Conflict of interest disclosures

I have no conflicts to disclose

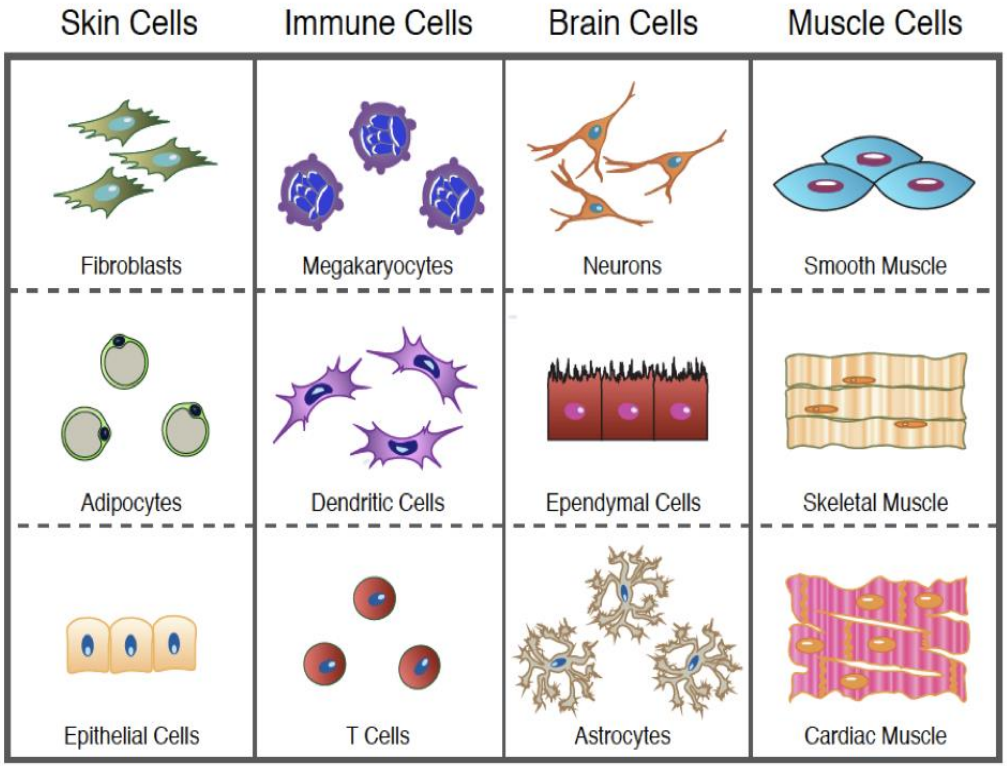
Identifying disease-critical cell types and cellular processes by integrating single-cell profiles and human genetics

Alkes Price
Harvard T.H. Chan School of Public Health

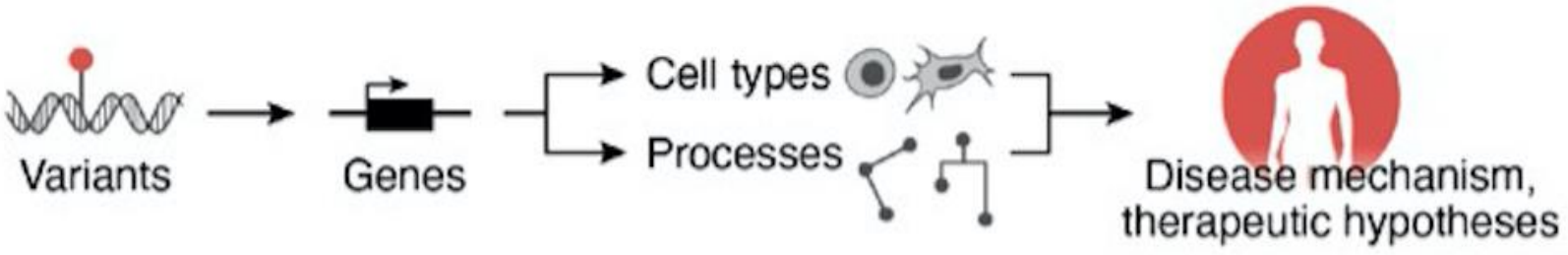
ASHG 2021 meeting
October 20, 2021

Jagadeesh*, Dey* et al biorxiv
(https://doi.org/10.1101/2021.03.19.436212)

Cellular dysfunction leads to disease



- Cells are the basic functional unit in biology.
- They are classified by structure, location, function, molecules
- Identifying disease-critical cell types is crucial to understanding disease biology

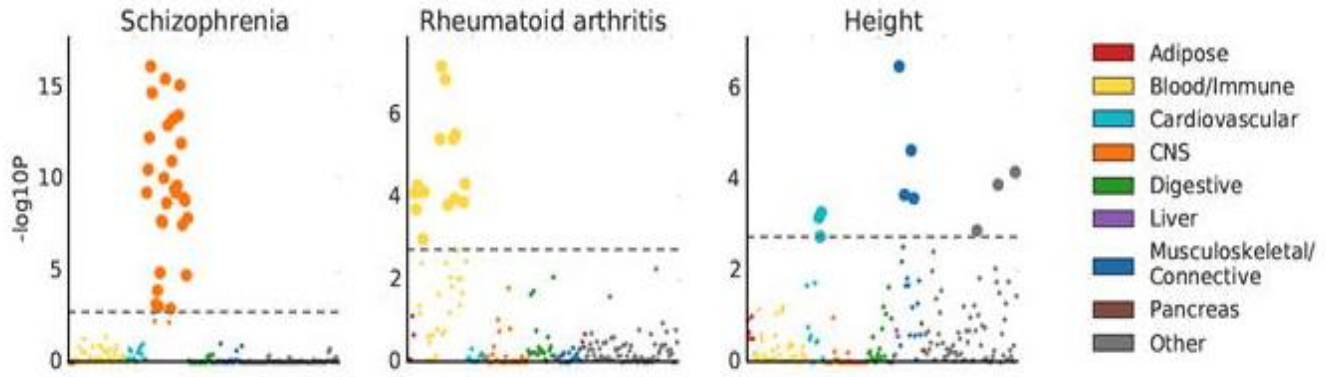


reviewed in Hekselman et al. 2020 Nat Rev Genet

Prior work linking cell types to disease using RNA-seq

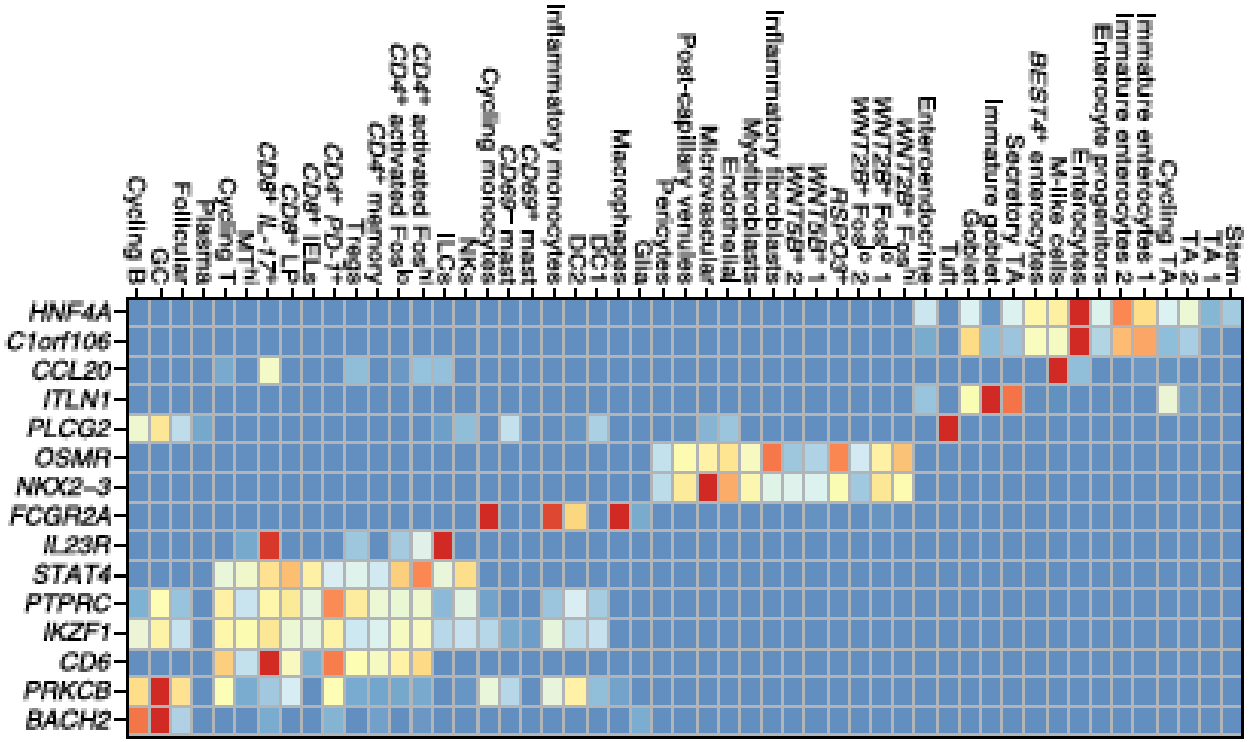
Finucane et al. 2018 Nat Genet

Bulk RNA-seq tissues and broad cell types + GWAS data



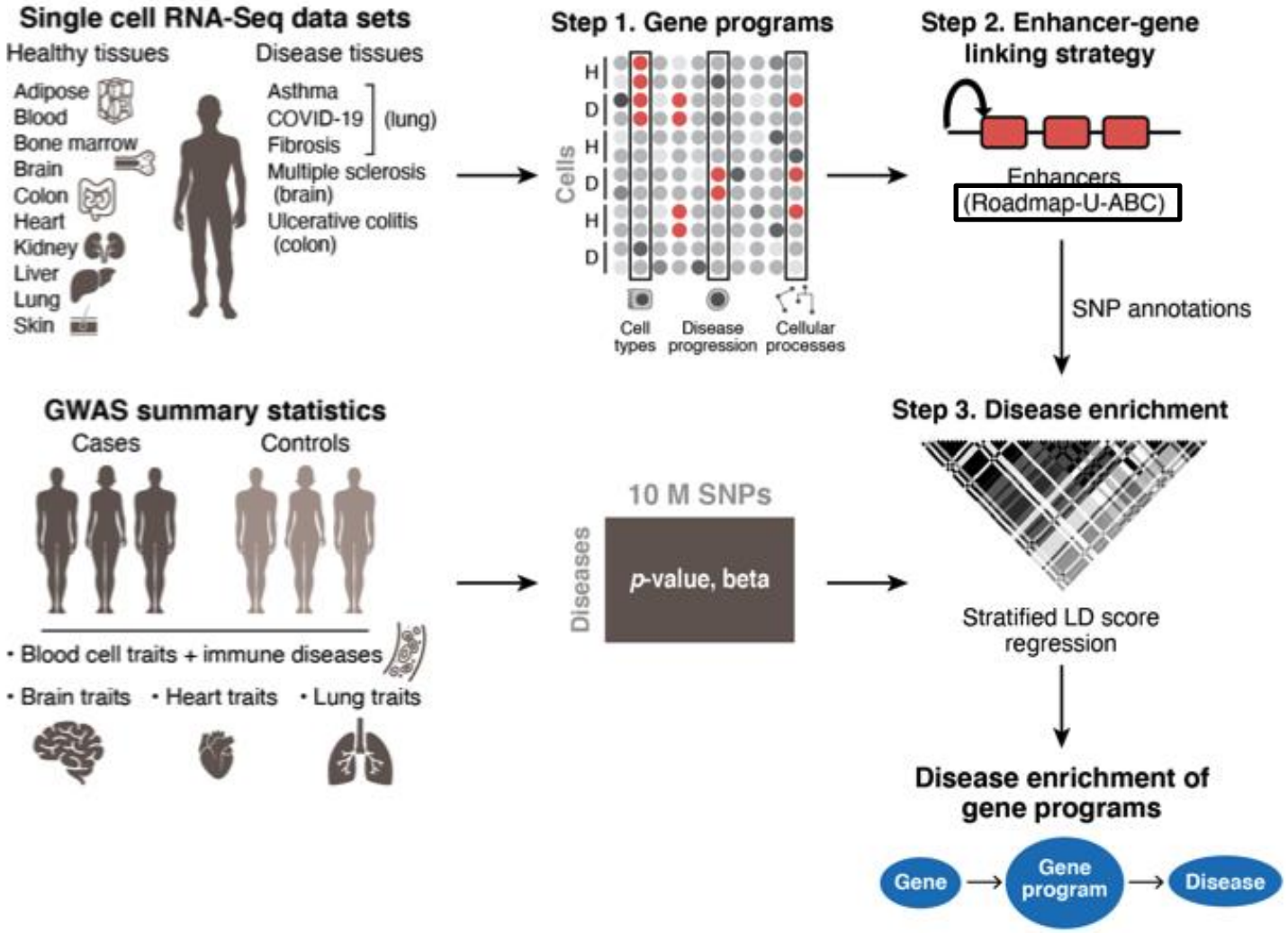
Smillie et al. 2019 Cell

UC scRNA-seq + individual GWAS genes



also see Bryois et al. 2020 Nat Genet (brain)

Integrating scRNA-seq with disease genetics

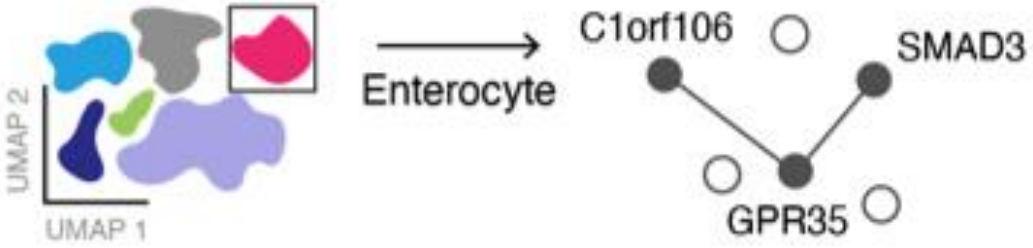


Primary metric: enrichment score corrected for all genes (**Escore**) and its pvalue

sc-linker: Jagadeesh*, Dey* et al biorxiv (https://doi.org/10.1101/2021.03.19.436212)

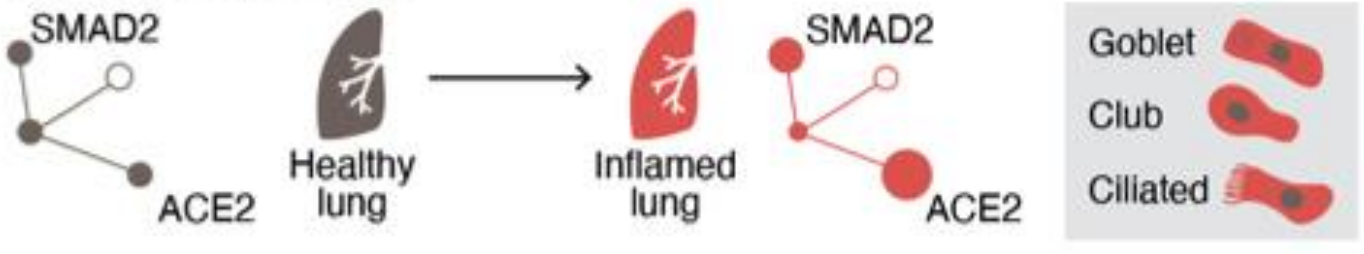
Outline: constructing gene programs from scRNA-seq

1. **Cell type**



2. **Cellular processes within/across cell types**

3. **Disease progression**



Outline

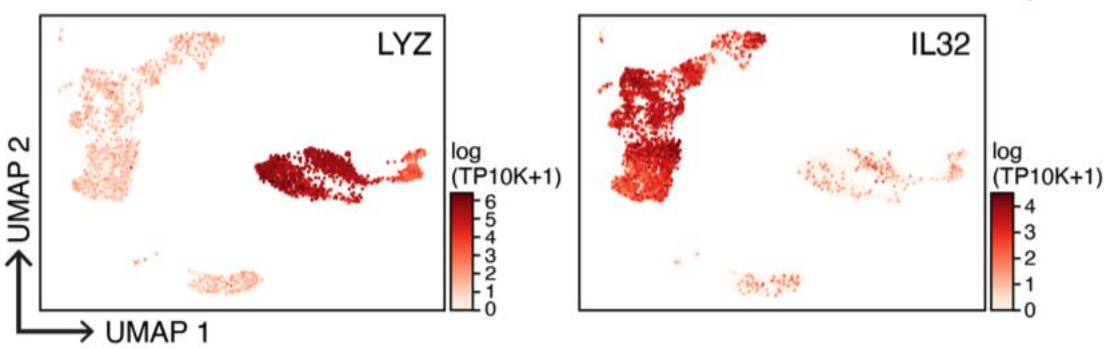
1. **Constructing cell type gene programs**
2. Constructing cellular process gene programs
3. Constructing disease progression programs

1. Constructing cell type gene programs

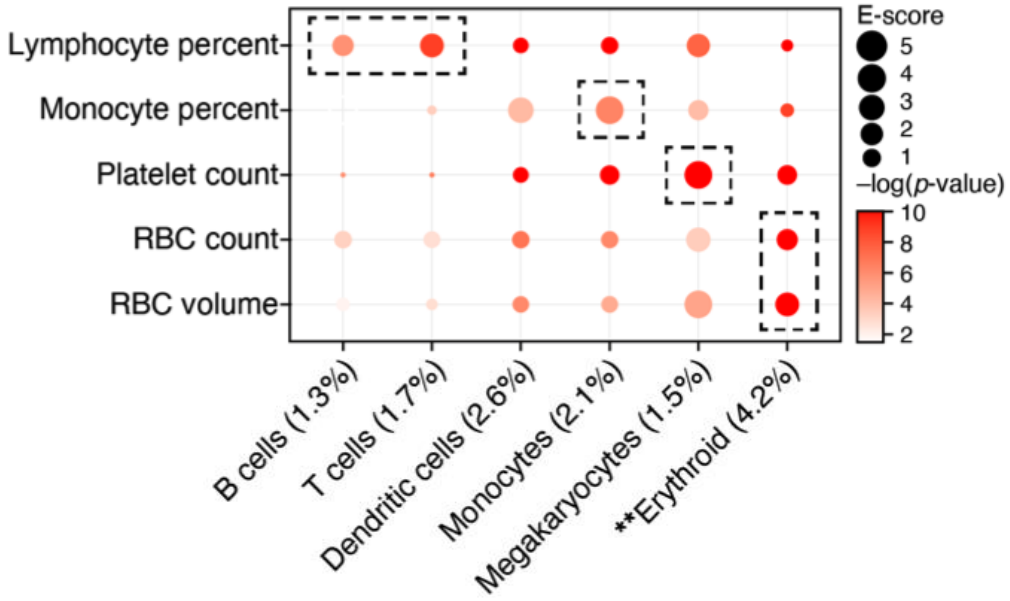
Cell type programs

Genes characterizing the most well understood functional unit – a cell type

Genes specifically expressed in an annotated cell type compared to other cell types in the same tissue

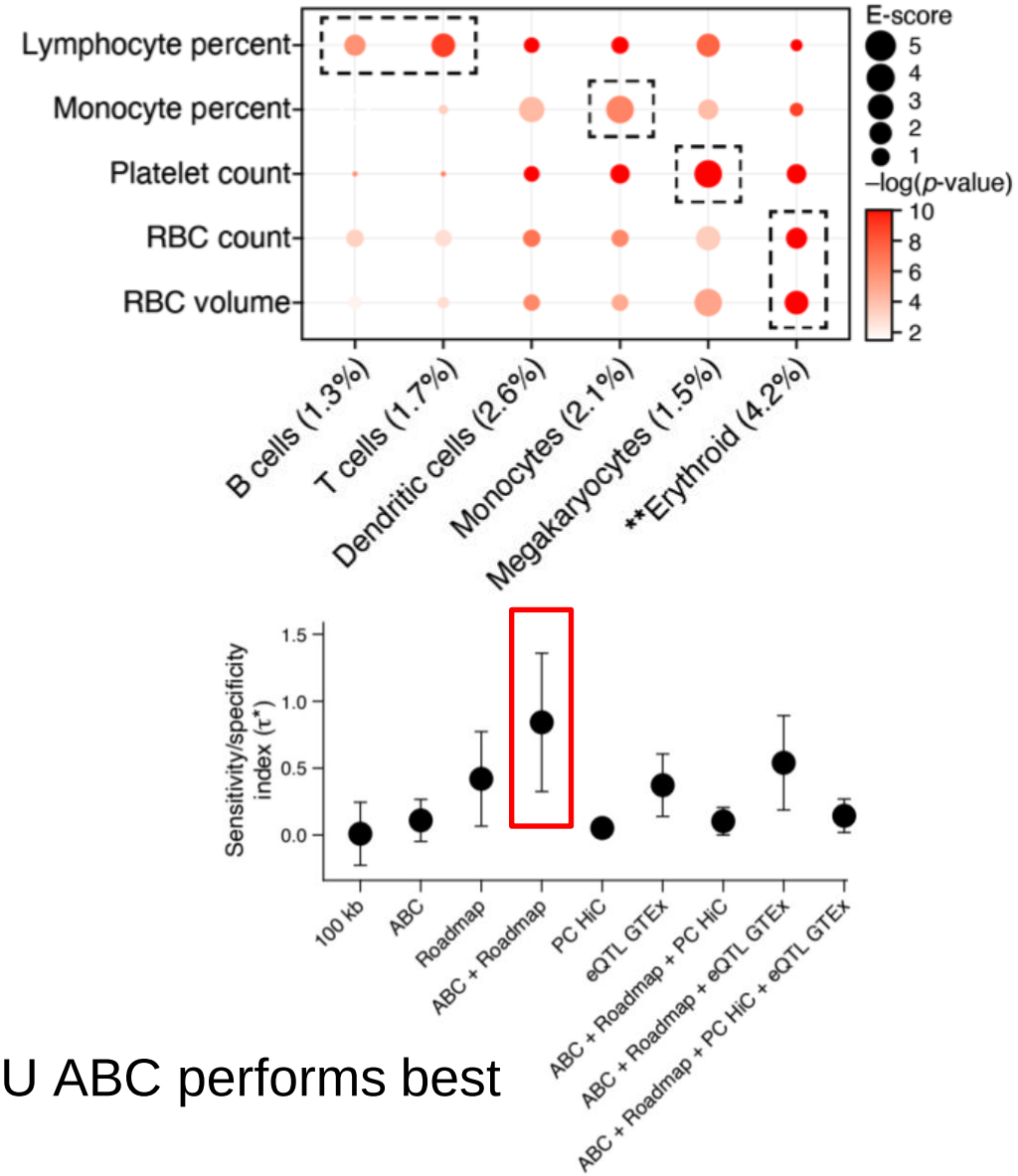


Linking blood cell types to blood cell traits



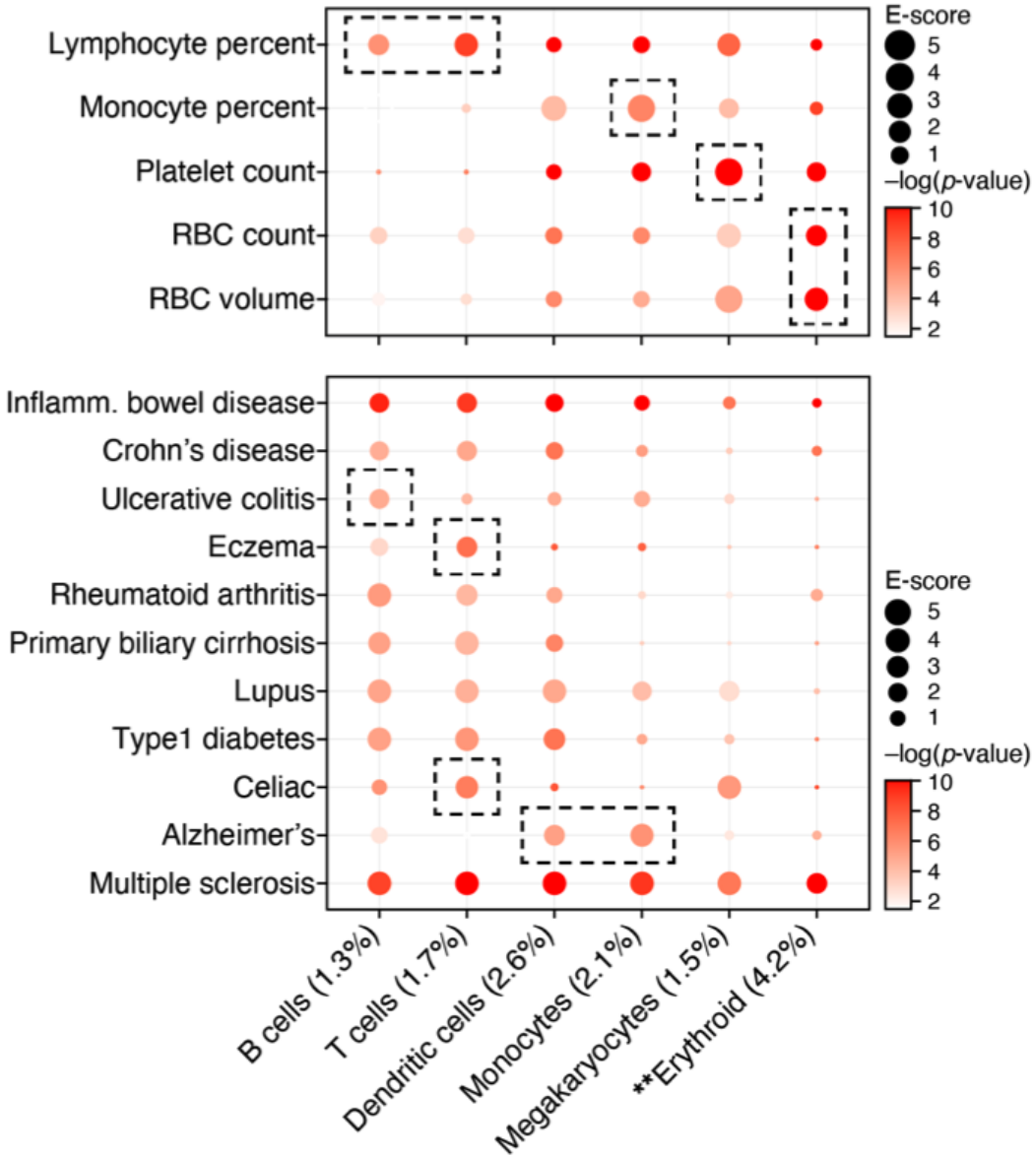
Confirmation of expected findings.

Validation of Roadmap U ABC S2G linking strategy

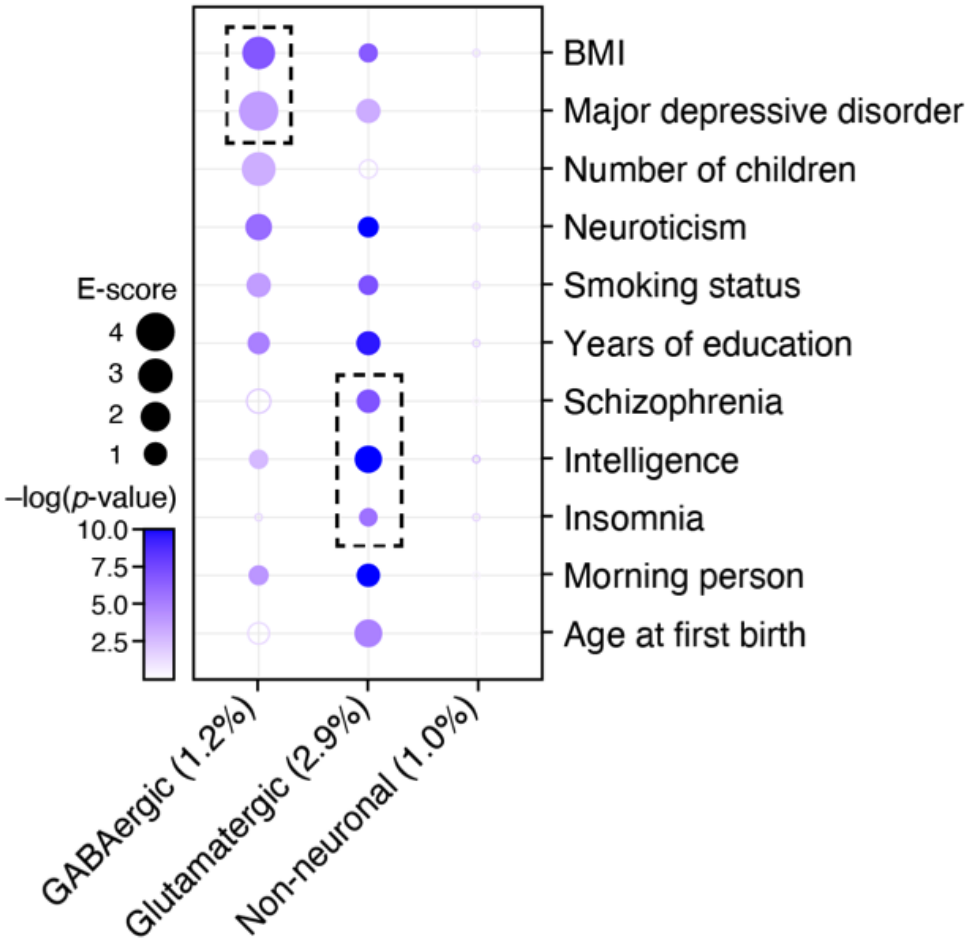


Roadmap U ABC performs best

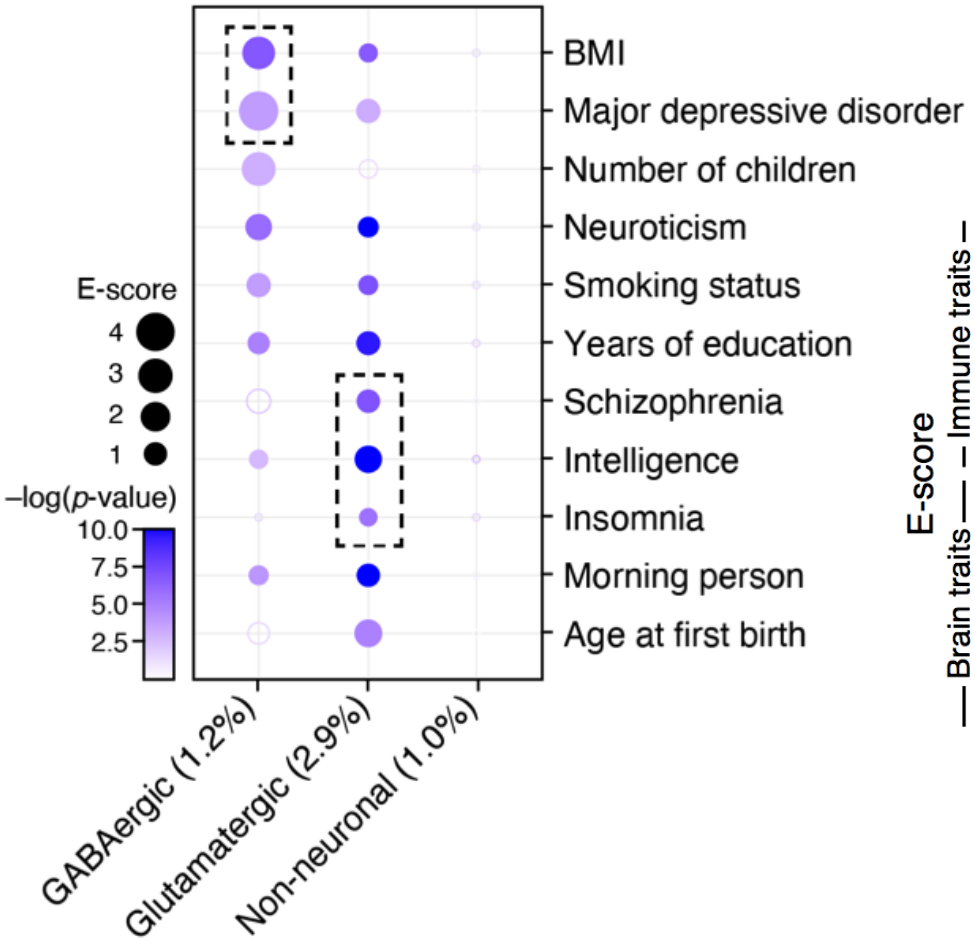
Linking blood cell types to immune diseases



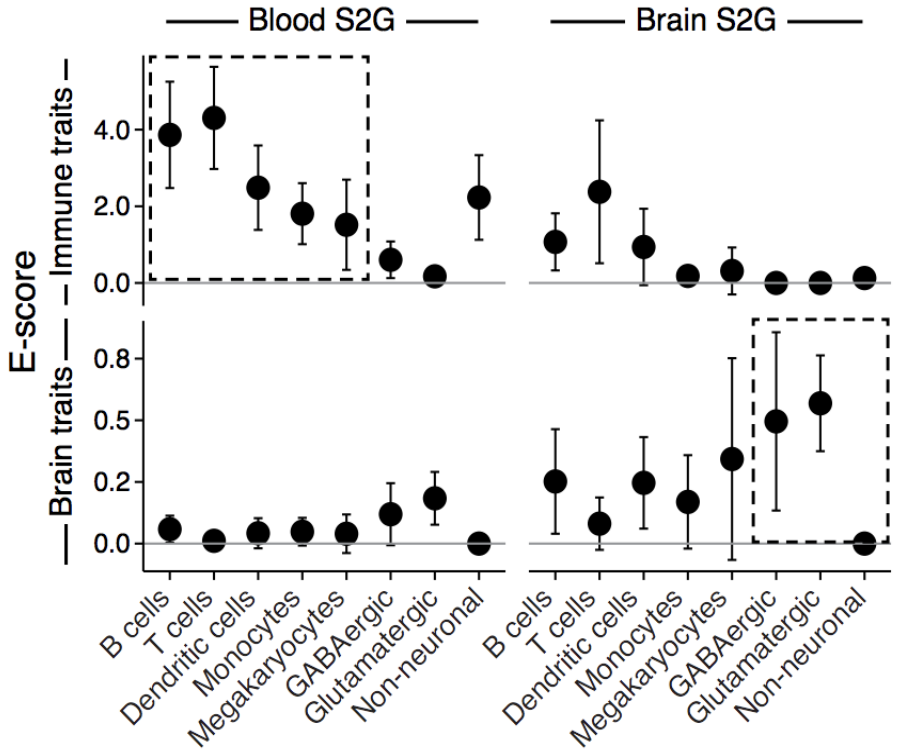
Linking brain cell types to brain diseases/traits



Importance of tissue-specific S2G linking strategy

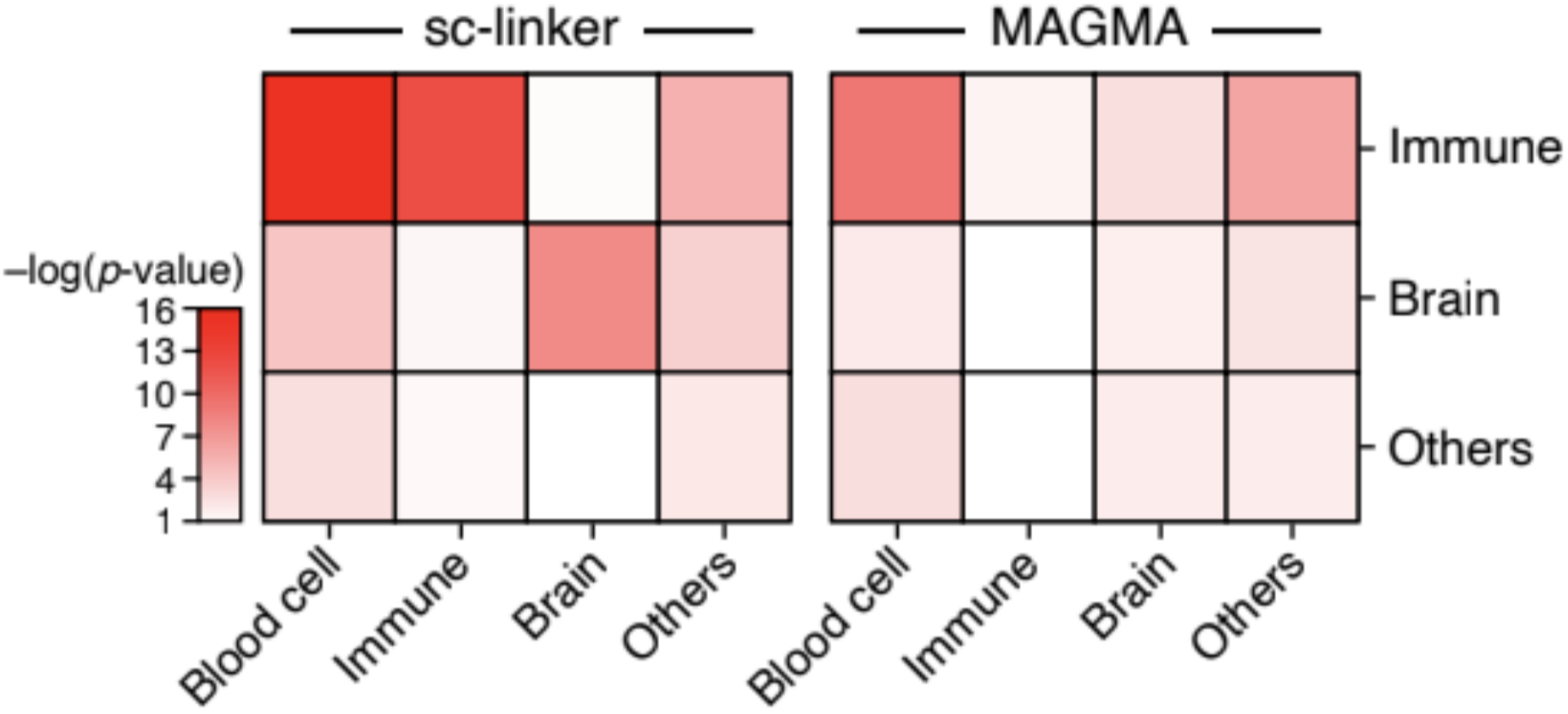


blood cell types x blood traits x <u>blood S2G</u>
brain cell types x brain traits x <u>brain S2G</u>



Comparing sc-linker vs. MAGMA gene set score

MAGMA: de Leeuw et al. 2015 PLoS Comput Biol



average $-\log_{10}$ P-values of 11.3 for sc-linker vs. 4.4 for MAGMA for cell type-disease/trait pairs in the most biologically plausible categories

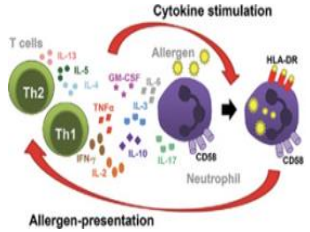
Outline

1. Constructing cell type gene programs
2. **Constructing cellular process gene programs**
3. Constructing disease progression programs

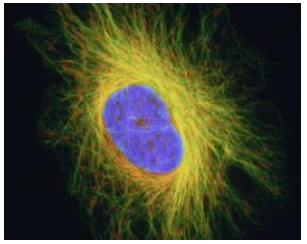
2. Constructing cellular process gene programs

Cellular process programs

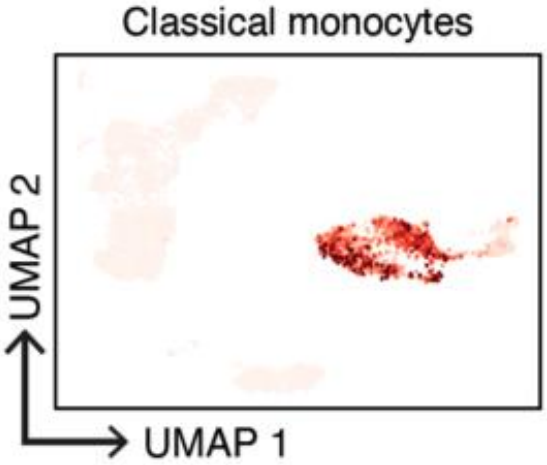
Allergen Response

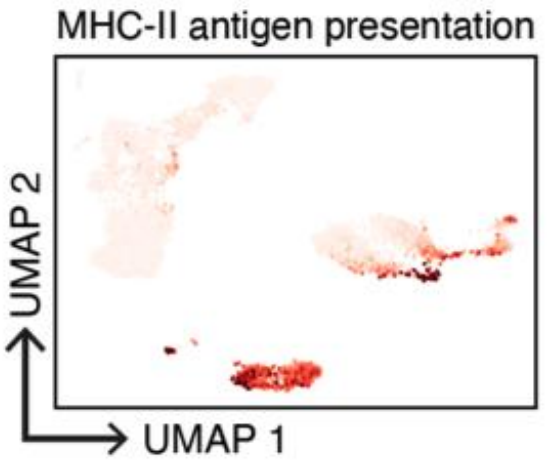


Cytoskeleton

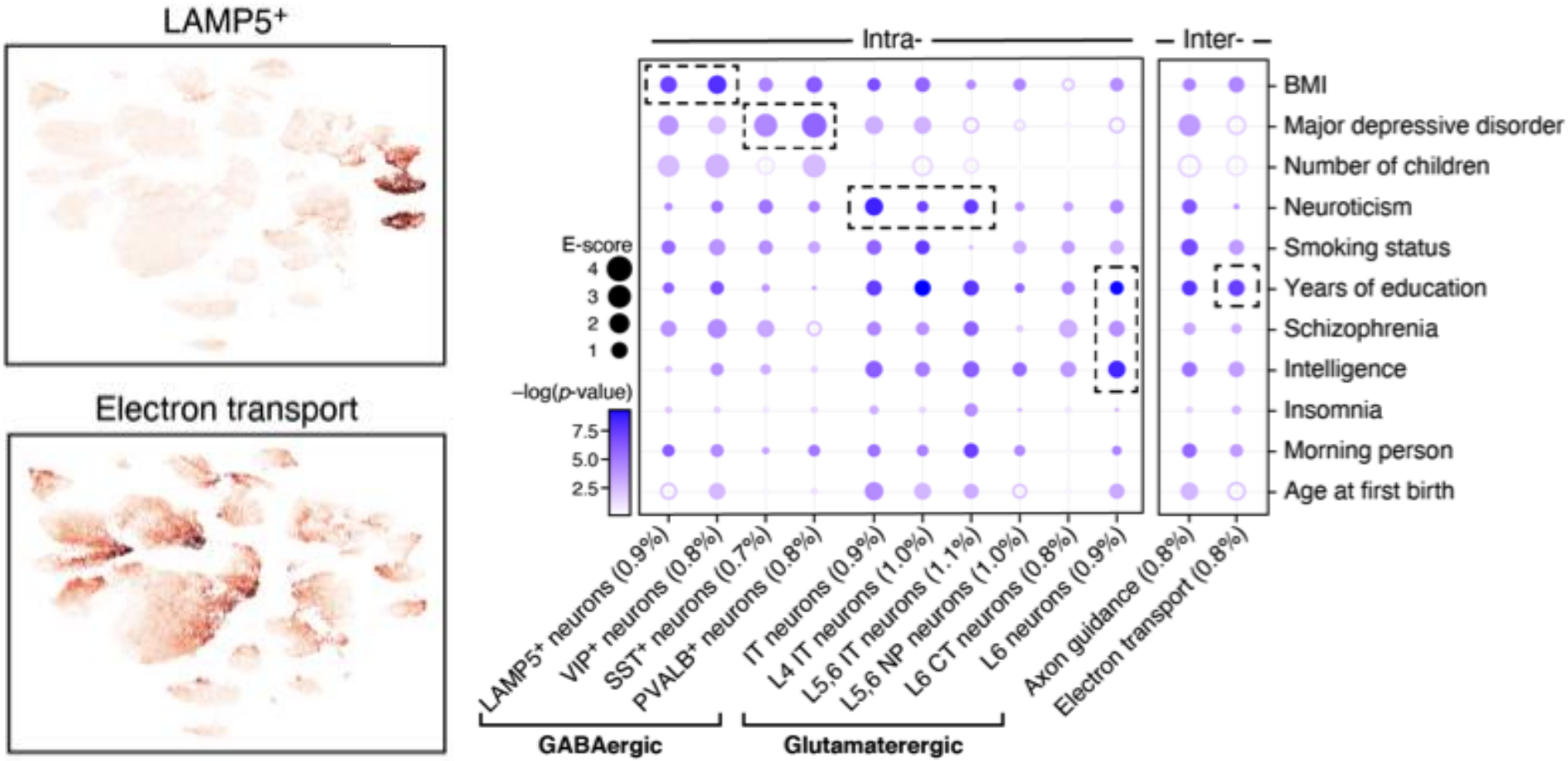


Genes characterizing cellular processes within or across cell types using unsupervised NMF-based approach (not using marker genes)





Linking brain cellular processes to disease



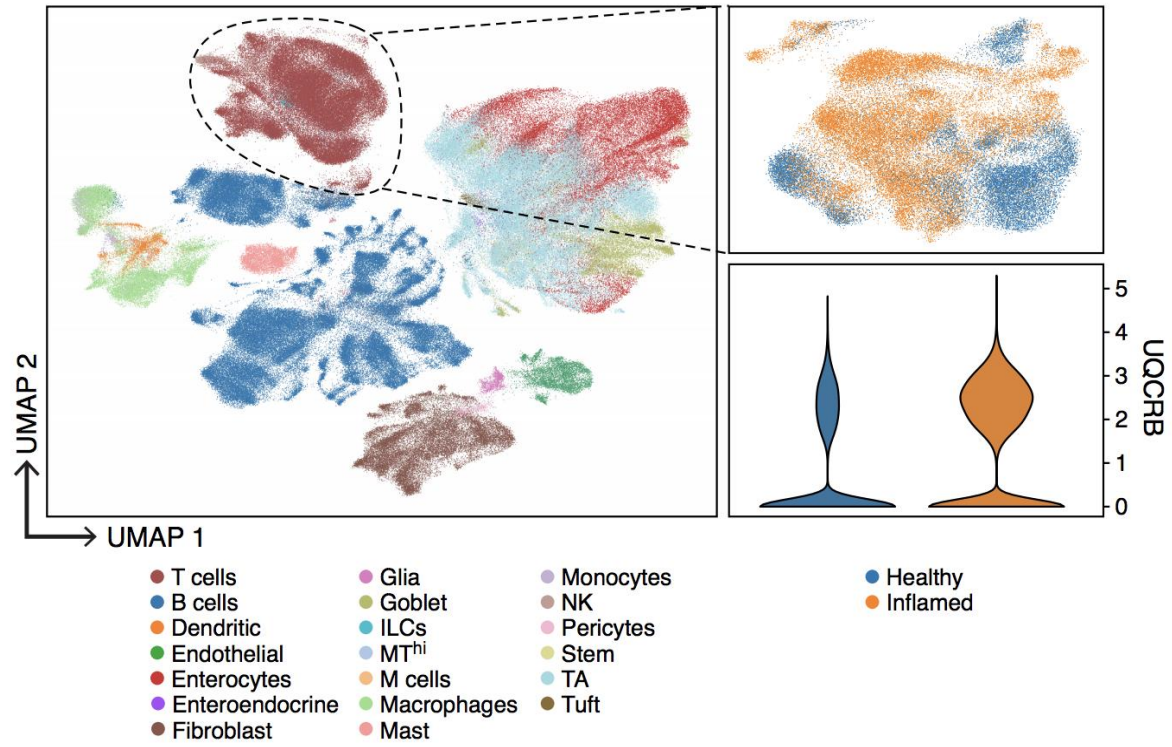
Outline

1. Constructing cell type gene programs
2. Constructing cellular process gene programs
3. **Constructing disease progression programs**

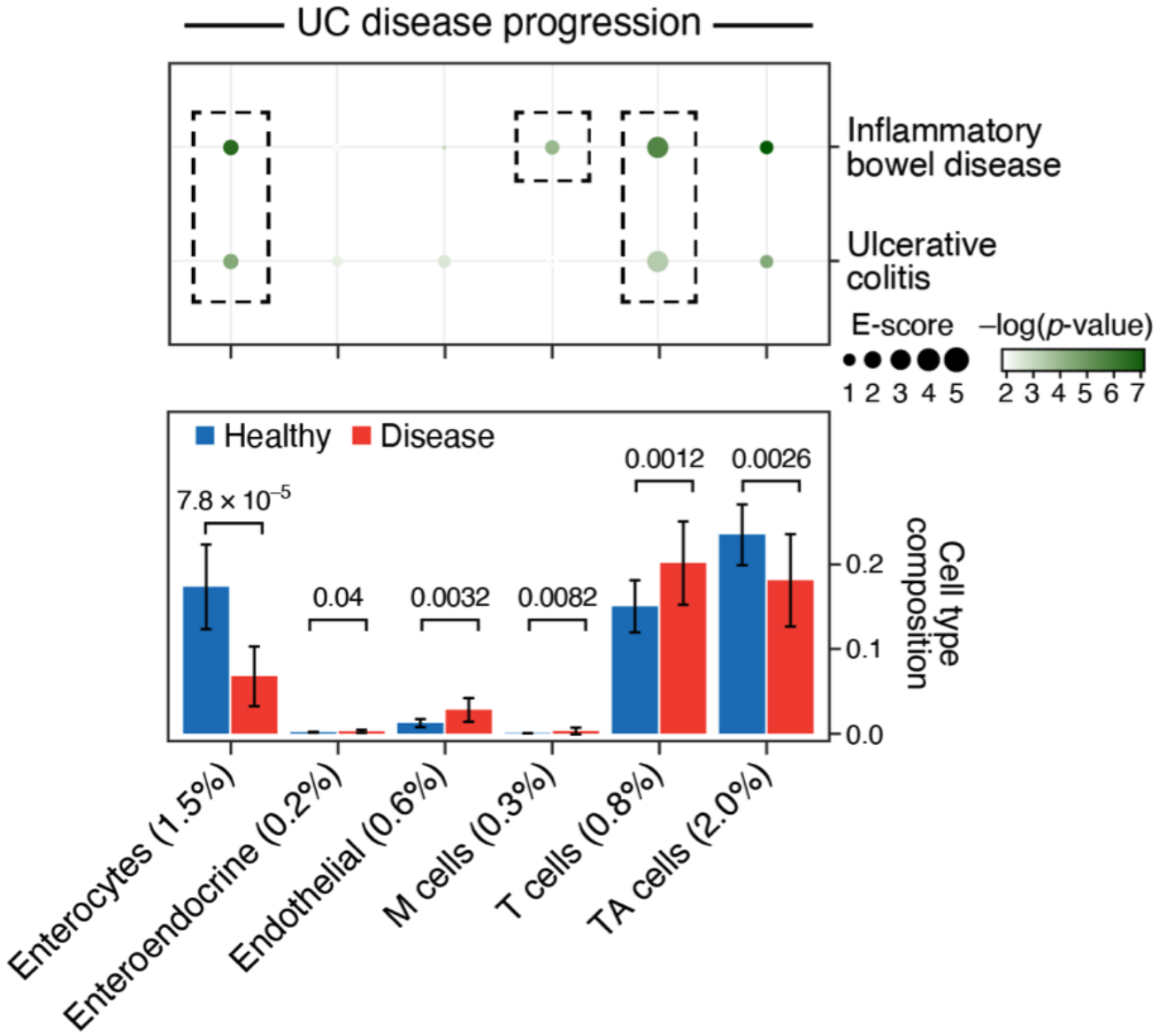
3. Constructing disease progression programs

Disease progression programs

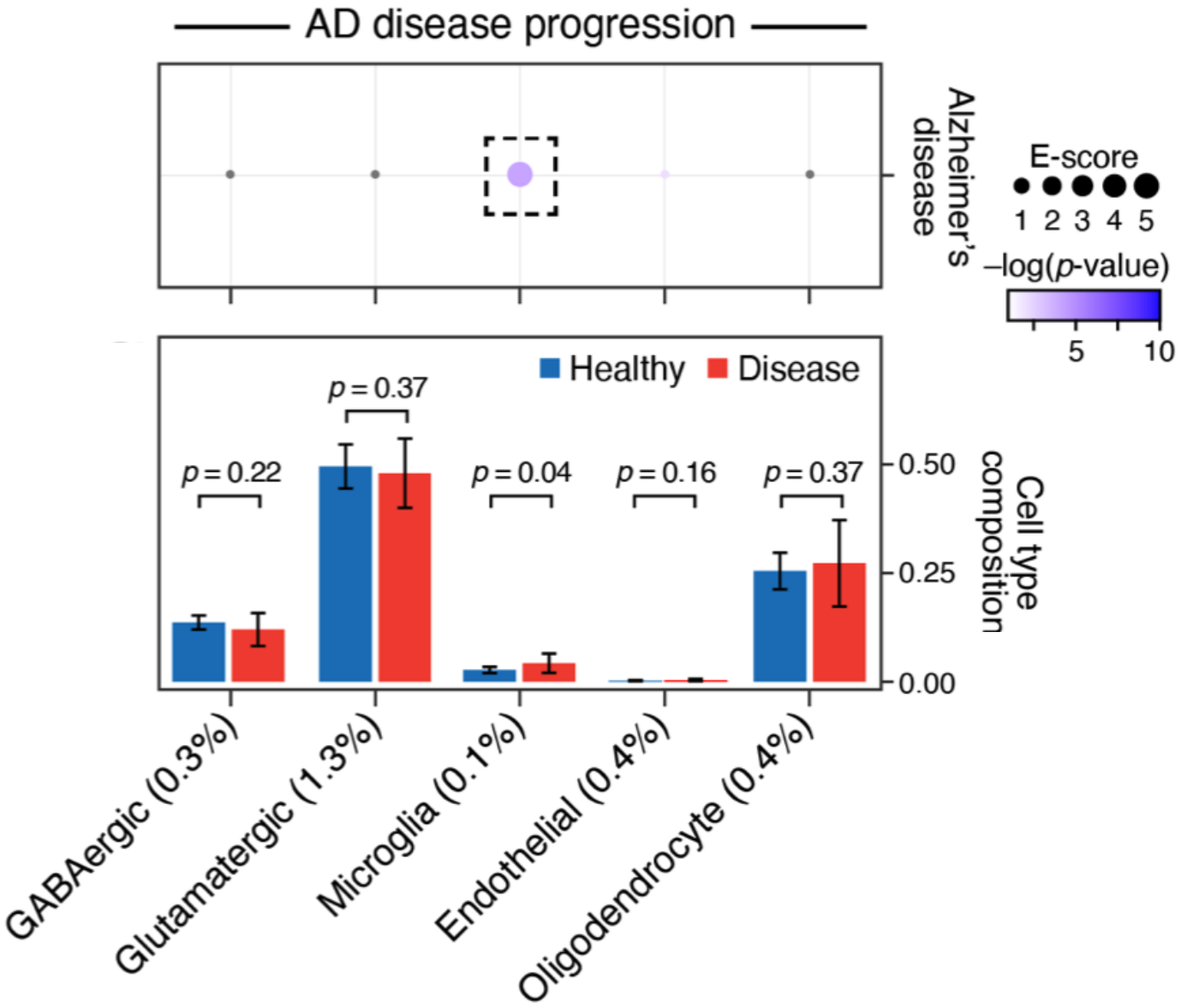
Genes specifically expressed in disease samples compared to healthy samples *in the same cell type*



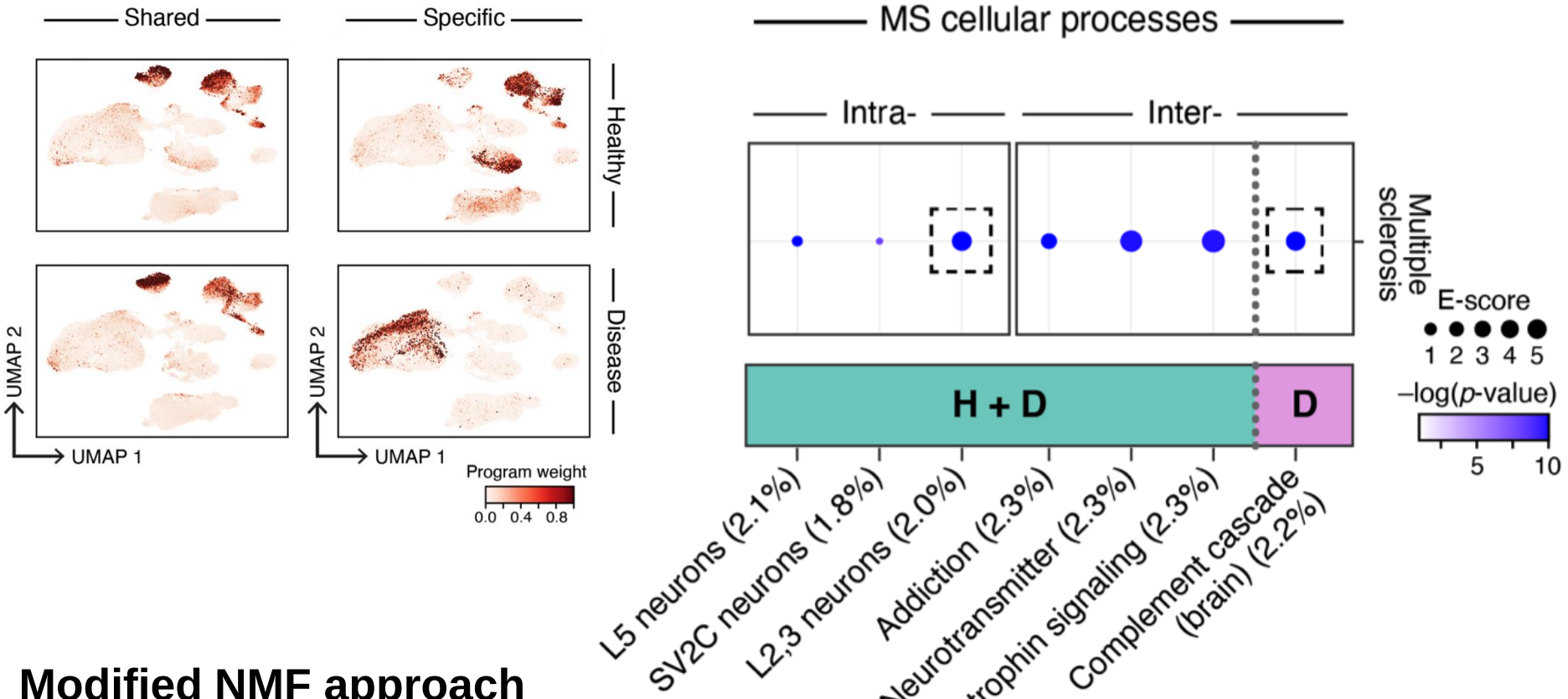
Enterocytes and M cells disease progression are critical for ulcerative colitis



Microglia disease progression program is critical for Alzheimer's disease



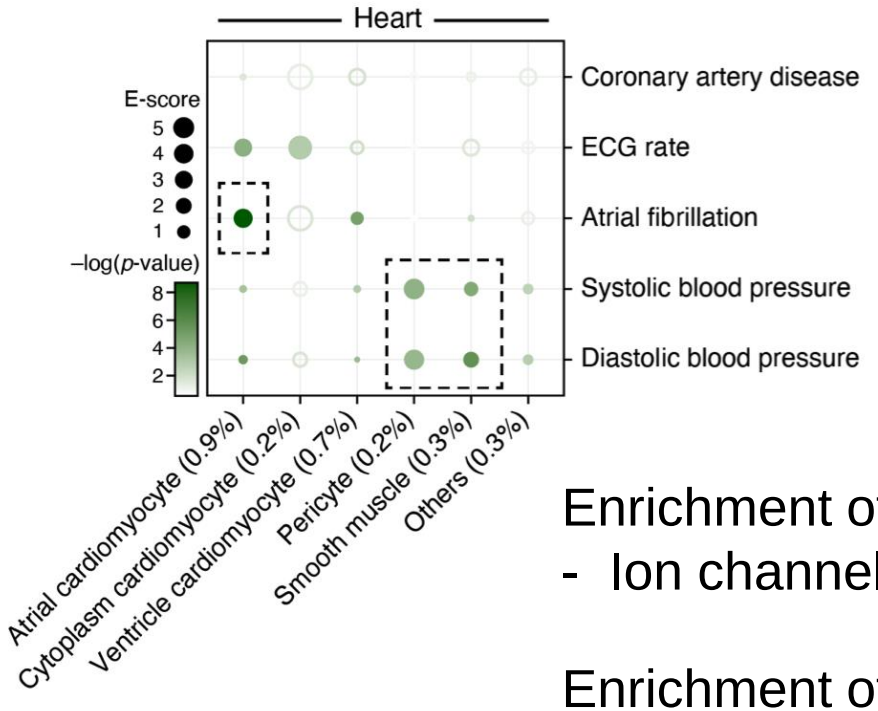
Constructing cellular process programs from case-control sc-RNA-seq data



Modified NMF approach
D = disease-specific cellular processes
H+D = shared cellular processes (healthy + disease)

Identifying disease-critical genes in top gene programs e.g. disease-critical (gene, cell type) pairs

- Prioritize genes based on grade > 0.8 in gene program
  + MAGMA gene-level score (de Leeuw et al. 2015 PLoS Comput Biol)

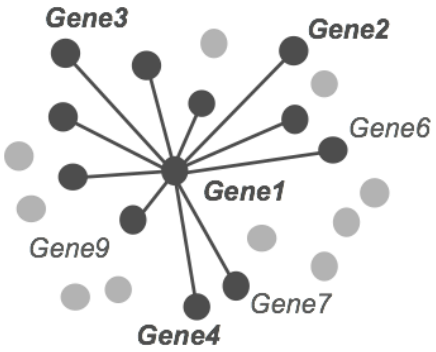


Enrichment of atrial fibrillation in atrial cardiomyocytes:
- Ion channel genes including *KD2L2*

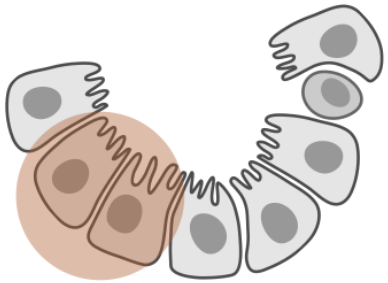
Enrichment of systolic blood pressure in pericytes:
- Adrenergic pathway genes including *PLCE1*
- Nitric oxide pathway genes including *GUCY1A3*

Summary: A refined vocabulary of disease

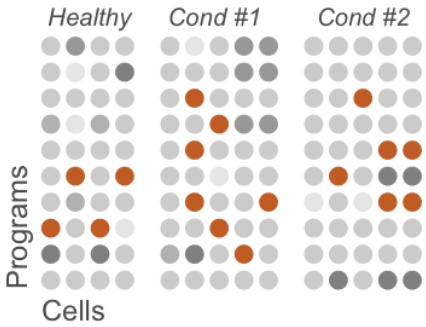
Learning gene programs: cell types and beyond



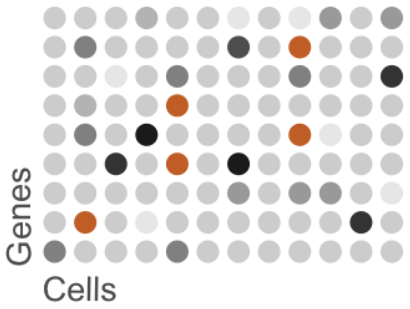
Identify cell type of action



Pinpoint disease specific gene programs



Prioritize genes in specific cellular context



Acknowledgements

Karthik Jagadeesh

Kushal Dey

Aviv Regev

Also:
Steven Gazal (HSPH/USC)
Daniel Montoro (Broad Institute)
Ramnik Xavier (Broad Institute)
Joseph Nasser (Broad Institute)
Jesse Engreitz (Broad/Stanford)

Jagadeesh*, Dey* et al biorxiv (https://doi.org/10.1101/2021.03.19.436212)