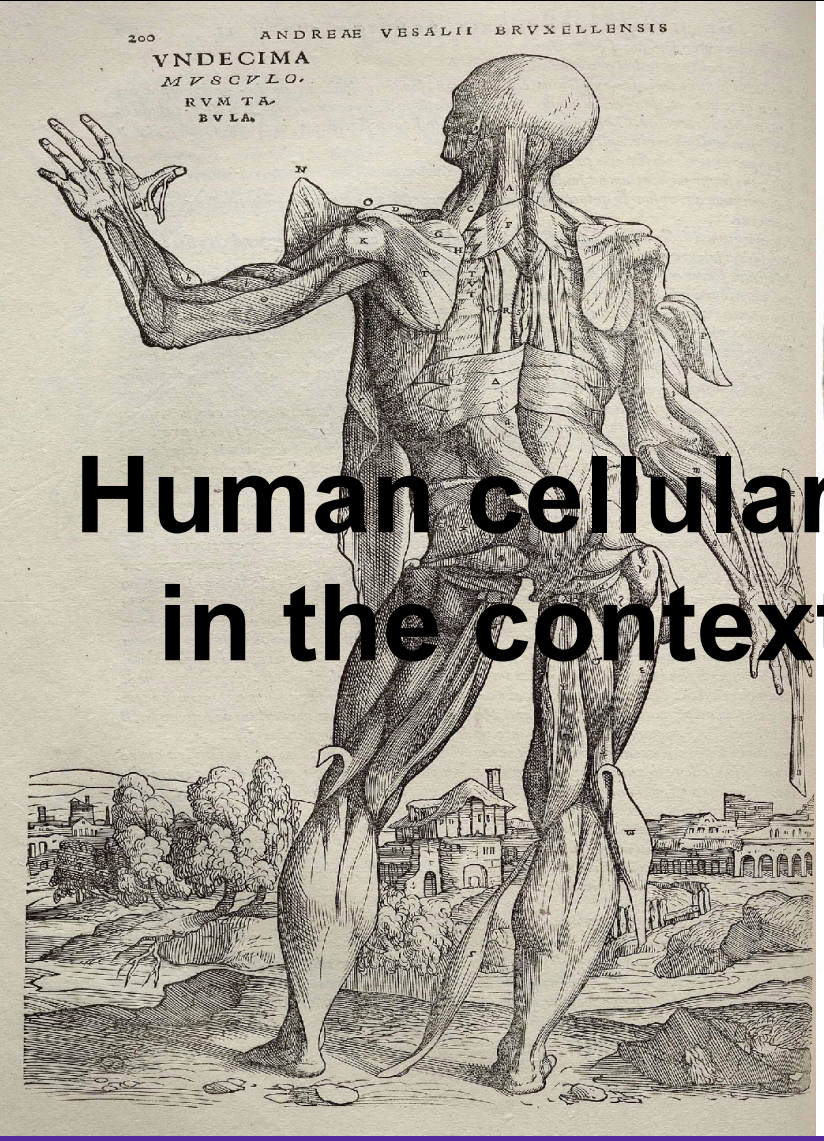
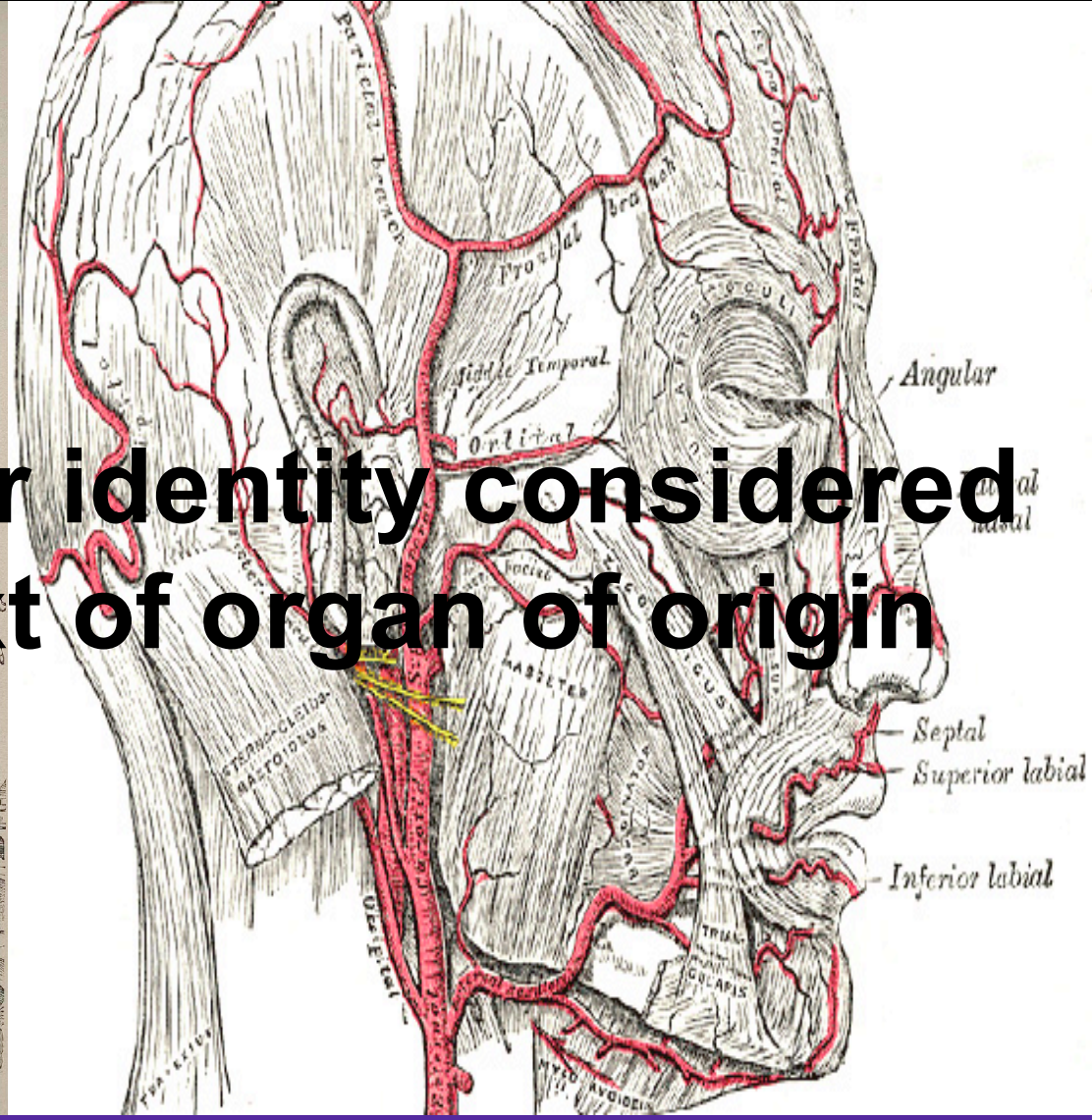




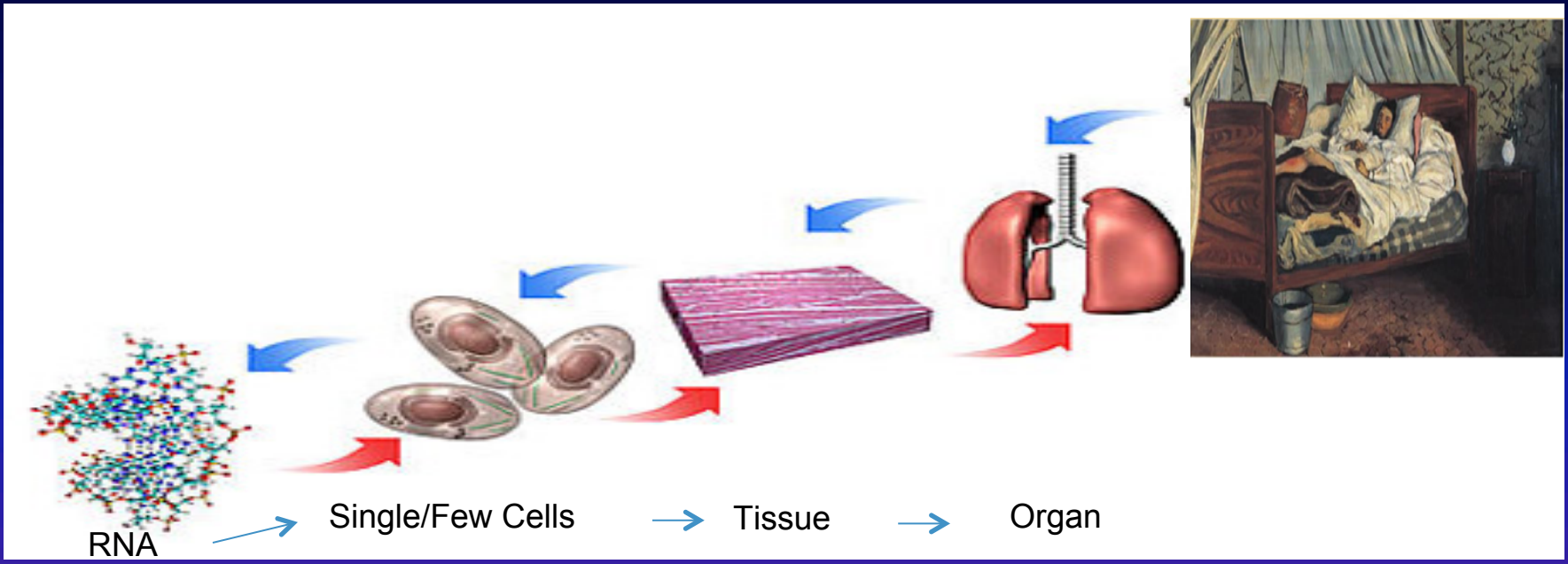
Human cellular identity considered in the context of organ of origin

June 8th, 2016

ENCODE Outreach Meeting

Stanford University, Stanford CA

A Recurrent Challenge of Precision Medicine: Biological Heterogeneity Of Samples



How accurate is transcriptional profiling tissue or organs in order to
Define "Normal"- Diagnose- Monitor Progression-Evaluate Rx Efficacy-Prognosis

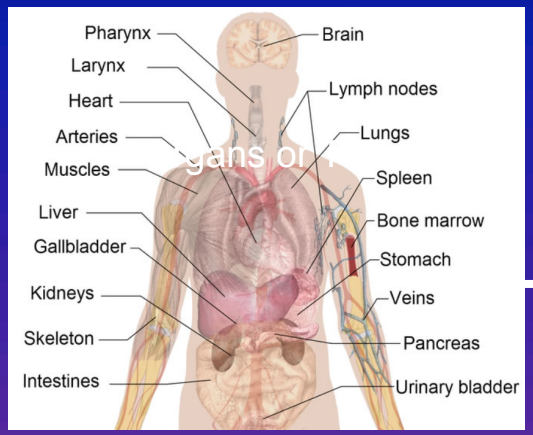
Will the analysis of enriched constituent primary cells or single cells be better ?

What do transcriptional landscapes of constituent primary cells of organs reveal?

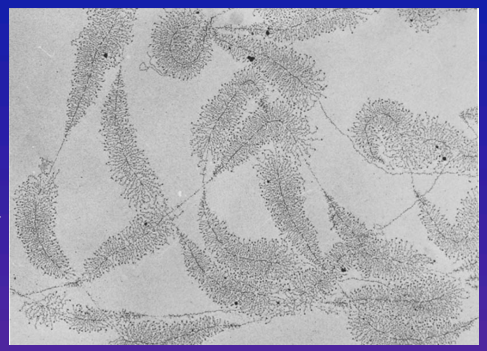
Transcription Profiles Used To Describe the Phenotypic State of Complex Tissues/Organs

(~50% of the 5K 2015 papers)

Sample Source RNA Types Tx Profiles Intrepret

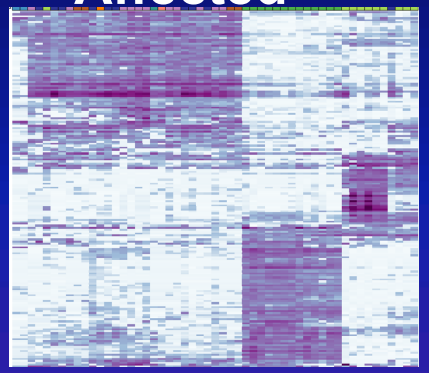


Human Organs/Tissues

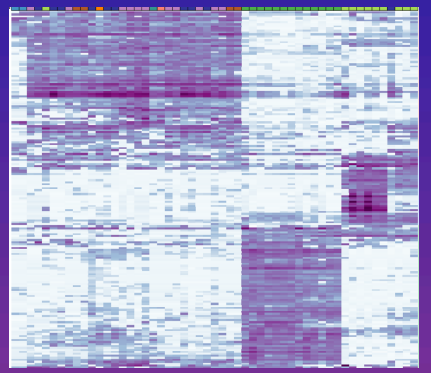


RNA transcripts
Miller and Beatty 1969

Affected



Normal



Differential Expression Clustering

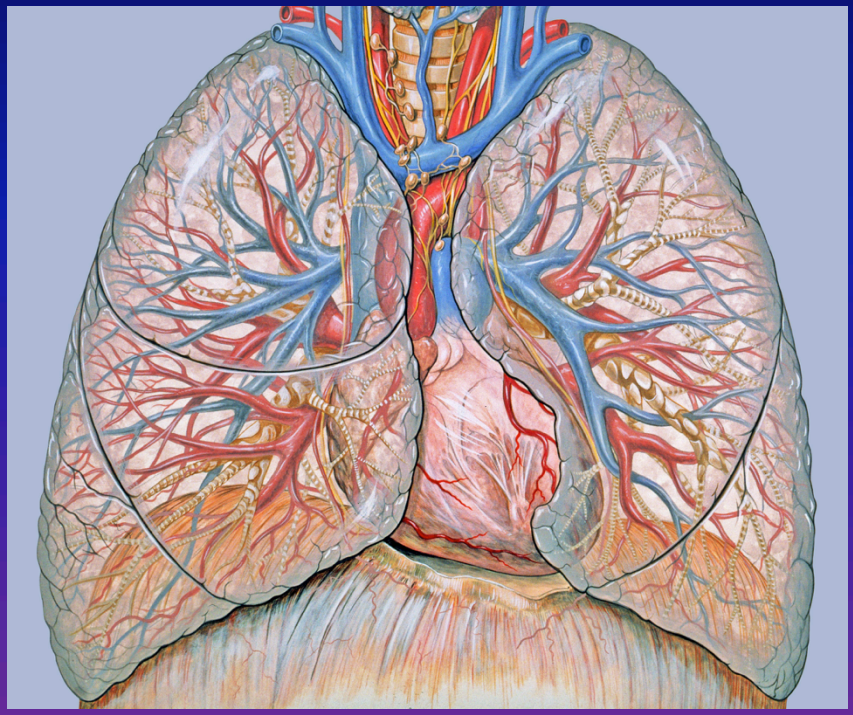
Good

Bad

Complexity of an Organ

Example: Lung/respiratory Tissues vs. Constituent Primary Cell Lines (10)

HPAEpiC - Alveolar Epithelial Cells
HPSAEpiC - Small Airways Epithelial Cells
HPF - Pulmonary Fibroblasts
HPMEC - Pulmonary Microvascular Endothelial Cells
HBEpC - Bronchial Epithelial Cells
HBSMC - Bronchial Smooth Muscle Cells
HBF - Bronchial Fibroblasts
HTEpC - Tracheal Epithelial Cells
HTSMC - Tracheal Smooth Muscle Cells
HNEpC - Nasal Epithelial Cells

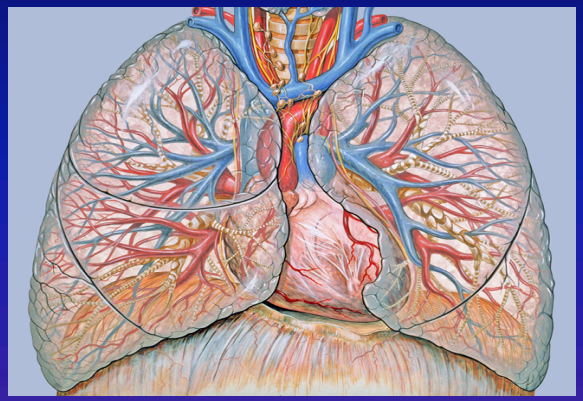


https://commons.wikimedia.org/wiki/File:Thoracic_anatomy.jpg

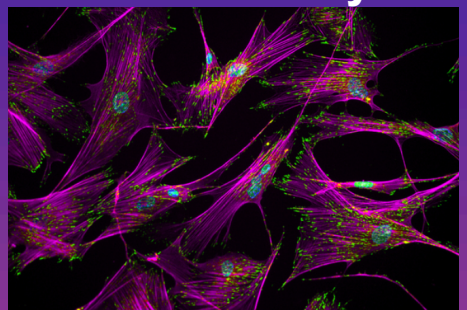
What Is Learned from Looking At The Transcription Profiles of Constitutent Primary Cells of Complex Tissues/Organs

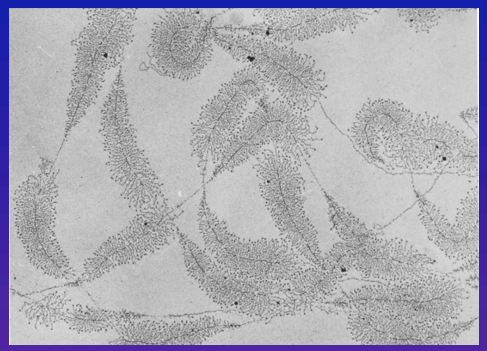
Sample Source RNA Types Tx Profiles Intrepret

Lung

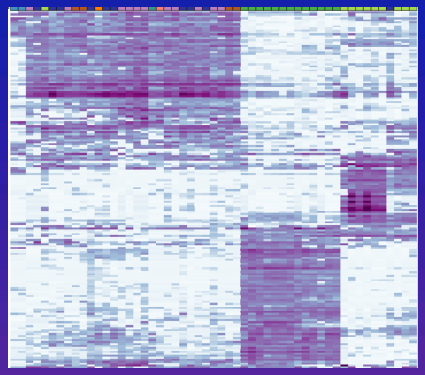


Constituent Primary Cells





RNA transcripts
Miller and Beatty 1969



Differential Expression
Clustering

Good

Bad

Long Range Focus:

- Considering the large number of cell types composing human tissue/organs what constitutes baseline “normal” Tx profile for the multicomponent tissue/organ sample ?

- If single cells or low number of primary cells are obtained from organ/tissue, will this allow for identification of “normal” for the organ/tissue and help more precisely identify the cell types involved in the affected condition?

Primary Cells

53 primary cell pellets or total RNAs from 10 organ/tissue systems

Cells were screened for cell-specific biomarkers

2 bio-replicates each

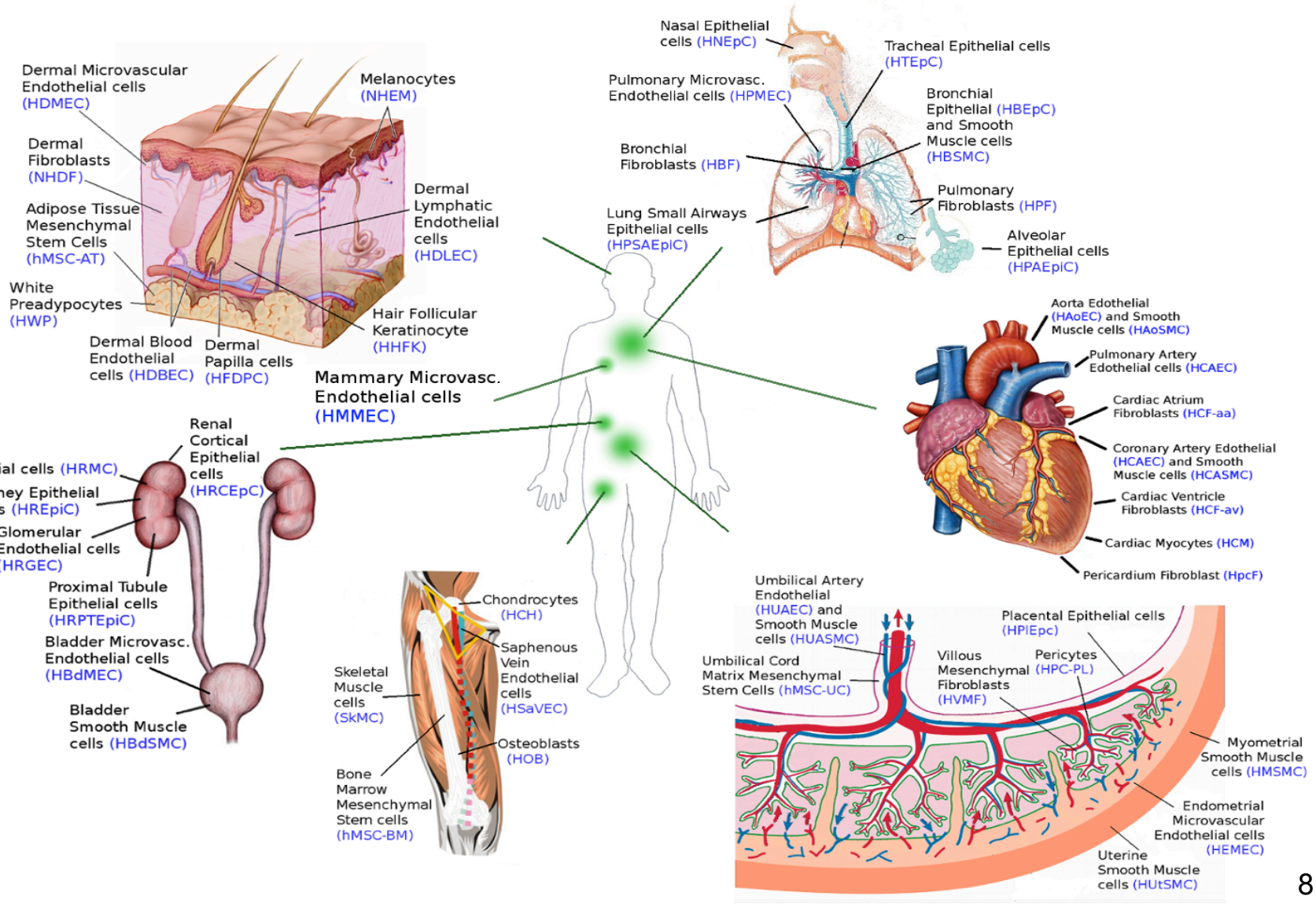
Ribo-depleted stranded RNA-seq
-> 279,676,766 average mapped reads/rep pair

Cell type	2x
Cardiomyocytes	1
Endothelial	13
Epithelial	10
Fibroblast	11
Melanocyte	2
Preadipocyte	1
Skeletal muscle	1
Smooth muscle	10
Mesenchymal Stem	4

Germ layer	2x
Ectoderm	6
Endoderm	4
Mesoderm	43

Body location	2x
Adipose	2
Bladder	2
Bone	4
Breast	1
Heart	11
Kidney	5
Lung/Breathing	10
Muscle	1
Plac/Umb/Uter	9
Skin	8

Issue of Location in Tissue of Isolated Primary Cells

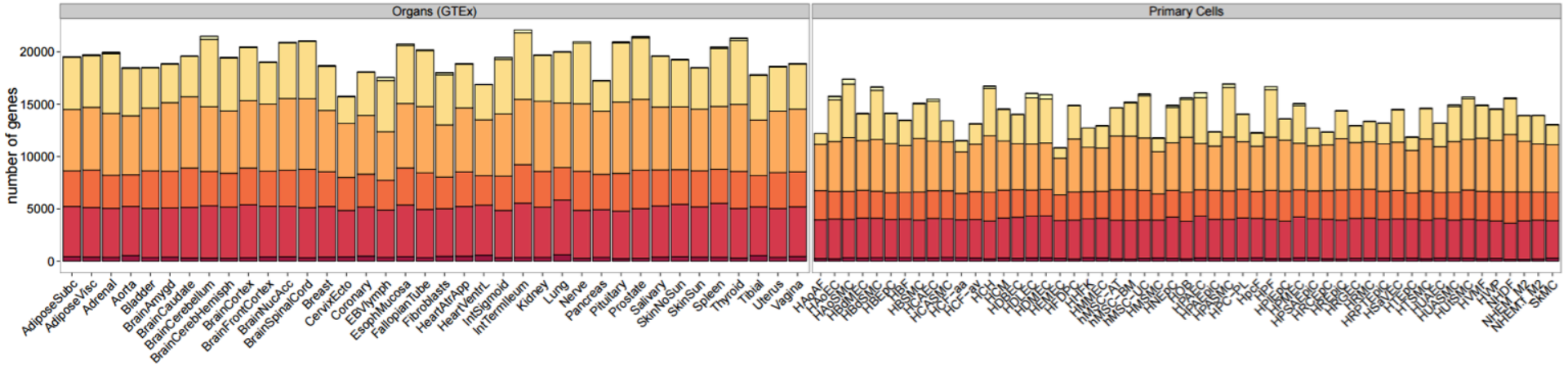


Not Surprisingly Individual Primary Cells Have a Less Complex Transcriptome Than Organs

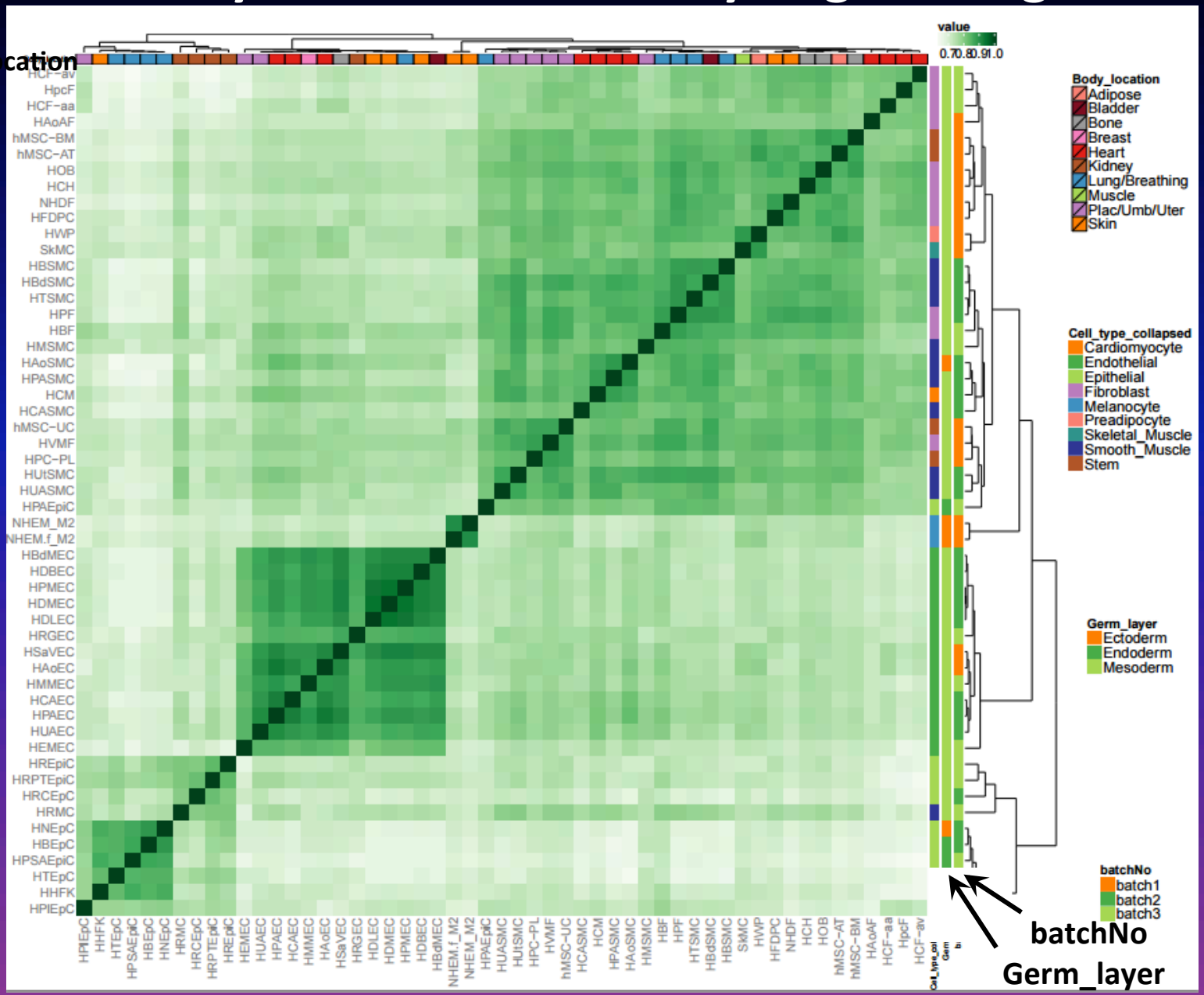
Organs express more annotated genes than primary cells

Organs (GTEx)

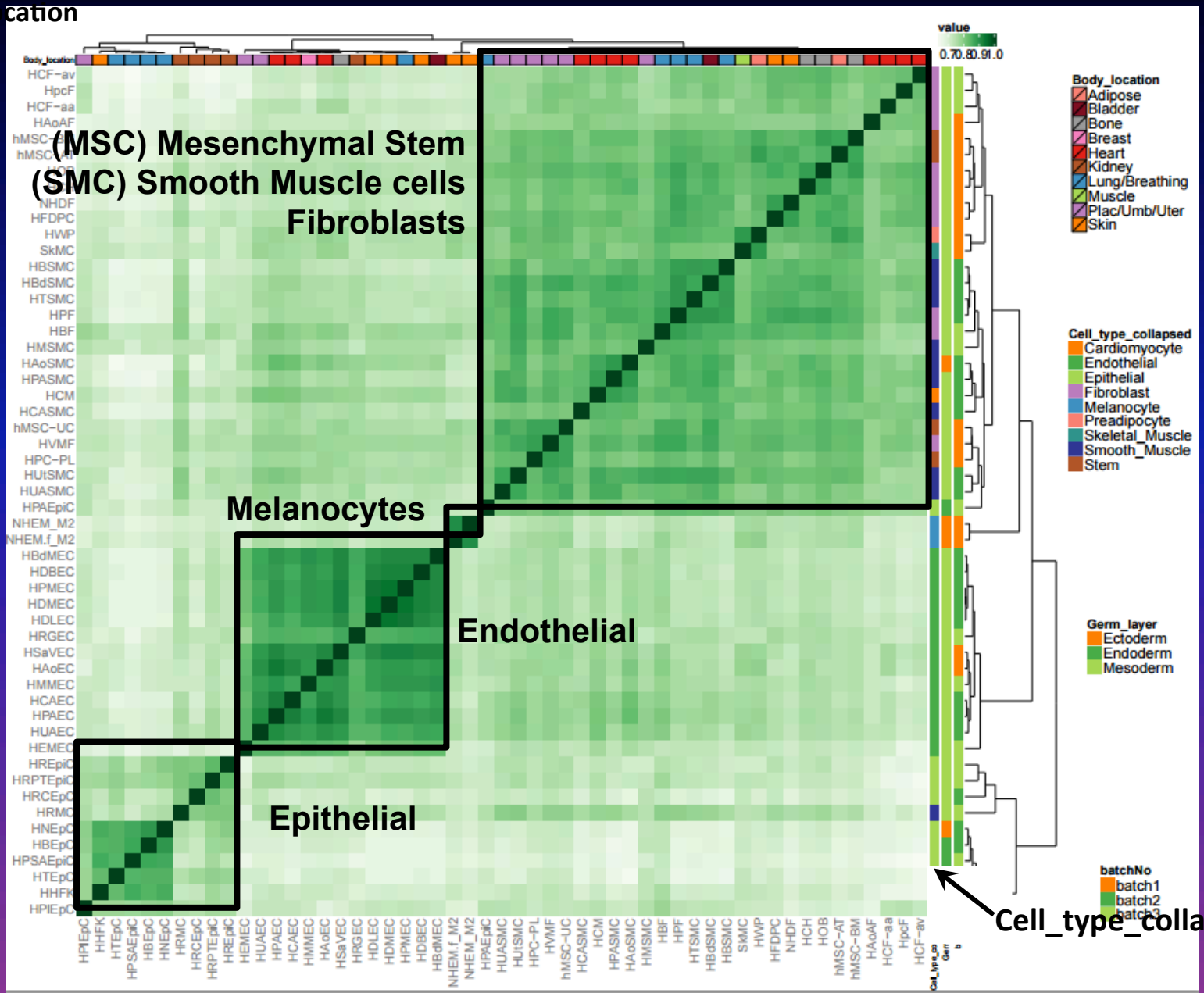
Primary Cells



Clustering of Primary Cells Does Not Reflect Body Location Nor Embryological Origin

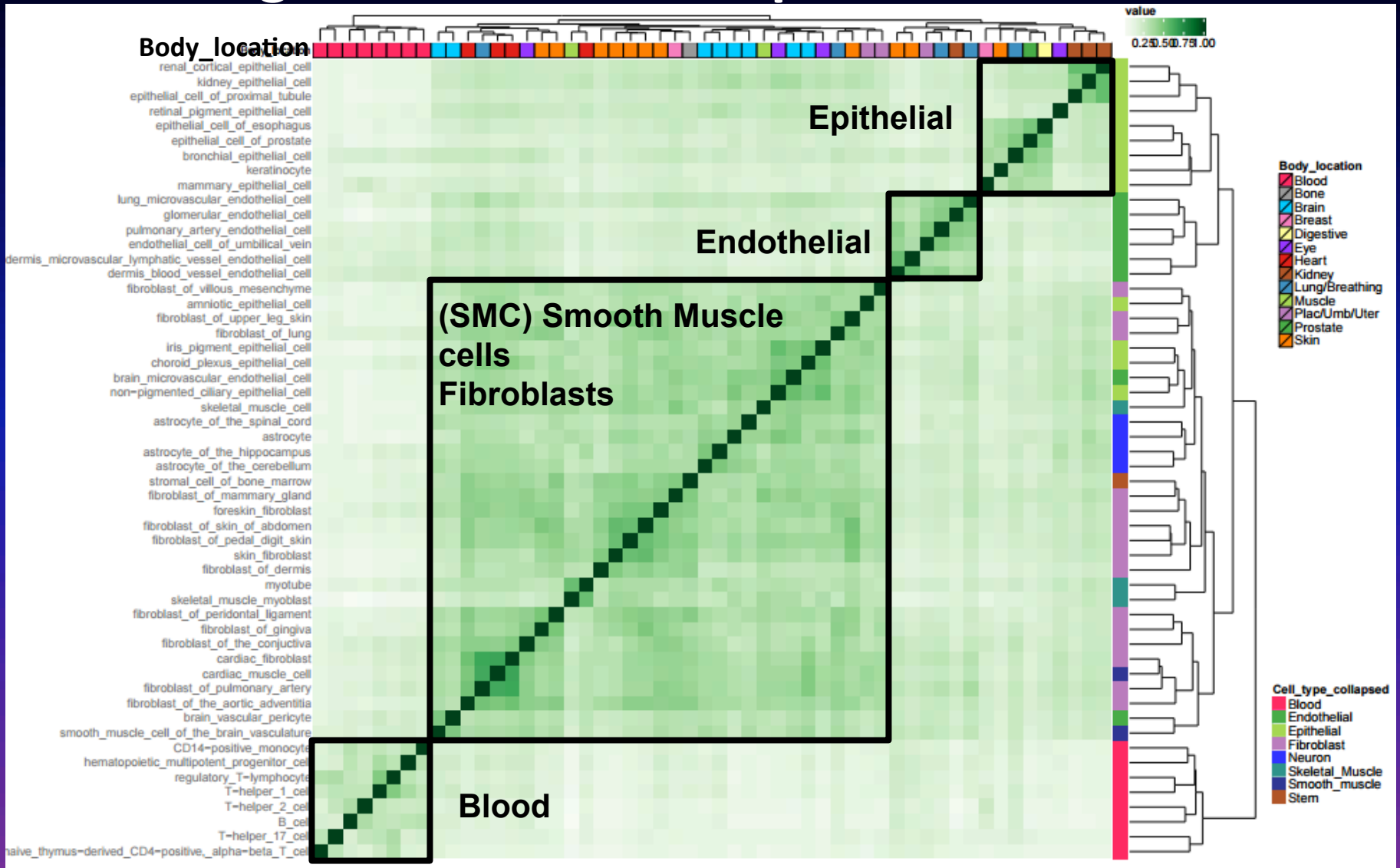


Clustering of DE of Primary Cells Is Based on Cell Type Not Organ/tissue or Embryological Origin



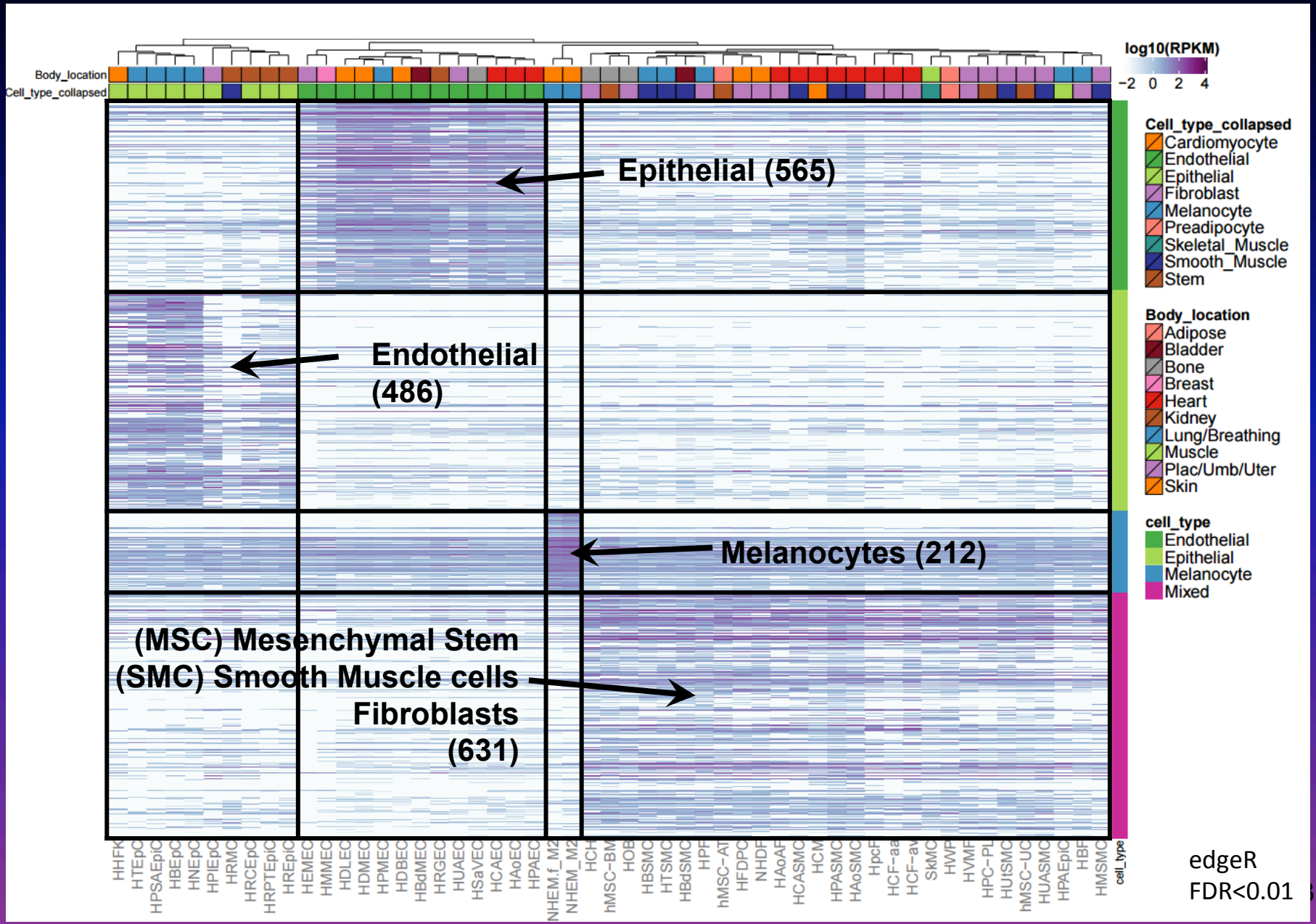
Pearson

Clustering Based on DNAse Peaks Supports the Clustering Based on Gene Expression



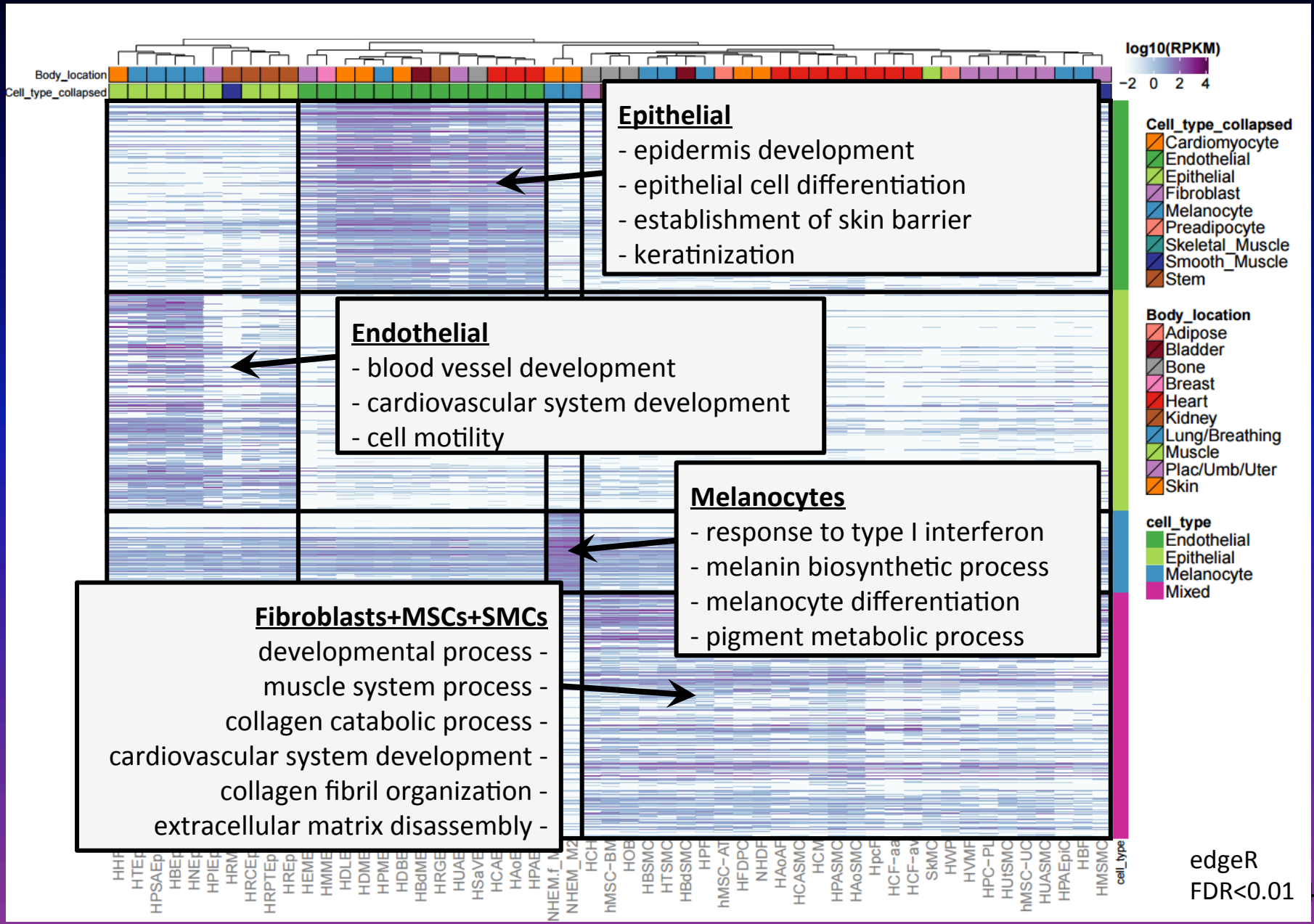
Sheffield, Nathan C., et al. "Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions."*Genome research* 23.5 (2013): 777-788.

2,873 DE Genes Specific of Each Cell Type Cluster

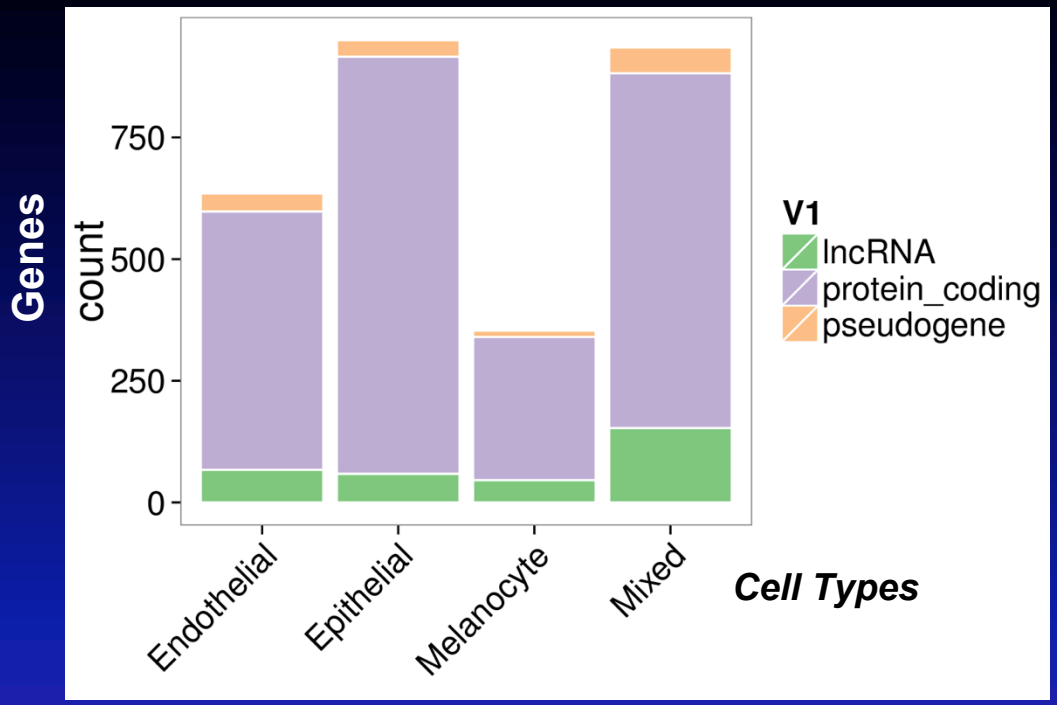


edgeR
FDR<0.01

2,873 DE Genes Specific of Each Cell Type Cluster

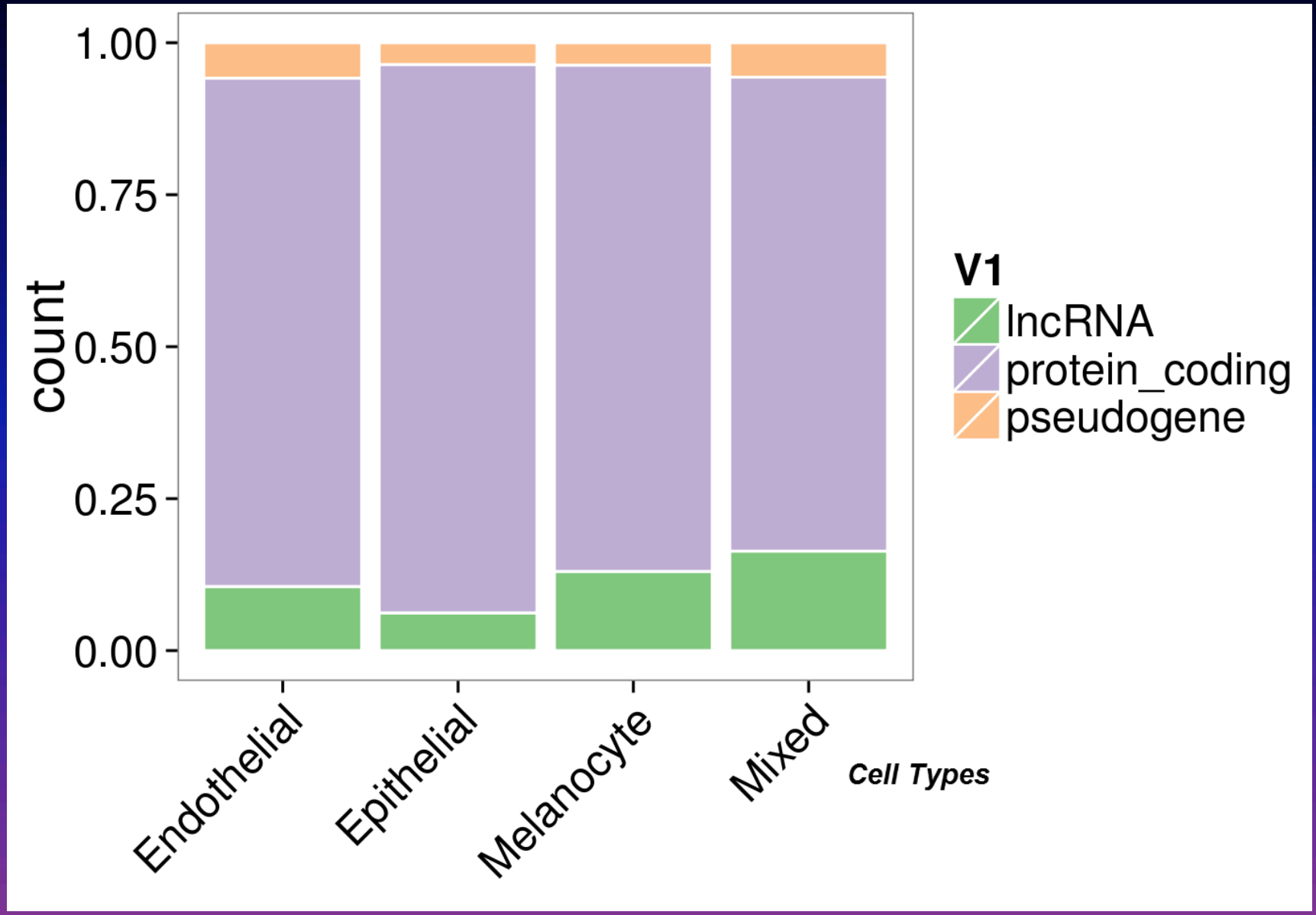


RNA Biotypes Summary (Numbers)

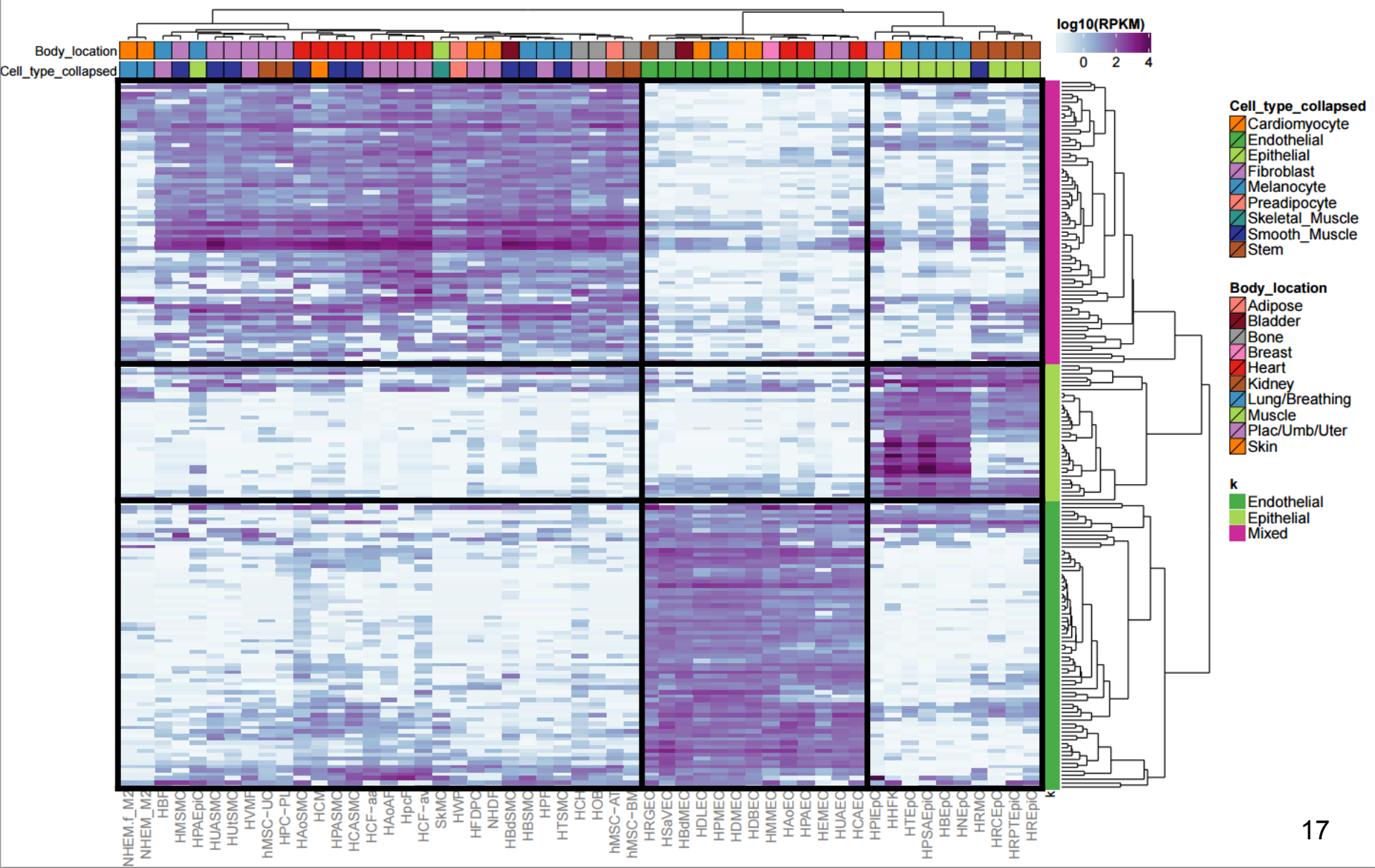


	Endothelial	Epithelial	Melanocyte	Mixed	Total
lncRNA	67	59	46	153	325
protein_coding	531	857	294	729	2411
pseudogene	37	34	13	53	137
Total	635	950	353	935	2873

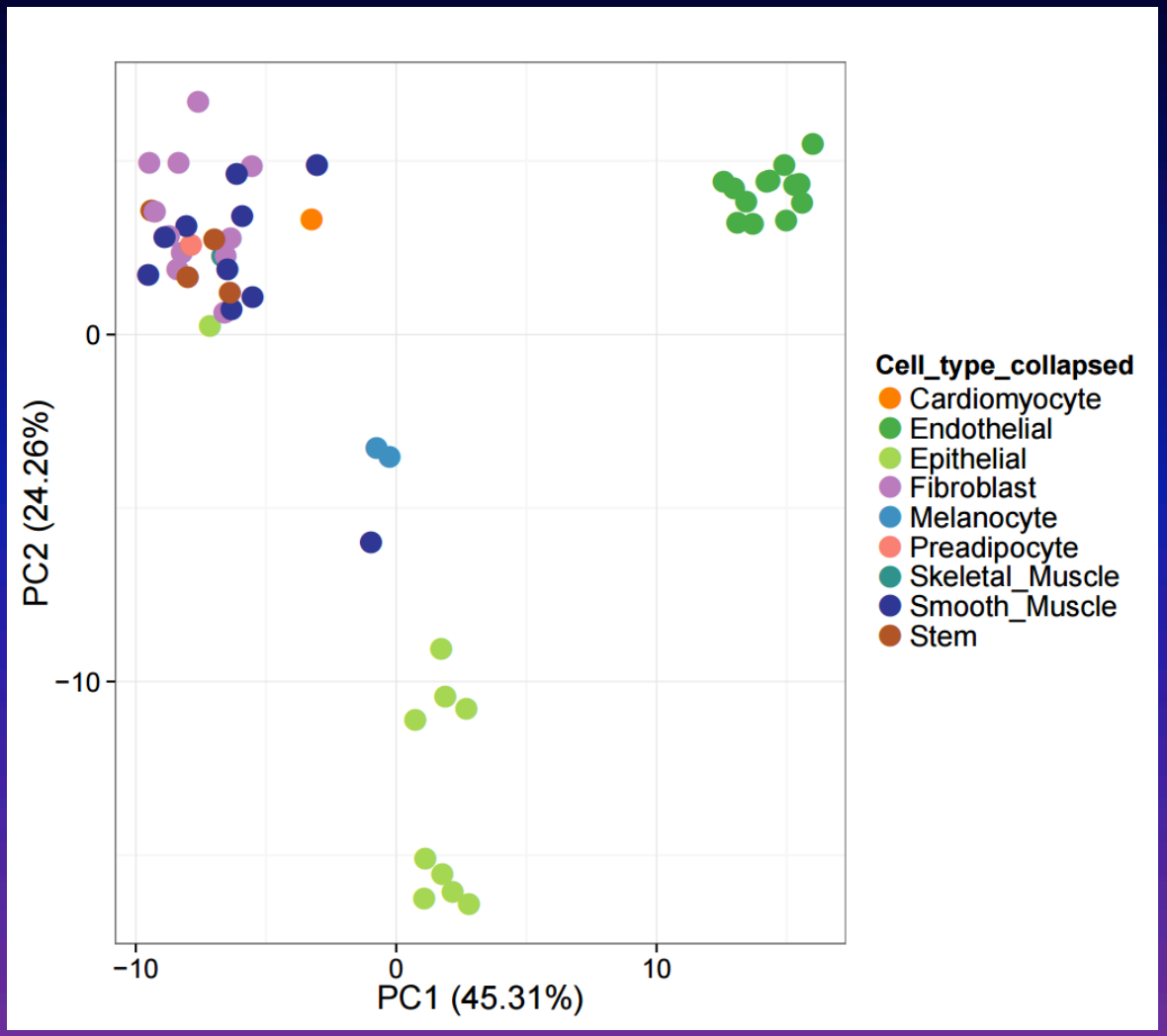
RNA Biotype Percent



Expression Profiles of the Core/Driver the180 Genes



Identifying Core Genes That Drive Clustering



180 genes:

- 177 protein coding
- 3 lncRNAs

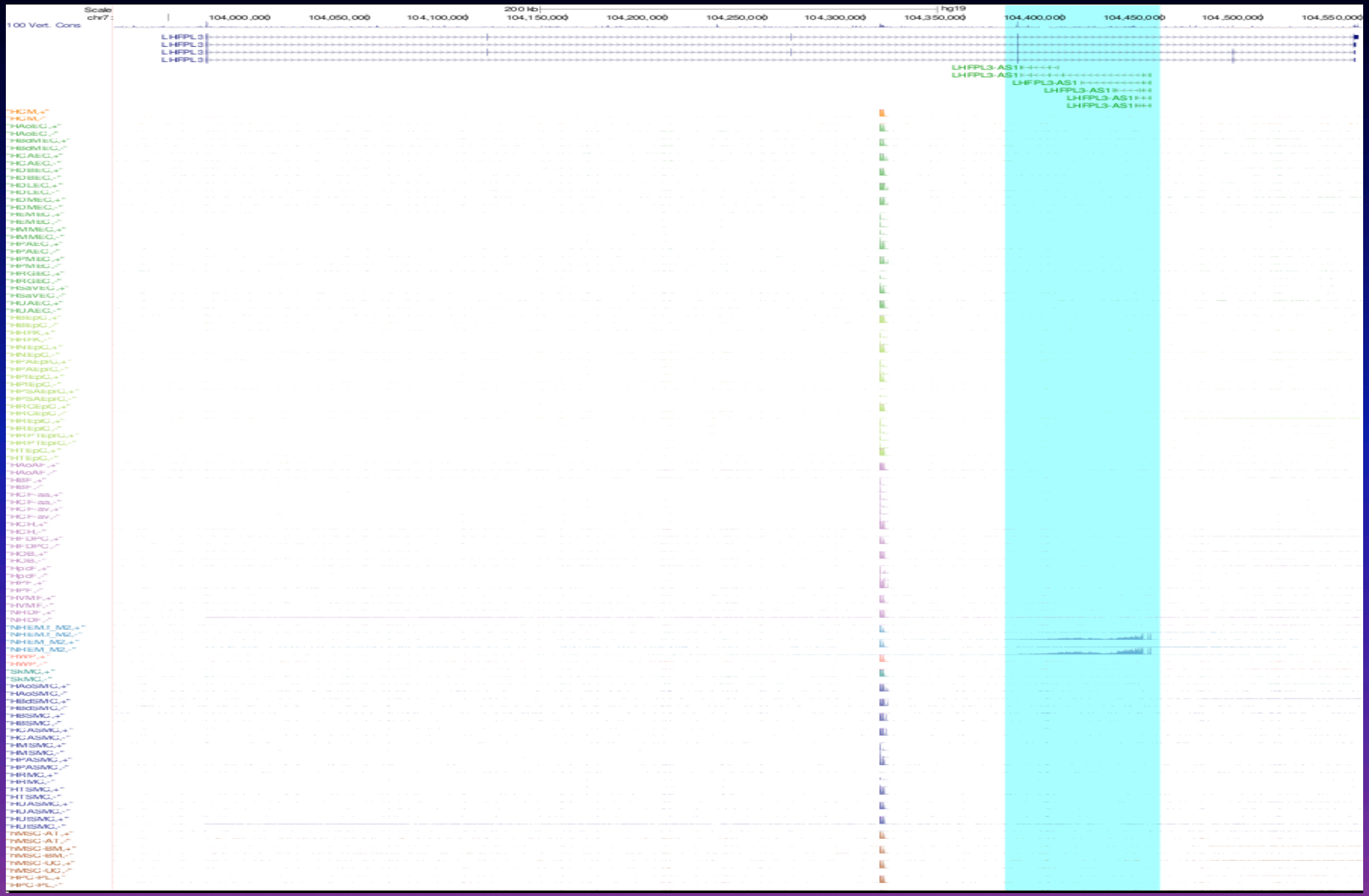
2873 genes:

- **325 lncRNA**
- **137 pseudogenes**

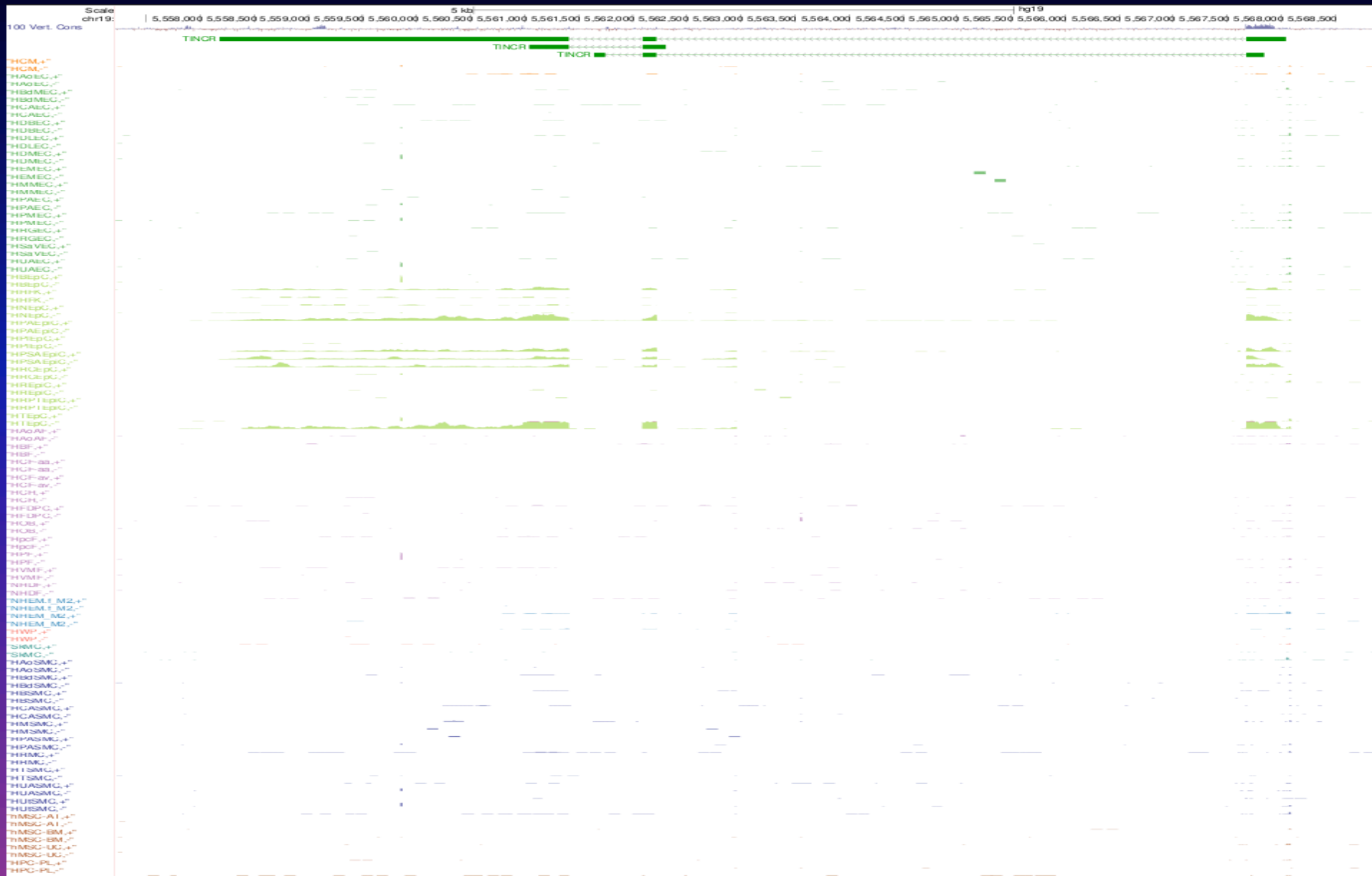
Fontes, Magnus, and Charlotte Soneson. "The projection score-an evaluation criterion for variable subset selection in PCA visualization." *BMC bioinformatics* 12.1 (2011): 307.

Endothelial-specific long noncoding RNA (LINC01235)

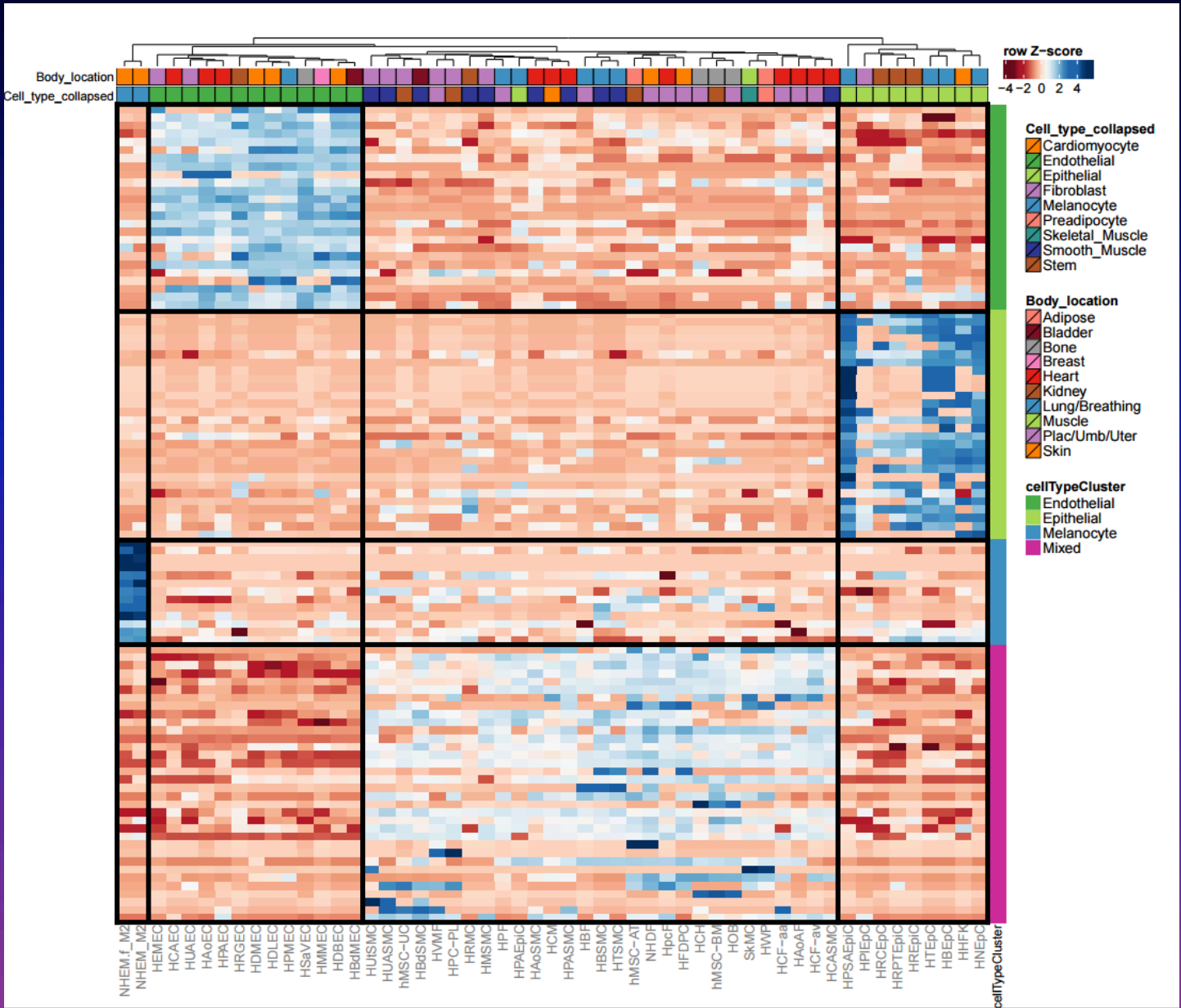
Melanocyte specific non-coding *antisense* RNA (LHPPL3)



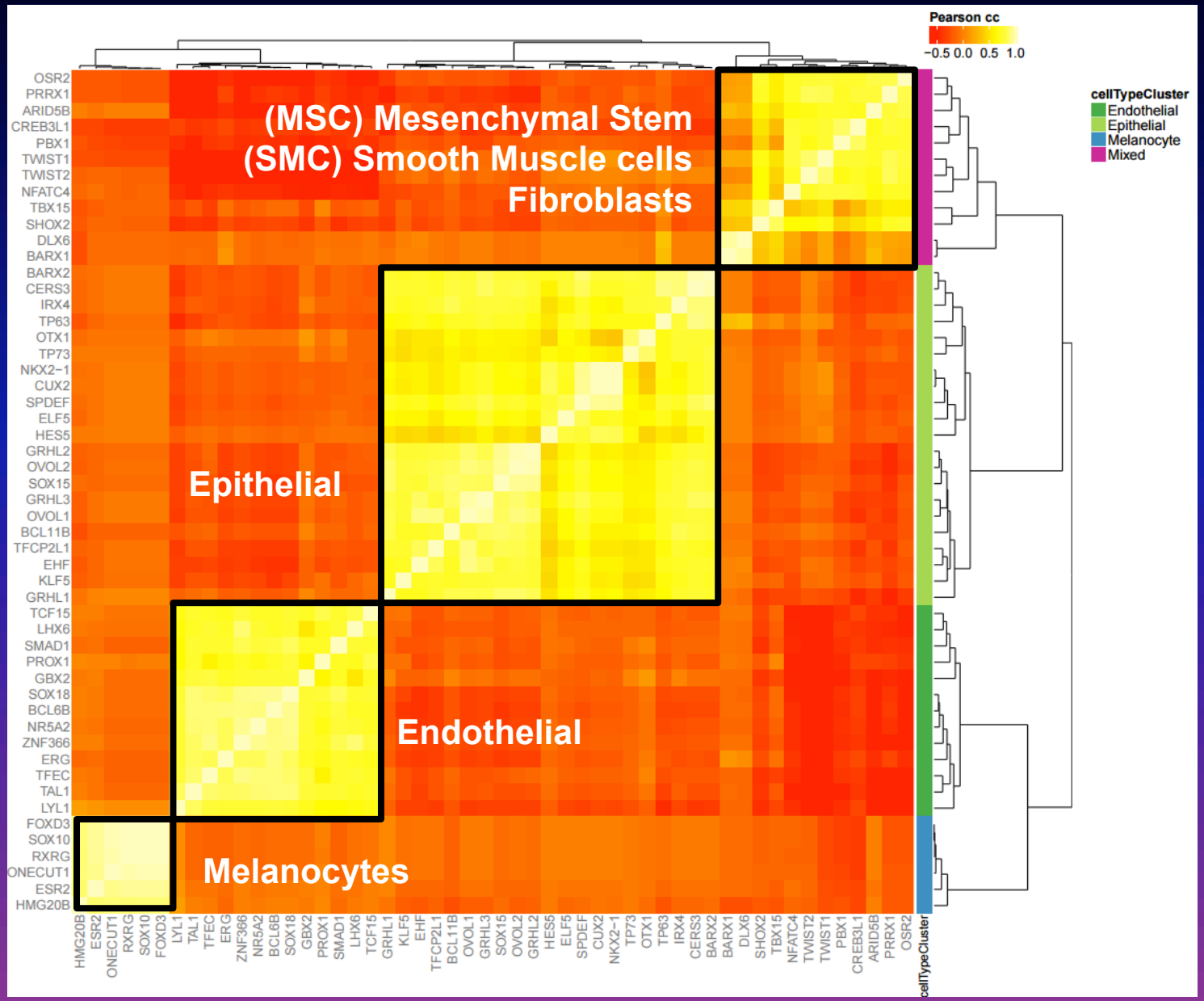
Epithelial Specific Long Non-coding RNA (TINCR)



Among 2,873 DE Genes There Are 100 Cell Type Cluster-Specific Transcription Factors (TFs)



A Subset of TFs Expression Is Highly Correlated and Underlines the Main Cell Type Clusters



Conclusions

1)The studied primary cells can be clustered into four major groups based on cell types: a) endothelial, b) epithelial, c) melanocytes and d) fibroblasts + SMCs + MSCs

2)The cell type clustering supercedes effect of body location and embryological origin (no batch effect)

3)There are ~2,000 genes specific to each cell type cluster. Approximately 180 genes are enough to separate the cell type clusters

4) DHS profiles mirror gene expression clustering

5) ~25- 50 transcription factors are distributed among each cluster and their clusters mimic gene expression of the cell type clustering

6) The correlaton of primary cell transcriptome with whole tissues/organ is poor.

7) Either the are missing primary cells that compose tissues (likely) and/or reconstruction of tissue profiles need complex integration of data from parts of primary cell profiles

Acknowledgments

Guigó lab

Roderic Guigó
Sarah Djebali
Anna Vlasova
Dmitri Pervouchine
Julien Lagarde
Barbara Uscynzka

Gingeras lab

Carrie Davis
Alex Dobin
Chris Zaleski
Alex Scavelli
Jorg Drenkow
Lei-Hoon See

Mortazavi lab

Ali Mortazavi