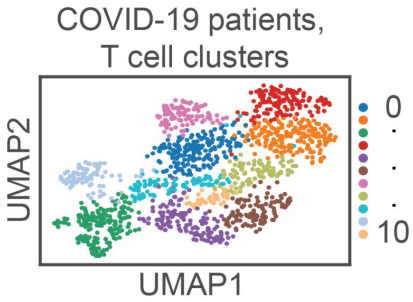
Subsetting cells important to COVID-19 severity using learned graphical representations

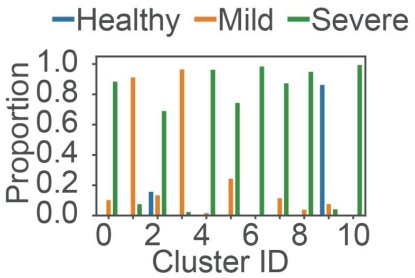


Visualizing attention using traditional unsupervised learning approaches shows model simultaneously discriminates by cell type and label

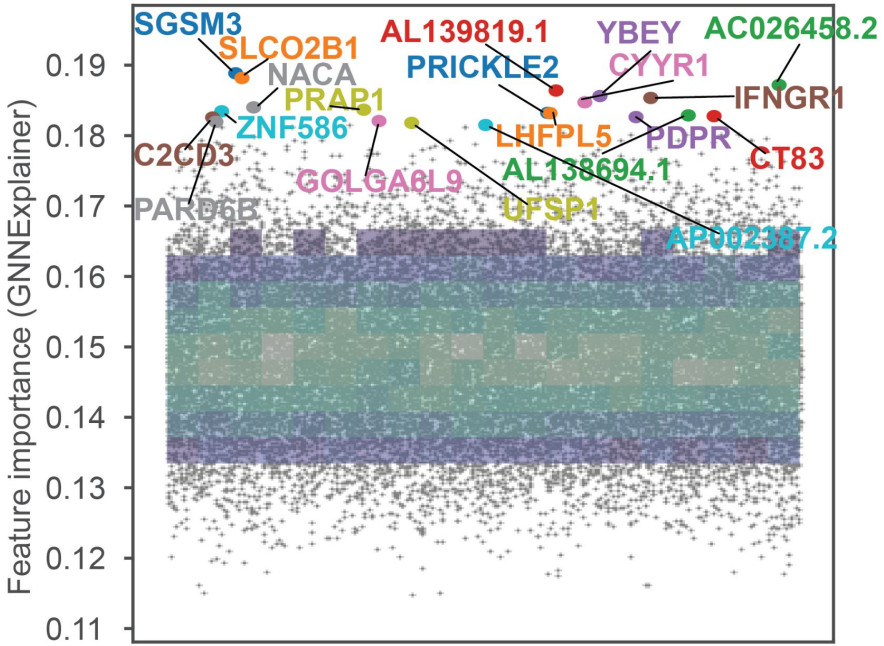
Model relies on genes involved in the innate immune system to discriminate between HBECs and cell types in heterogeneous clinical samples

Genes driving an active cluster of T cells in severe pts





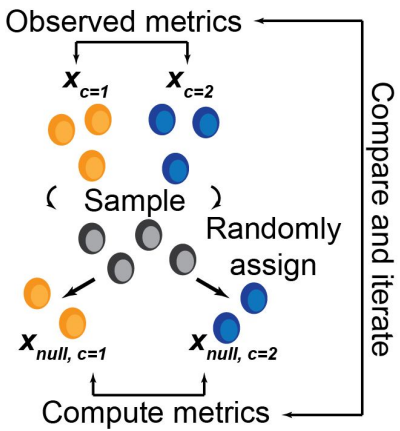
Top 20 important genes for predicting severe COVID-19 in cluster 10

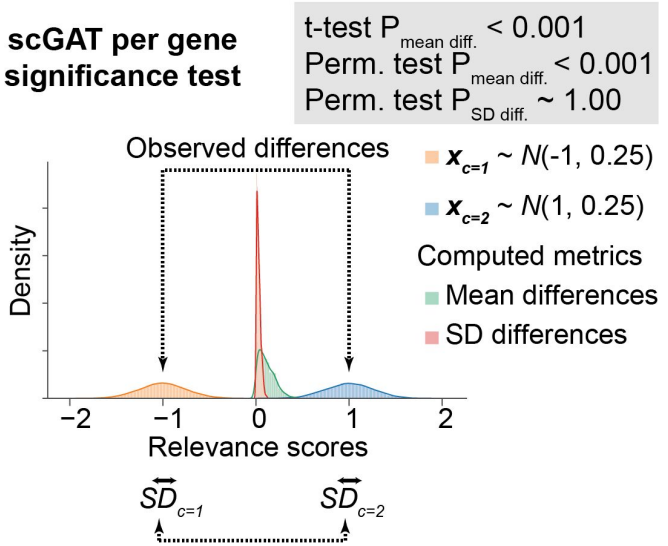


Combining interpretability via attention and feature attribution methods allows us to simultaneously gain insight into cells and genes driving model's association with label

Defining *significant* features associated with cell state

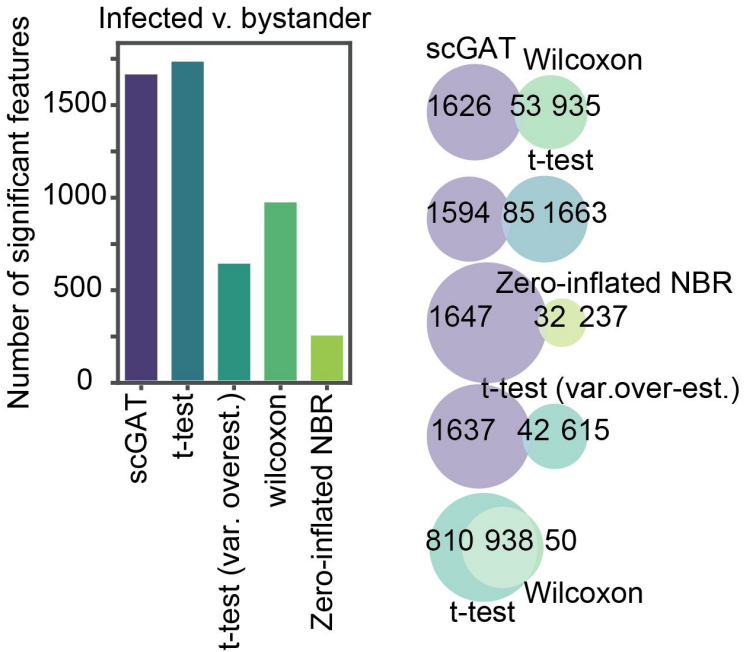
SmoothGrad scores

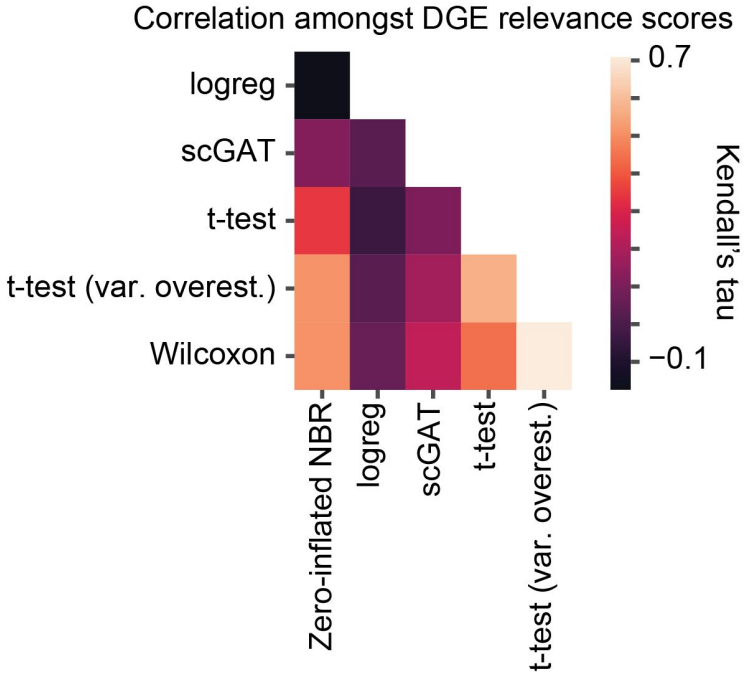




Goal: going beyond top-k features

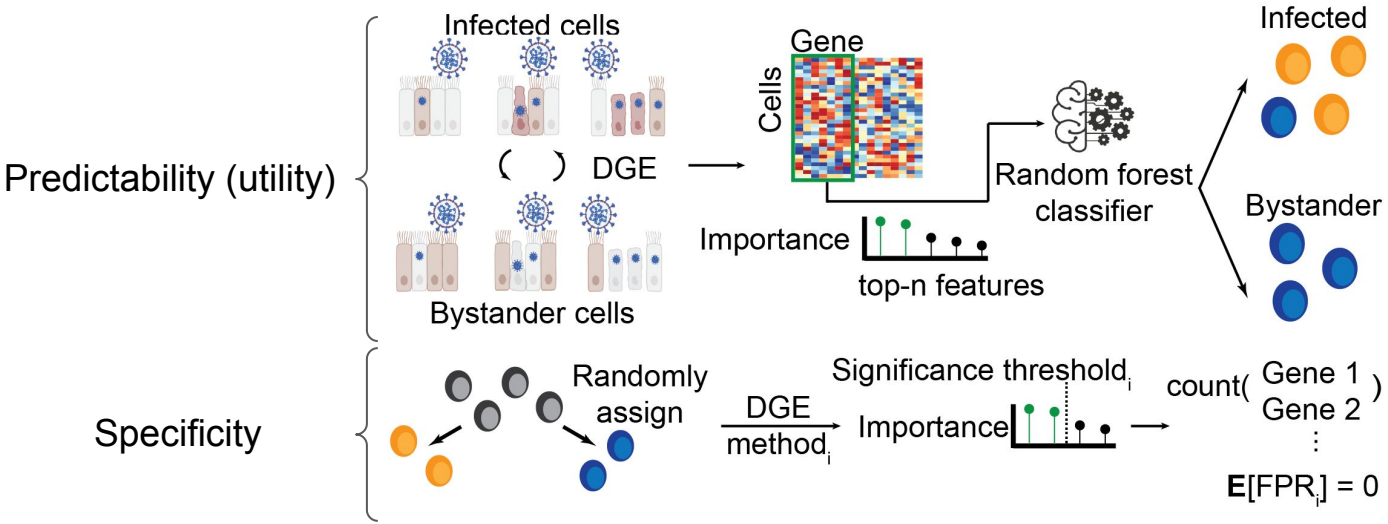
scGAT identifies unique infected v. bystander differences



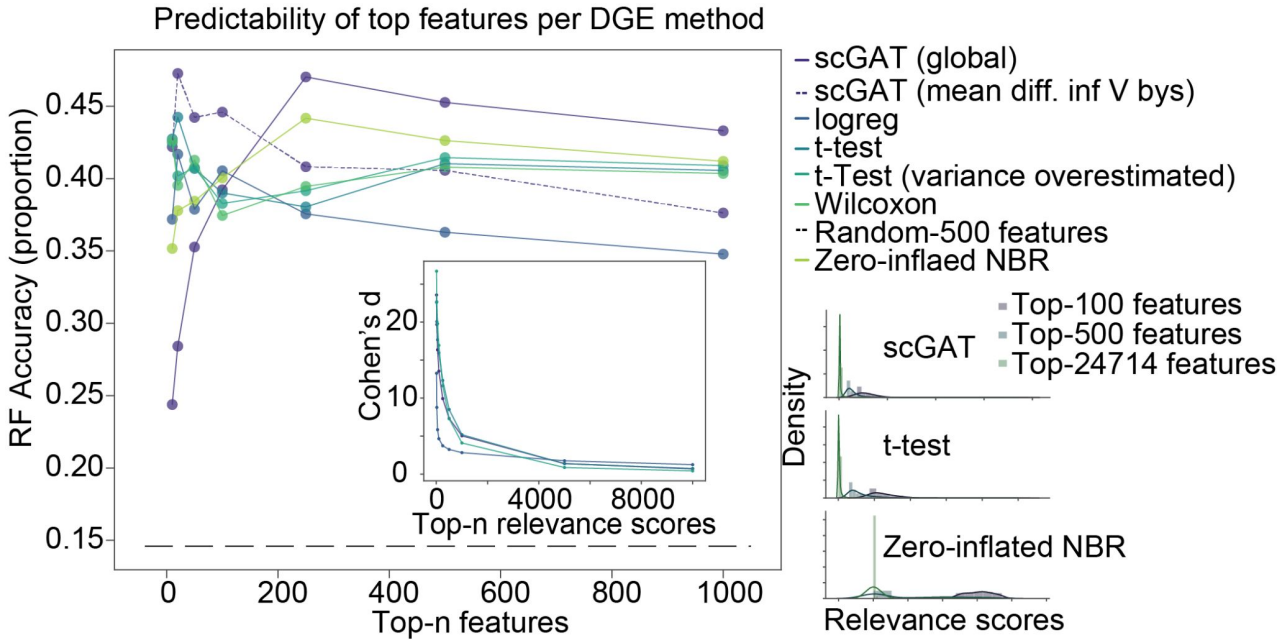


Lack of overlap with scGAT's *significantly* important features may suggest many genes are missed by standard DGE methods

Comparing scGAT features to other DGE methods

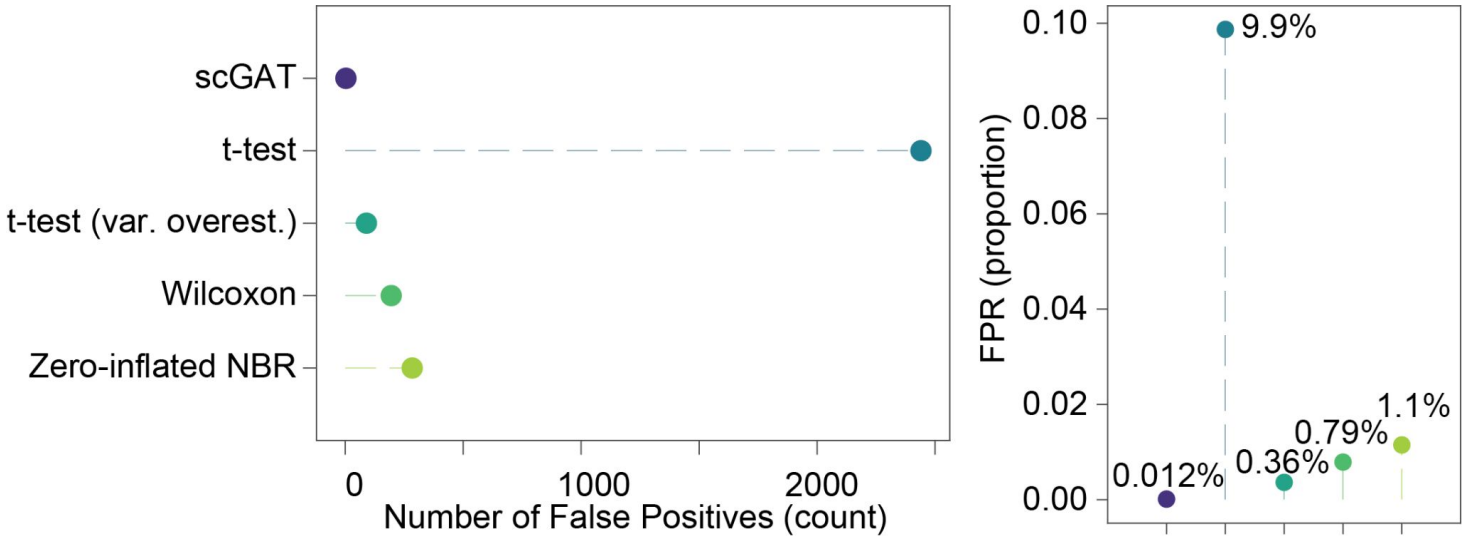


Predictability of top important features, scGAT v. DGE



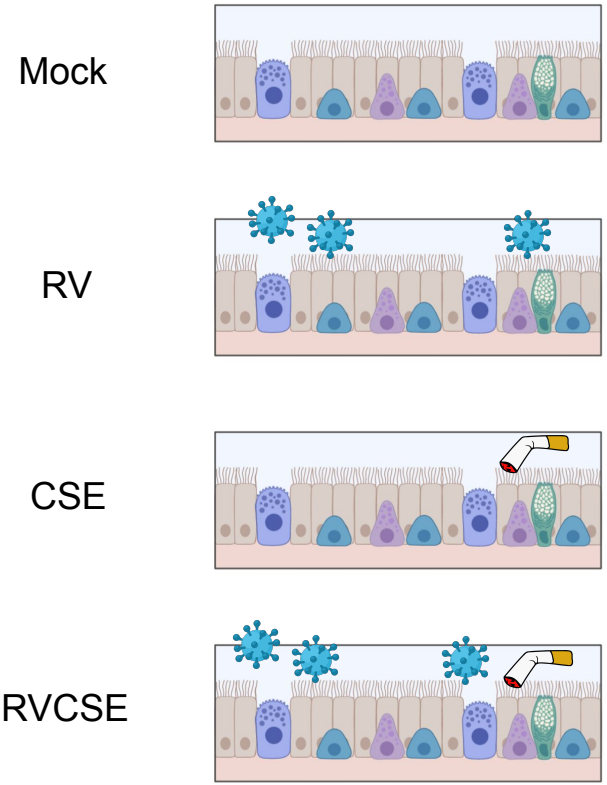
scGAT features are more predictable than features deemed "important" by standard single-cell DGE methods

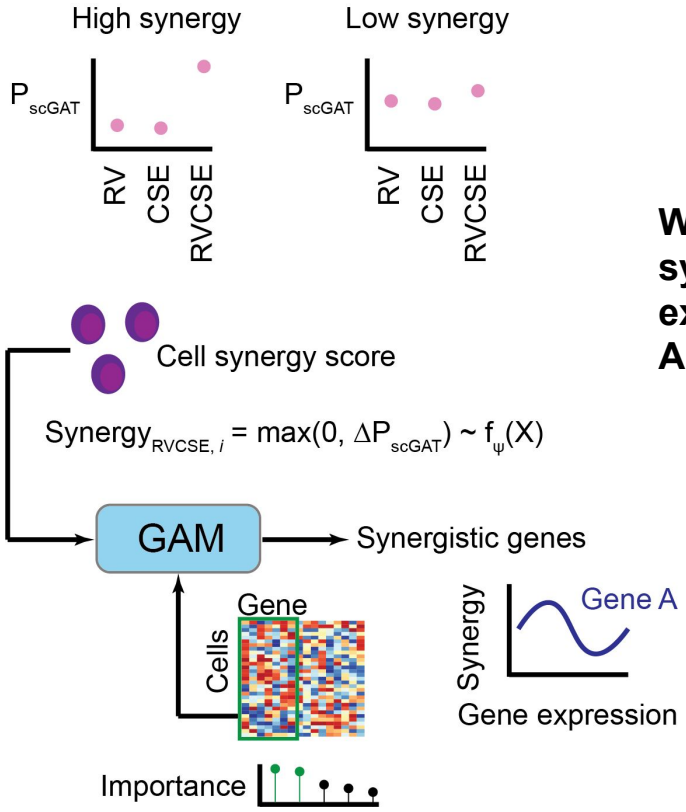
scGAT feature importance yields the lowest FPR



scGAT has a virtually 0 false-positive rate, which can be quite high for DGE methods, especially large datasets (post hoc inference tests use ~10k cells)

Identifying synergistic effects and associated genes





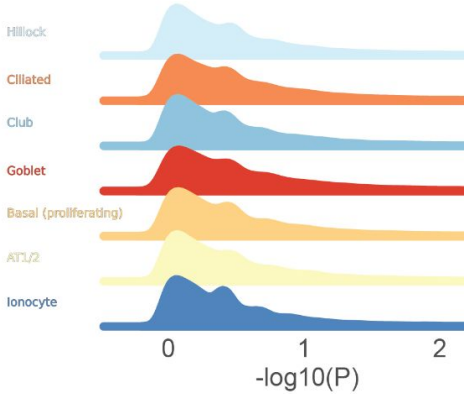
What genes/cells have synergism in combined exposure,
A v. B. v. AB?

Bao Wang and Ellen Foxman

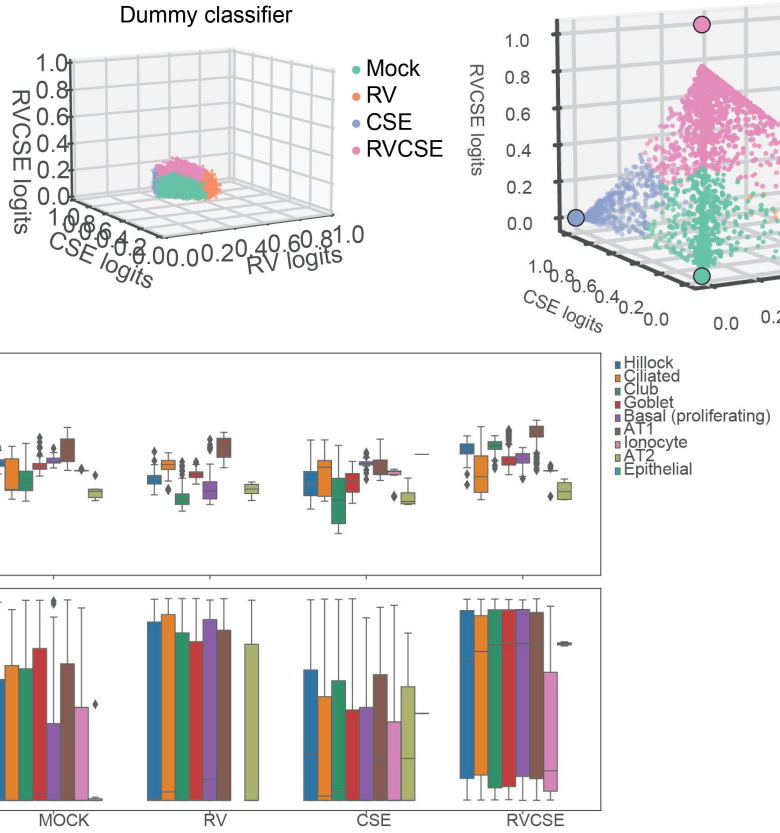
Using scGAT logits to define synergy scores per cell

Cell synergy score

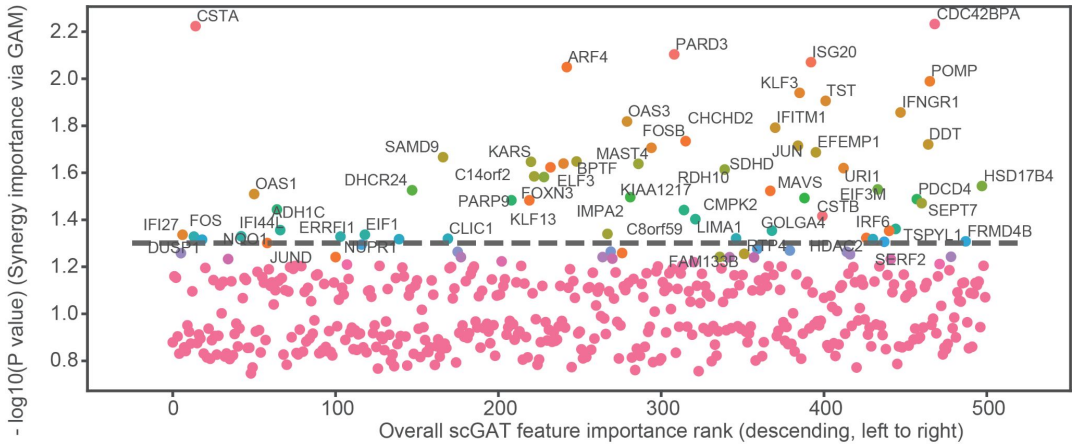
$$Synergy_{RVCSE,\,i} = \max(0, \Delta P_{scGAT}) \sim f_{\psi}(X)$$

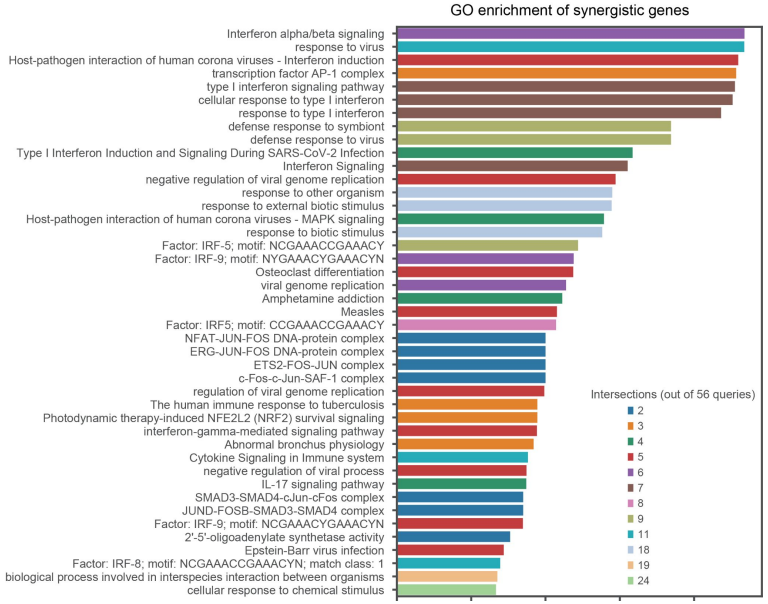


Flexible, non-linear, *learned* scoring



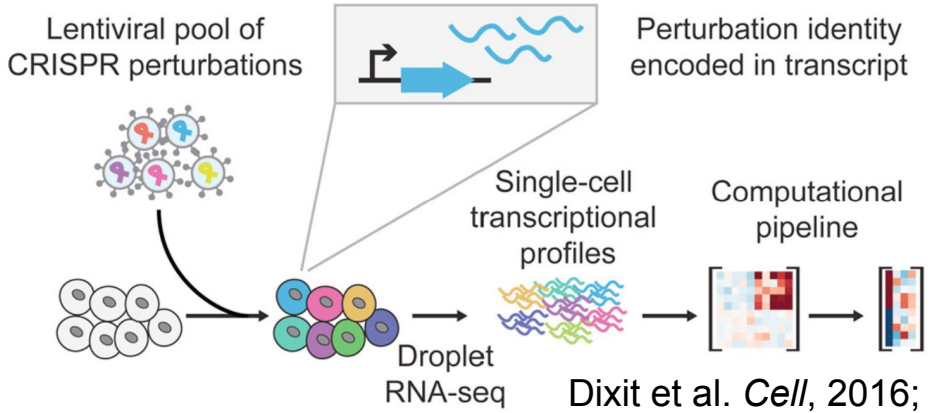
Synergistic genes: attributing genes to cell synergy score





scGAT can be used to identify semantically meaningful features associated with various complex and overlapping experimental designs or environmental contexts

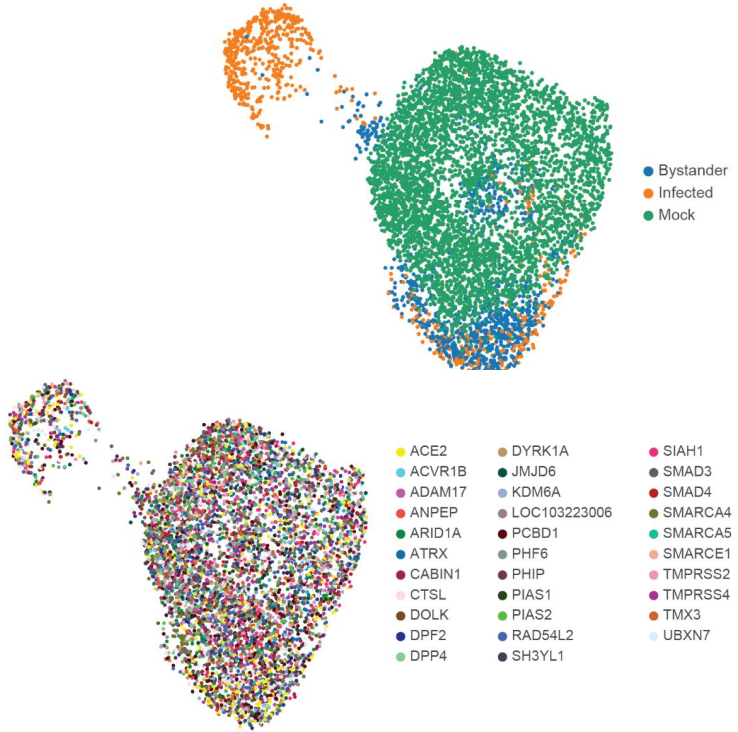
Single-cell Perturb-seq data in multiple contexts



Dixit et al. *Cell*, 2016;
Adamson et al. *Cell*, 2016

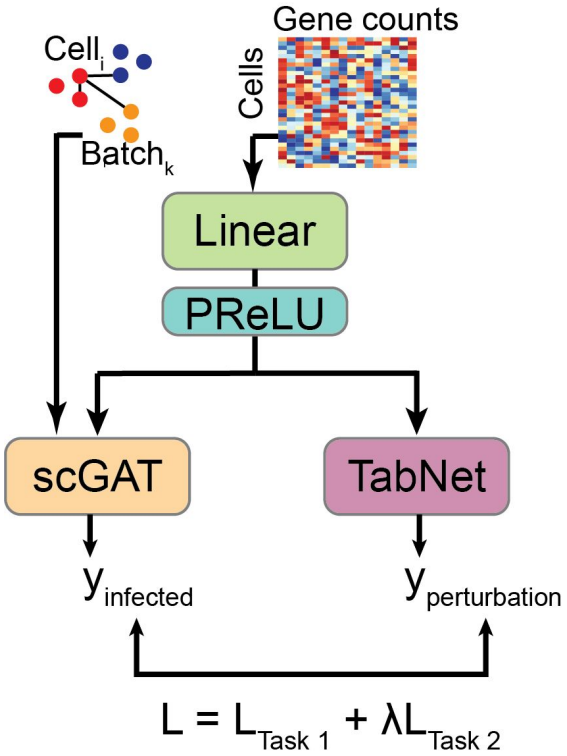
Want to find which genes *and* perturbations are associated with infection susceptibility or response to SARS-CoV-2

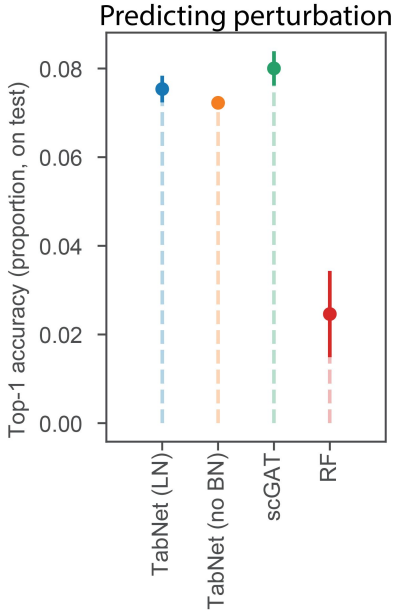
Want to know which genes respond to CRISR/Cas-9 mediated perturbation but only have cell-cell attention

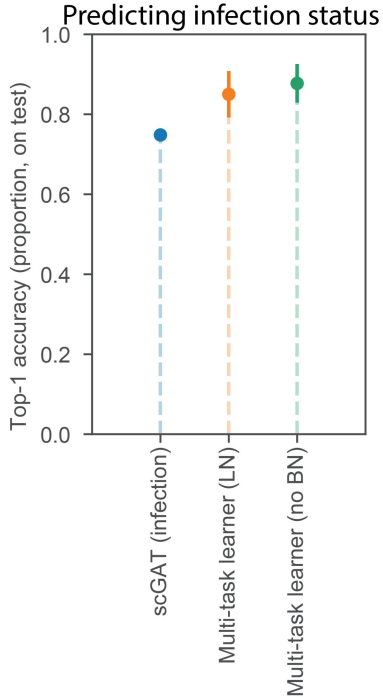


Vero E6 cells (monkey kidney), <10k cells, **MMA & CBW**

Multi-task learning to combine axial and feature attention

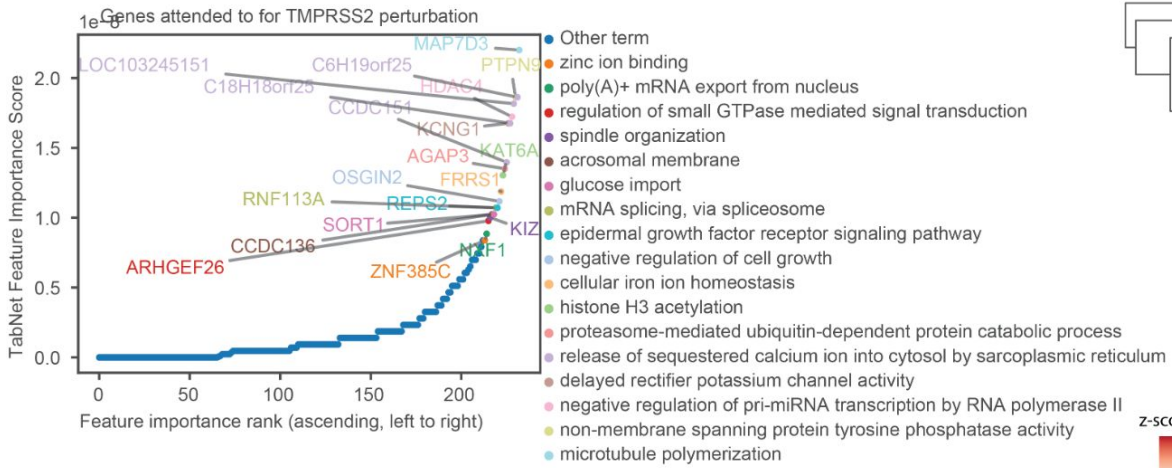


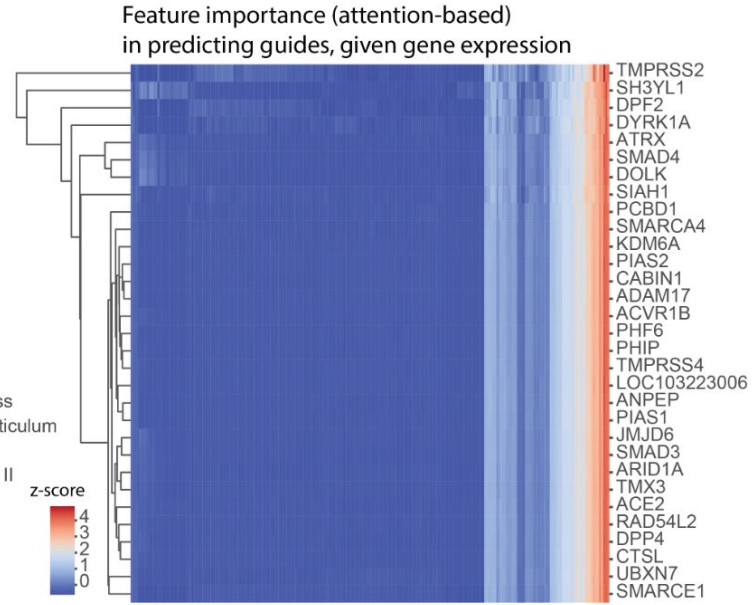




Results | scGAT

Comparing feature-wise attention for infected & bystander



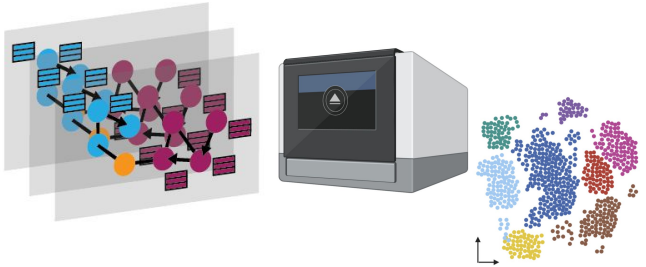


Multi-task, attention-based learning can provide insight into how perturbations differ between conditions, separating what responds to perturbation alone from what is responsive *and* differs between conditions

Overview

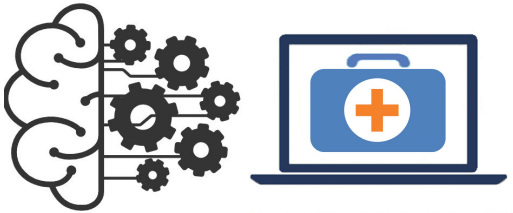
Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

- Dynamical genes from landmark time-points
- single-cell Graph Attention Networks (scGAT)



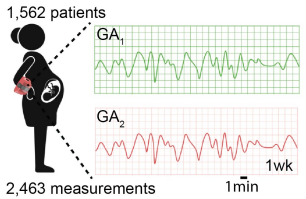
XAI to create clinically useful and parsimonious models

- **qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department**

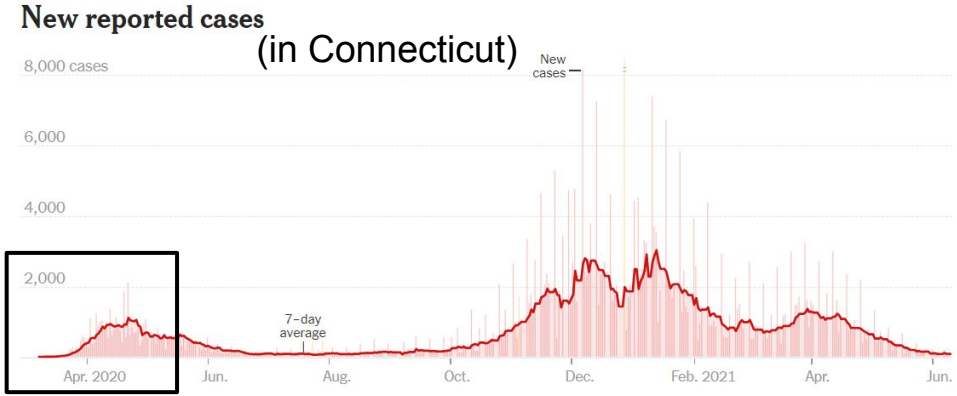


Moving from applications of ML/AI to fundamental research

- actigraphy2GA: sleep and activity disruptions and their relation to preterm birth
- Permutation invariant networks to encode distributions
- sc2drug: perturbation modeling to align similar but disparate distributions



Prognostic challenge for the fast and perplexing onset of respiratory deterioration in SARS-CoV-2+ patients



Inappropriate inpatient dispositions lead to increased provider contacts and is associated with higher morbidity

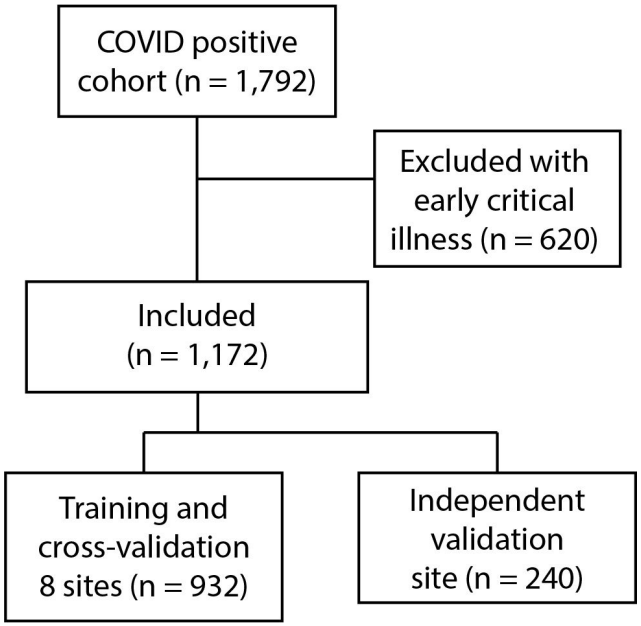
Sequential [Sepsis-related] Organ Failure Assessment (qSOFA), Elixhauser, and CURB-65 inadequate

Relatively few patients (1,172 COVID-19+ across 9 EDs), high-dimensional and messy EHR data

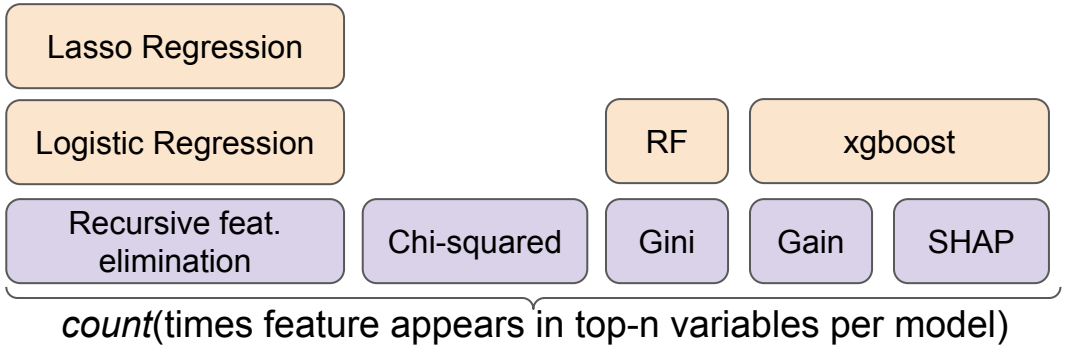
Adrian Haimovich, Andrew Taylor

Visual from *NYTimes*

Cohort selection, data preprocessing, and ML



Outcomes w/in 24h based on first 4h of data



Ensemble approach: aggregating over multiple feature selection methods to overcome individual strengths and weaknesses

(q)CSI performance v. other clinical decision support tools

Model	AU-ROC	Accuracy	Sensitivity	Specificity	AU-PRC	Brier score	F1	Average Precision
CURB-65	0.66 (0.58,0.78)	0.79 (0.56,0.94)	0.67 (0.29,1.00)	0.62 (0.27,0.93)	0.26 (0.09,0.44)	0.10 (0.06,0.15)	0.20 (0.00,0.36)	0.20 (0.10,0.33)
qSOFA	0.76 (0.69,0.86)	**0.88 (0.82,0.95)**	0.79 (0.62,1.00)	0.70 (0.60,0.80)	0.35 (0.09,0.62)	0.09 (0.05,0.14)	0.21 (0.00,0.46)	0.26 (0.13,0.42)
Elixhauser	0.70 (0.62,0.80)	0.71 (0.40,0.86)	0.73 (0.47,1.00)	0.67 (0.33, 0.88)	0.20 (0.09, 0.36)	0.10 (0.06, 0.15)	0.30 (0.15,0.43)	0.22 (0.11, 0.36)
qCSI	0.90 (0.85,0.96)	0.84 (0.72,0.94)	0.90 (0.70,1.00)	0.79 (0.59,0.94)	0.54 (0.27,0.76)	**0.07 (0.04,0.11)**	0.49 (0.30,0.67)	0.52 (0.30,0.72)
CSI	**0.91 (0.86,0.97)**	0.83 (0.70,0.94)	**0.94 (0.77,1.00)**	**0.82 (0.67,0.95)**	**0.56 (0.25,0.80)**	0.25 (0.25,0.28)	**0.51 (0.29,0.70)**	**0.58 (0.31,0.81)**

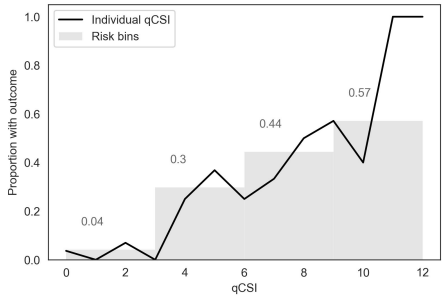
Some overfitting but parsimony helps to reduce generalization error and qCSI outperforms other popular tools, without high FPR, which would reduce system burden

Leveraging interpretable ML to create a COVID-19 severity index for ED disposition decisions, w/easy entry

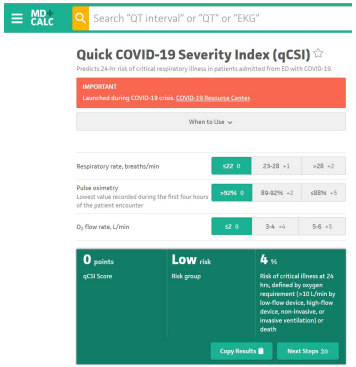
24h respiratory decompensation in admitted pts

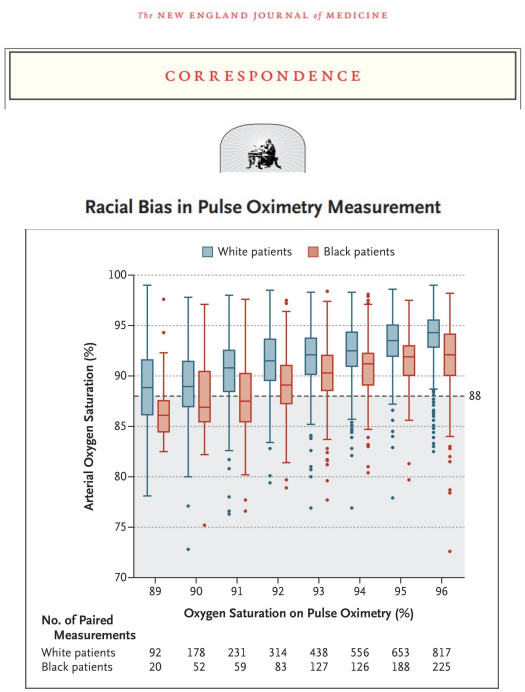
American College of Emergency Physicians official COVID-19 triage workflow

qCSI variable	Points	Additional CSI variables
Respiratory rate, breaths/min		Aspartate transaminase
≤22	0	Alanine transaminase
23–28	1	Ferritin
>28	2	Procalcitonin
Pulse oximetry, %*		Chloride
>92	0	C-reactive protein
89–92	2	Glucose
Oxygen flow rate, L/min		
≤88	5	Urea nitrogen
		WBC count
≤2	0	Age
3–4	4	
5–6	5	



Symbolic regression?

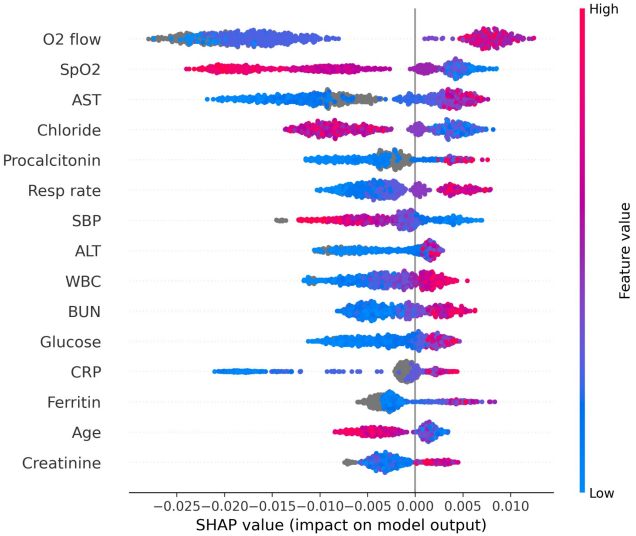


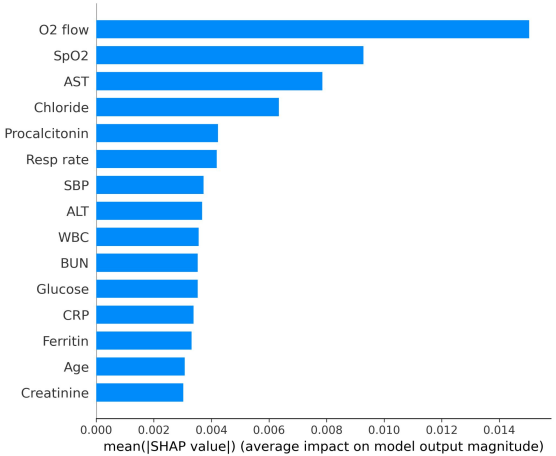


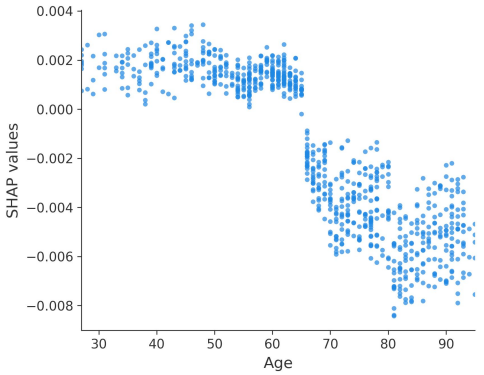
Haimovich A*, **Ravindra NG***, … van Dijk D, Taylor RD. *Annal. Emer. Med.* 2020

Sjoding et al. *NEJM*, Dec 17, 2020

Interpreting CSI to gain insight into resp. decomp. course







Scale CSI model output by isotonic regression so that SHAP values reflect a relative weighting of contributions

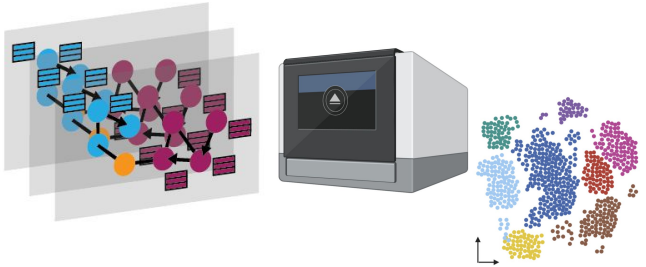
Consistent finding that inflammatory markers are suggestive of clinical course and low oxygen flow rates and high pulse oximetry values are protective but, in these EDs, younger pts are at heightened risk

Lundberg et al. *ICML*, 2017; Niculescu-Mizil & Caruana. ICML, 2005

Overview

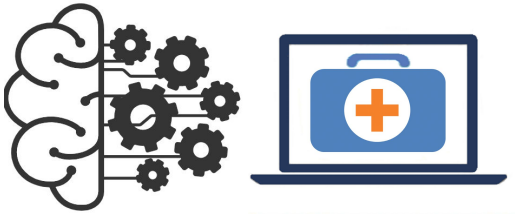
Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

- Dynamical genes from landmark time-points
- single-cell Graph Attention Networks (scGAT)



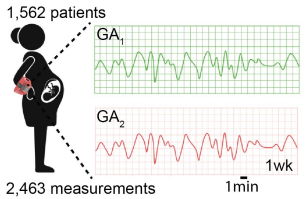
XAI to create clinically useful and parsimonious models

- qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department

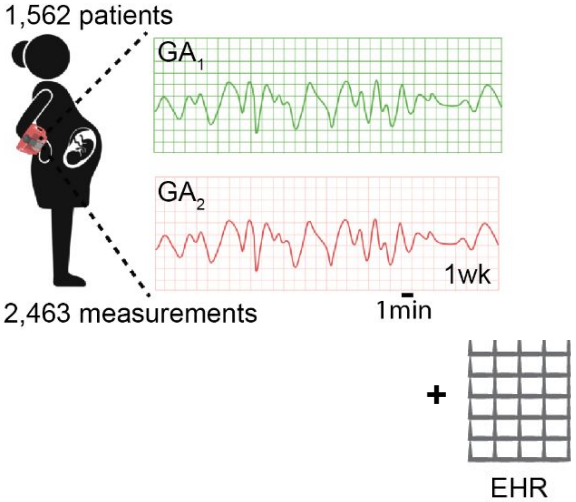


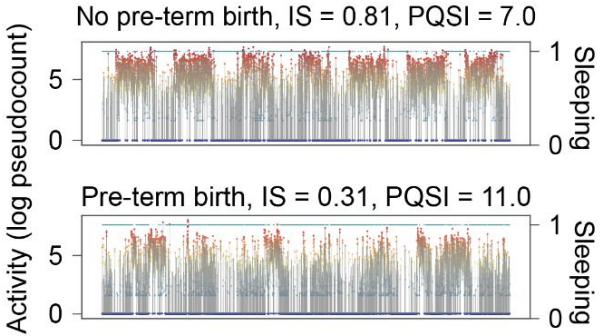
Moving from applications of ML/AI to fundamental research

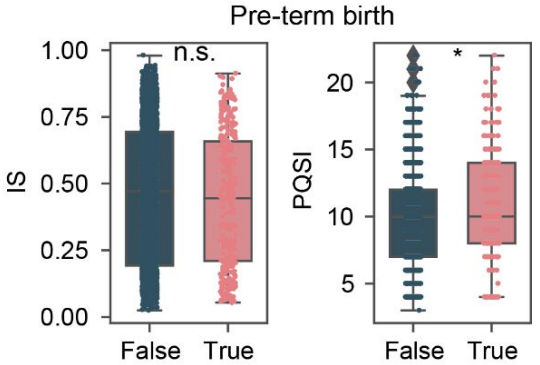
- **actigraphy2GA: sleep and activity disruptions and their relation to preterm birth**
- Permutation invariant networks to encode distributions
- sc2drug: perturbation modeling to align similar but disparate distributions



Standard analyses fail to indicate risk of preterm birth



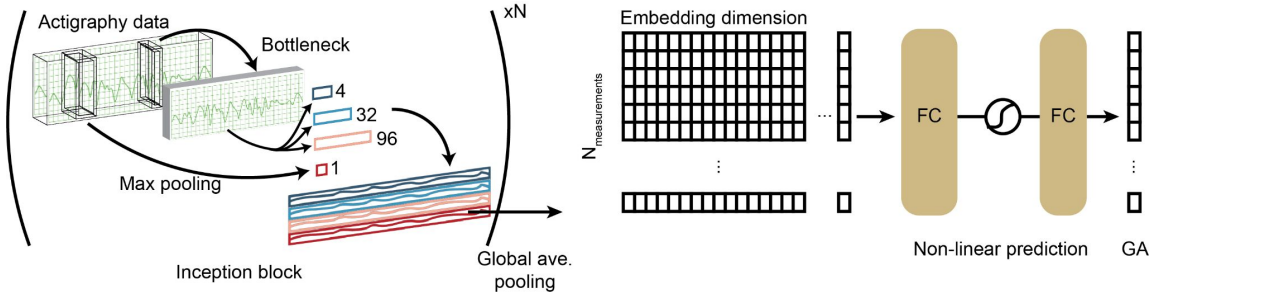


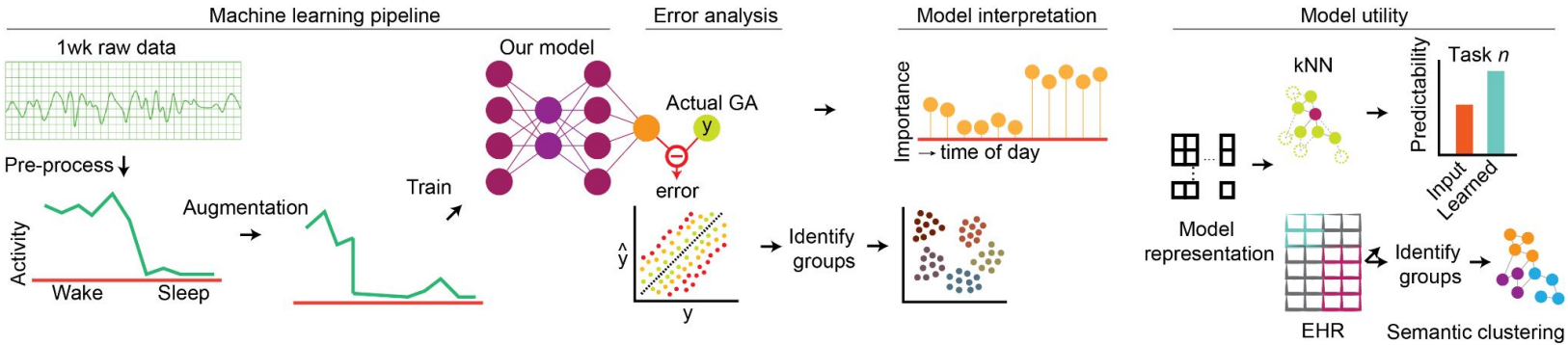


Ravindra, NG… Angst MS, Shaw GM, Stevenson D, Herzog E, Aghaeepour N. ResNet-inspired Deep Learning and post-hoc Inference Analyses of Physical Activity and Sleep Patterns During Pregnancy Identifies Model Deviations Associated with Prematurity. 2022. (*under review at Nat. Mach. Int.*)

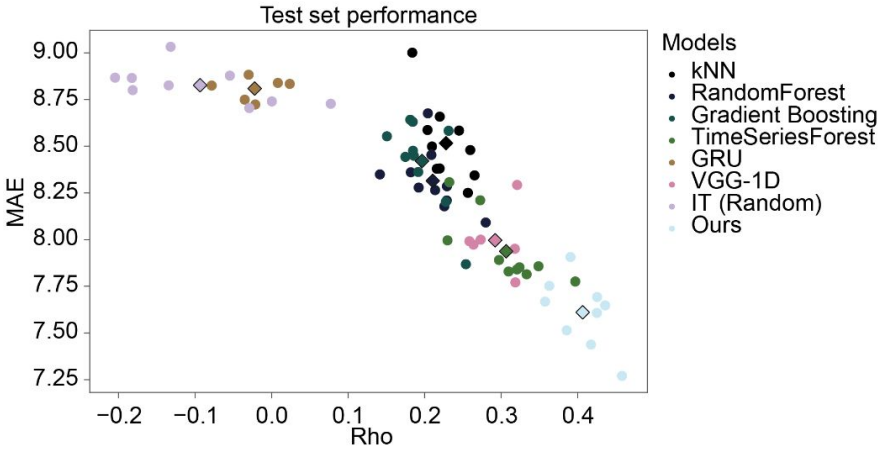
actigraphy2GA monitors pregnancy by applying deep learning to time series data monitoring activity and sleep

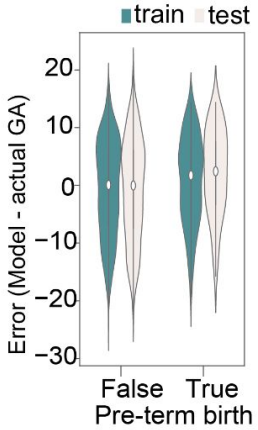
ResNet-inspired architecture

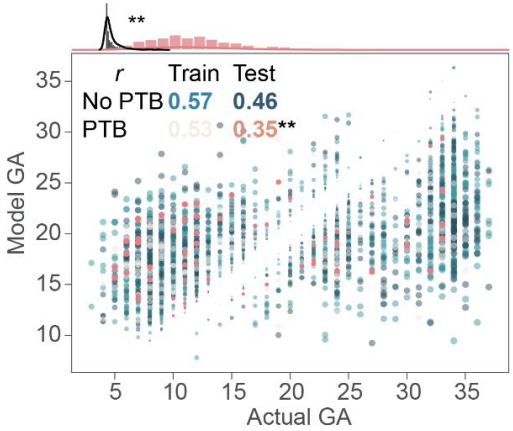




SOTA prediction of GA from 1wk of actigraphy data

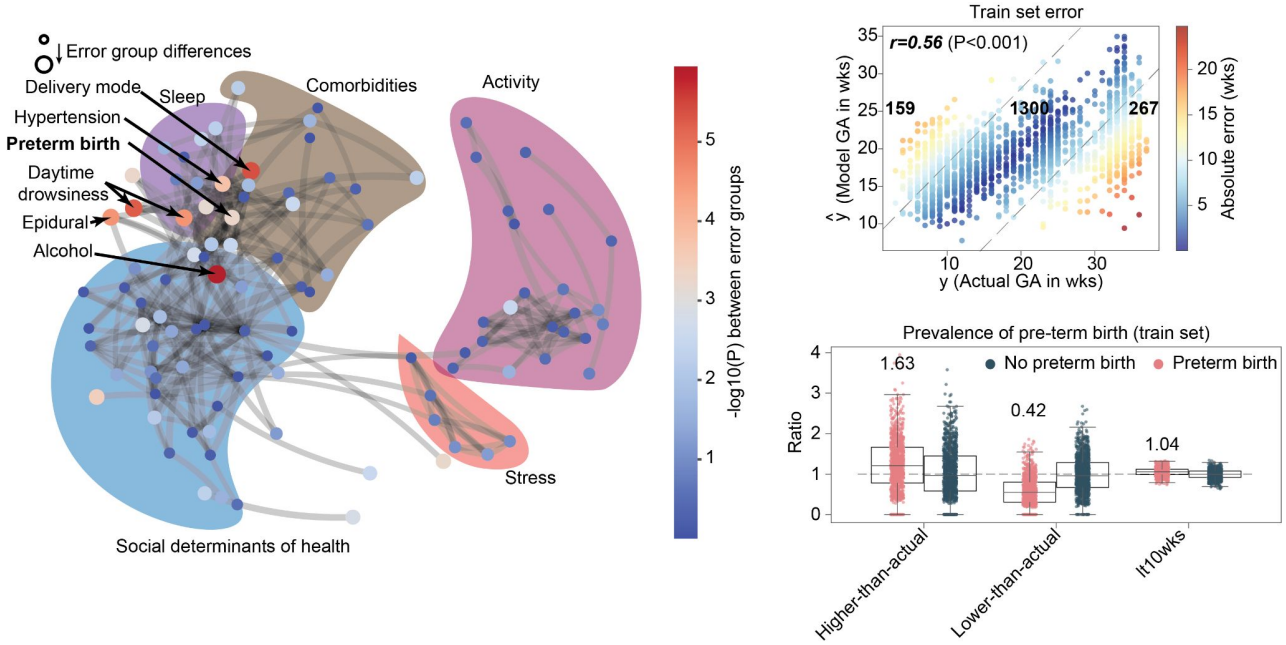




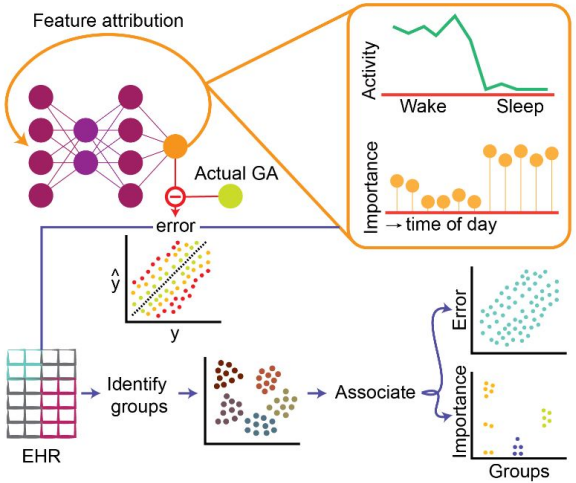


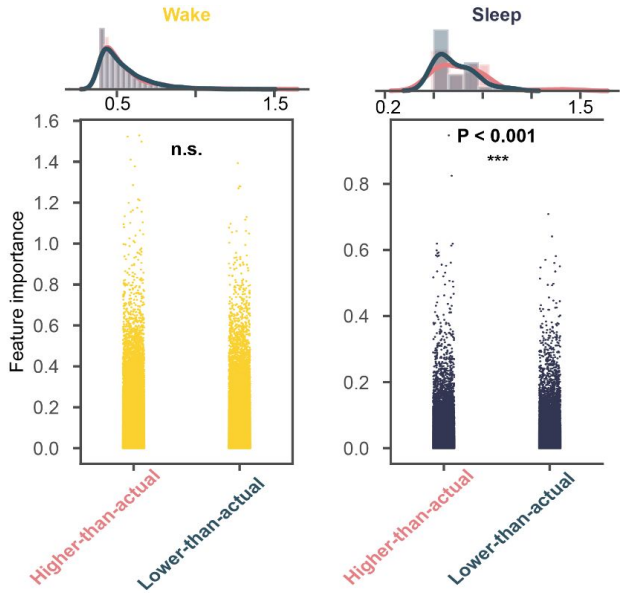
Results | wearables

Model error analyses reveals that actigraphy-GA indicates increased likelihood of adverse pregnancy outcomes

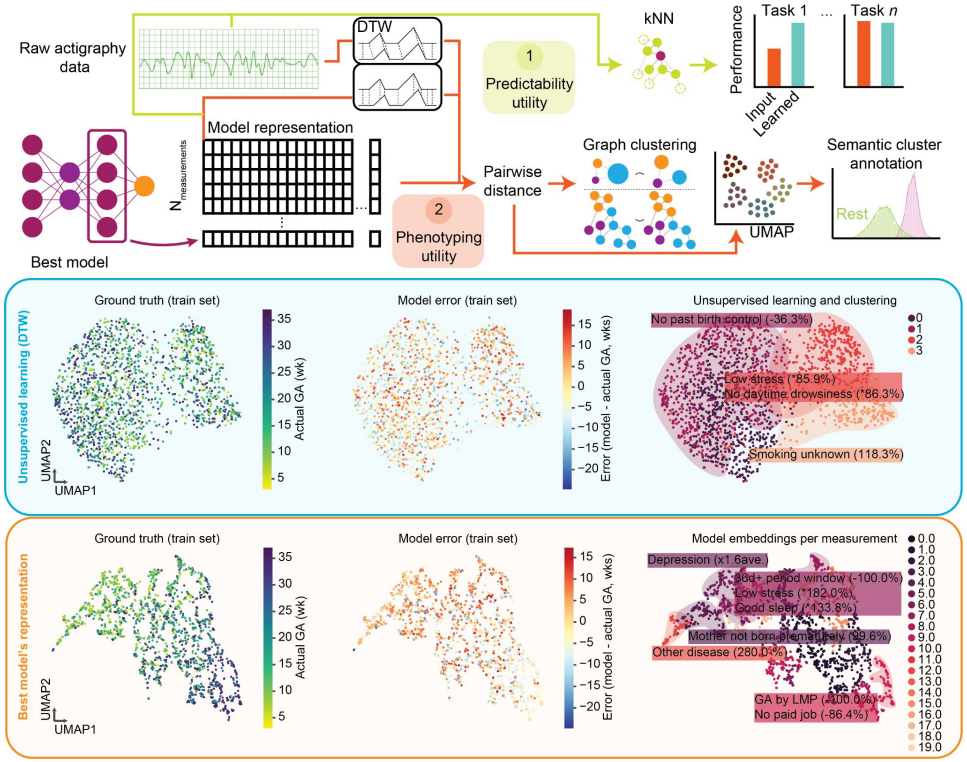


actigraphy2GA relies on deviations to sleep and activity in predicting higher- or lower-than actual GA





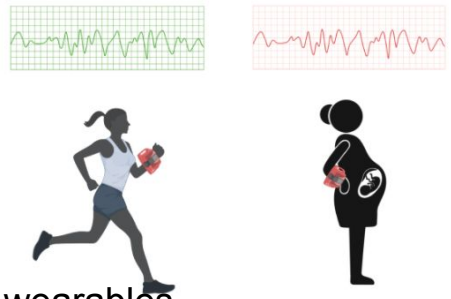
actigraphy2GA emeddings are useful for predicting ancillary tasks and semantic phenotyping



Characterizing biology of actigraphy-GA signaled patients

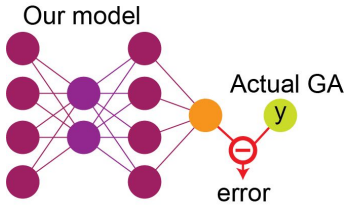
1

We can use actigraphy data to monitor pregnancy and identify risky behaviors



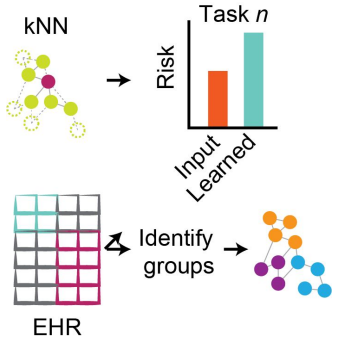
2

Sleep and activity disruptions cause the model to make errors



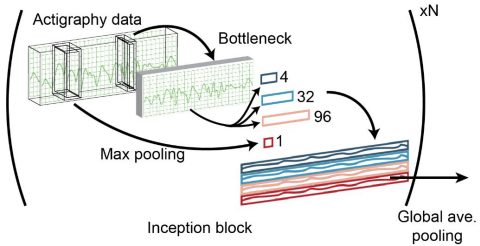
3

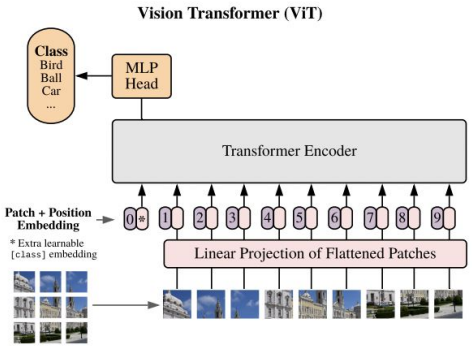
Interpreting the model shows we may be able to target inexpensive interventions

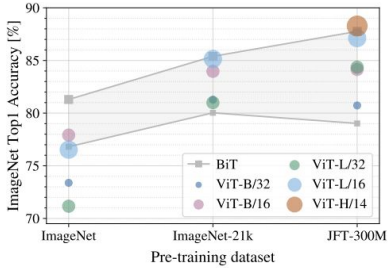


Mixing convolutions with attention & exploring pre-training

Vision Transformers for time-series

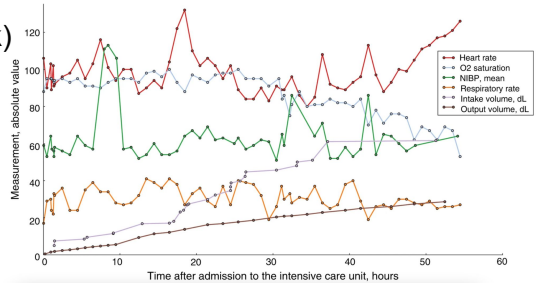






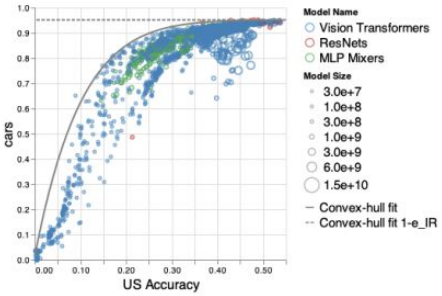
Experiments on queryable database of time-series data

MIMIC-IV (~40k)
UCR (<5k)
Financial data
⋮



Johnson et al. *Scientific Data*, 2016

Abnar et al. Exploring the Limits of Large Scale Pre-training. *ICLR'22*



Dosovitskiy et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR'21*

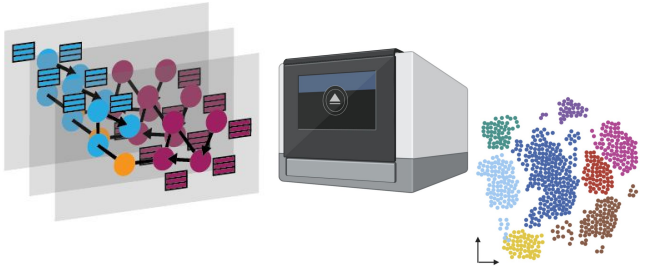
Future directions | wearables

Overview

Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

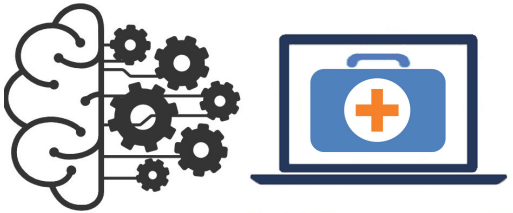
Dynamical genes from landmark time-points

single-cell Graph Attention Networks (scGAT)



XAI to create clinically useful and parsimonious models

qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department

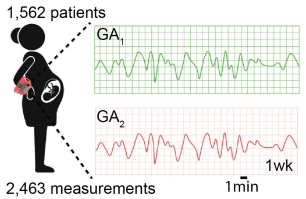


Moving from applications of ML/AI to fundamental research

actigraphy2GA: sleep and activity disruptions and their relation to preterm birth

Permutation invariant networks to encode distributions

sc2drug: perturbation modeling to align similar but disparate distributions



Wasserstein metric: principled way to compare distributions

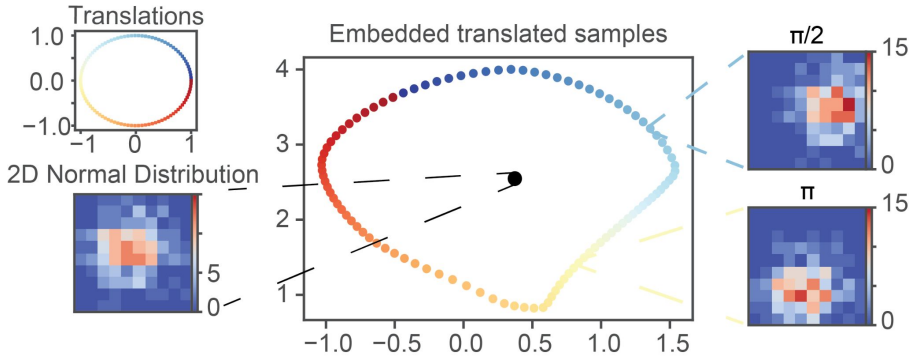
High-computational cost, not robust, non-differentiable distance

$$W_p(\mu,\nu) := \mathbf{inf}_{X\sim\mu,Y\sim\nu}\mathbb{E}(|X-Y|)^{1/p}$$

Can we learn a metric space of order *p*? What properties of the measures can we learn?

Do we learn something about the moments?

What properties of the original Wasserstein space can we preserve?

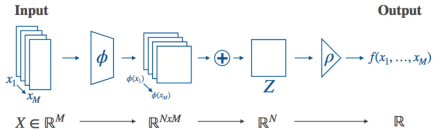


Sehanobish A*, **Ravindra NG***, van Dijk D. Permutation Invariant Networks to Learn Wasserstein Metrics. *TDA Workshop* at *NeurIPS'21*.

Encoding distributions for an optimal-transport map

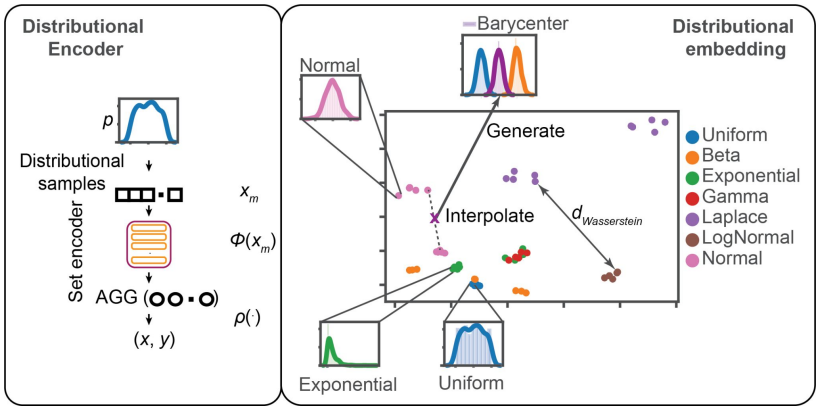
Draw samples from distributions in $\mathbb{P}(\mathcal{X})$

Encode with DeepSets



Train the encoder s.t.

$$\mathrm{SD}_p^{\lambda}(x, y) = ||H_{\theta(x)} - H_{\theta}(y)||$$

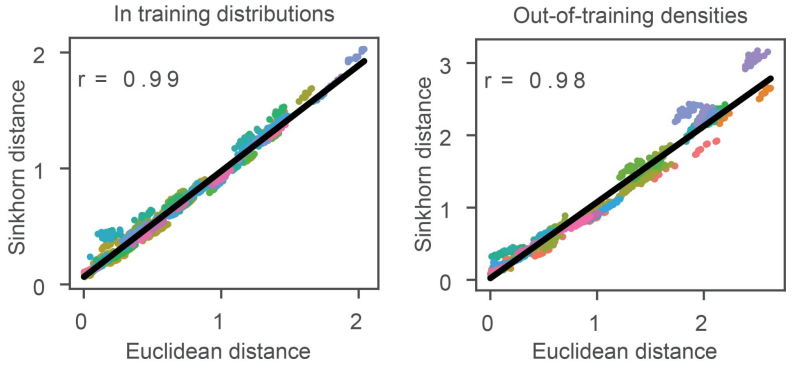


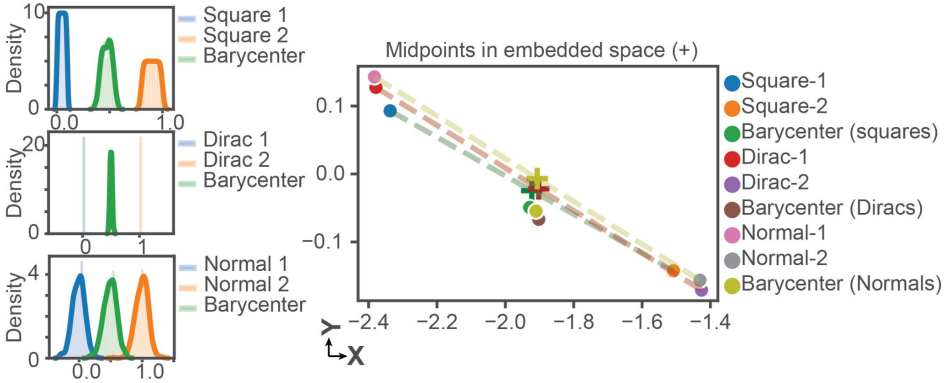
Regularize to enforce translation + scaling laws

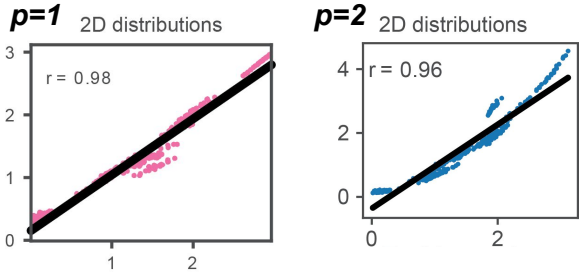
$$W_p(aX, aY) = |a| W_p(X, Y)$$

$$W_p(X + x, Y + y) = W_p(X, Y) \; \forall x \in \mathcal{X}$$

Barycenters from interpolation in learned space and generalization to out-of-training distribution samples







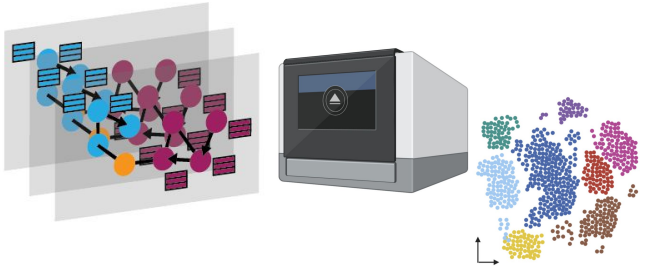
Optimizing transfer learning for small data via dataset similarity

OOD generalization and higher-dimensional mappings

Overview

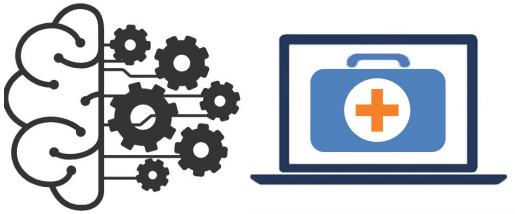
Interpretable ML to study molecular & cellular mechanisms of disease and cell state based on single-cell omics data

- Dynamical genes from landmark time-points
- single-cell Graph Attention Networks (scGAT)



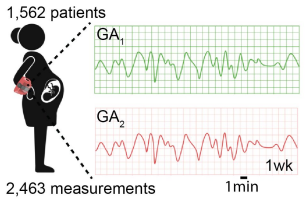
XAI to create clinically useful and parsimonious models

- qCSI from a custom COVID-19 Severity Index model for triaging patients in the emergency department

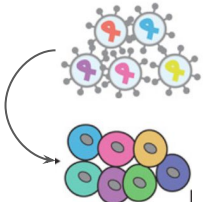


Moving from applications of ML/AI to fundamental research

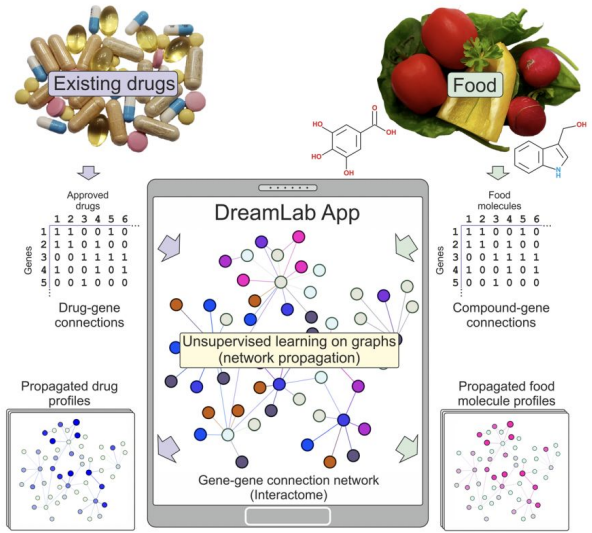
- actigraphy2GA: sleep and activity disruptions and their relation to preterm birth
- Permutation invariant networks to encode distributions
- **sc2drug: perturbation modeling to align similar but disparate distributions**



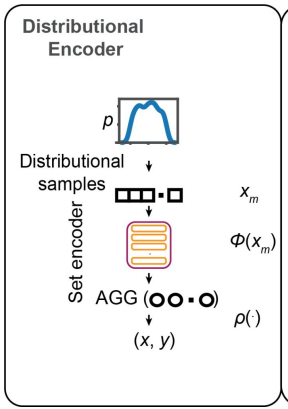
Preponderance of targets and determinants: combining recommender systems, representation and metric learning

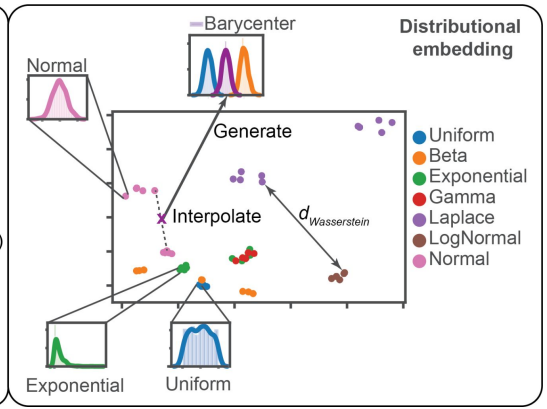


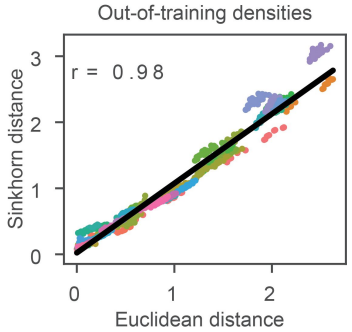
Perturbation modeling



Veselkov, K… Bronstein M et al. HyperFoods. *Sci. Rep.*, 2019

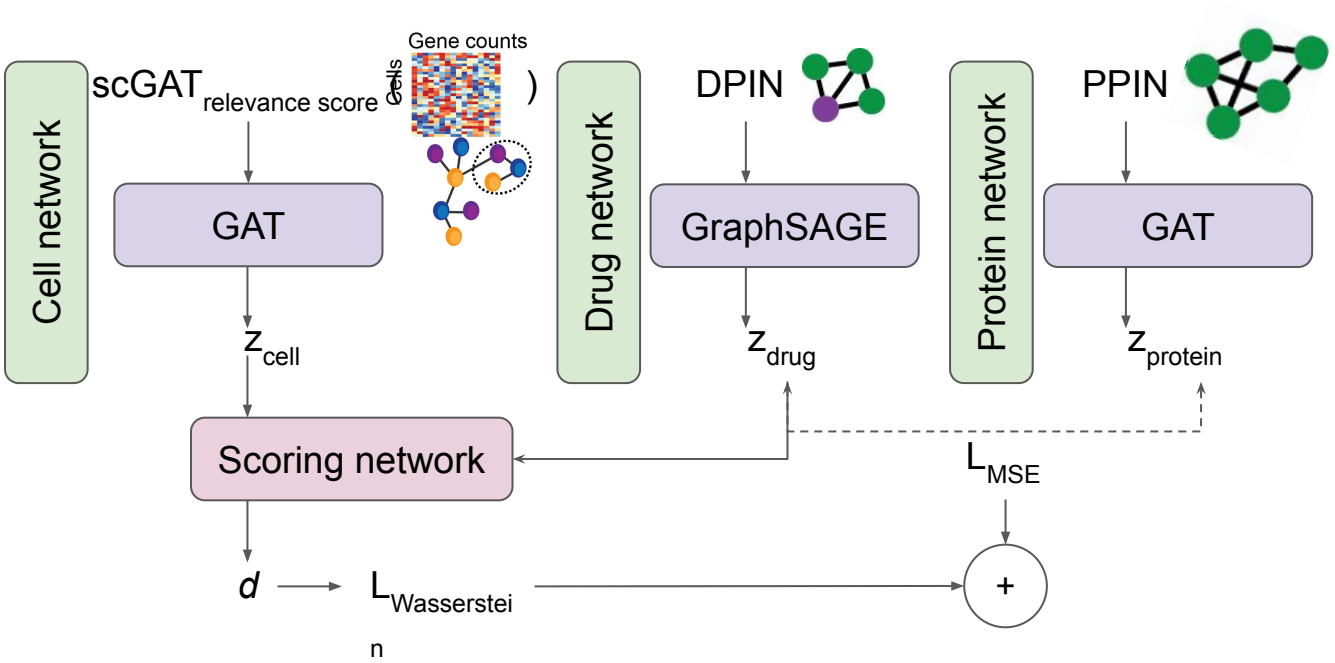






Sehanobish A*, **Ravindra NG***, van Dijk D. Permutation invariant networks to learn Wasserstein metrics. TDA workshop at *NeurIPS'20*

Constraining interpretability for omics-backed hypotheses



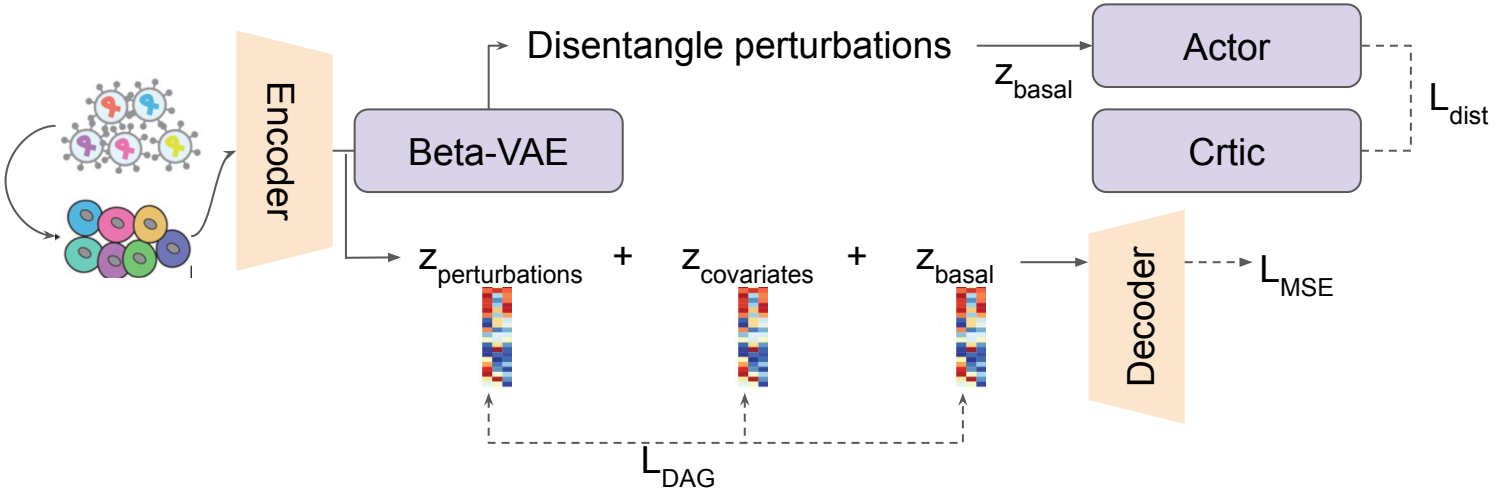

```
Rank 1: certoparin
Rank 2: dilazep
Rank 3: 2Q4
Rank 4: MHPP
Rank 5: chlordecone
Rank 6: AGN-PC-076G3F
Rank 7: acylsaligenin
Rank 8: Phenyl-1,2-Octanedione
Rank 9: adiphenine
Rank 10: AGN-PC-078YTI
Rank 11: nonachlor
Rank 12: ethoxyquin
```

?

Model	(1, 1)
sc2drug (DFA)	0.73 (0.40)
sc2drug (MSE loss)	0.34 (0.03)
sc2drug (Hinge loss)	0.34 (0.03)

Single-cell drug recommender systems: align representations of protein-protein, drug-protein, and cell-gene perturbations to suggest drugs to modulate genes varying across a particular cell subset; NLP to evaluate

Constraining interpretability for omics-backed hypotheses



Build on recent developments in DAG learning to infer causal dependencies between perturbed genes, perturbations, and covariates

Future directions

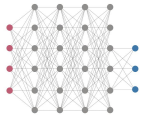
Acknowledgments

van Dijk Lab

Arijit Sehanobish
Victor Gasque
Jason Bishai
Antonio Fonesca
Aagam Shah
Yavuz Nuzumlali

Aghaeepour Lab

Eloise Berson
Joe T.P.
Davide de Francesco
Martin Becker
Shivam Saboo
Rishabh Gupta
Mingze Dong
Juanru Guo
David van Dijk



Nima Aghaeepour
Camilo Espinosa
Samson Mataraso
Ivana Maric

Collaborators

Craig B. Wilen, Yale
Mia Madel Alfajaro, Yale
Janghoo Lim, Yale
Leon Tejwani, Yale
Akiko Iwasaki, Yale
Adrian Haimovich, Yale
Andrew Taylor, Yale
Kristan Studemeyer, Stanford
Mike Snyder, Stanford
Jure Leskovec, Stanford
Trevor Hastie, Stanford
Stephanie C. Eisenbarth, Yale
Anna M. Pyle, Yale
Tamas L. Horvath, Yale
Bao C. Wang, Yale
Ellen F. Foxman, Yale
Richard W. Pierce, Yale
Tariq Ahmad, Yale
Nihar Desai, Yale
Erik Herzog, WashU
David Stevenson, Stanford
Gary Shaw, Stanford

Code: github.com/nealgravindra

Stanford Yale SCHOOL OF MEDICINE

NIH