

Identifying the origin of cancer at cell-type resolution by modeling the relationship between scATAC genome accessibility and the tumor mutational landscape

Mohamad Daniel Bairakdar (RA in Tsankov Lab)

Department of Genetics and Genomic Sciences

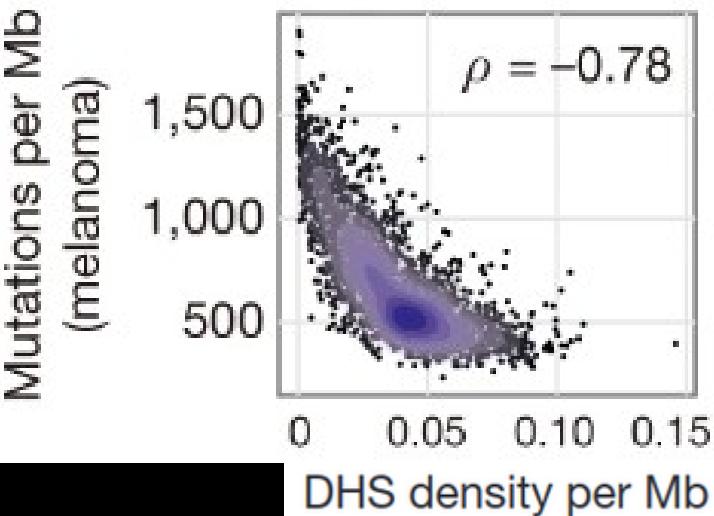


Icahn
School of
Medicine at
**Mount
Sinai**

Background

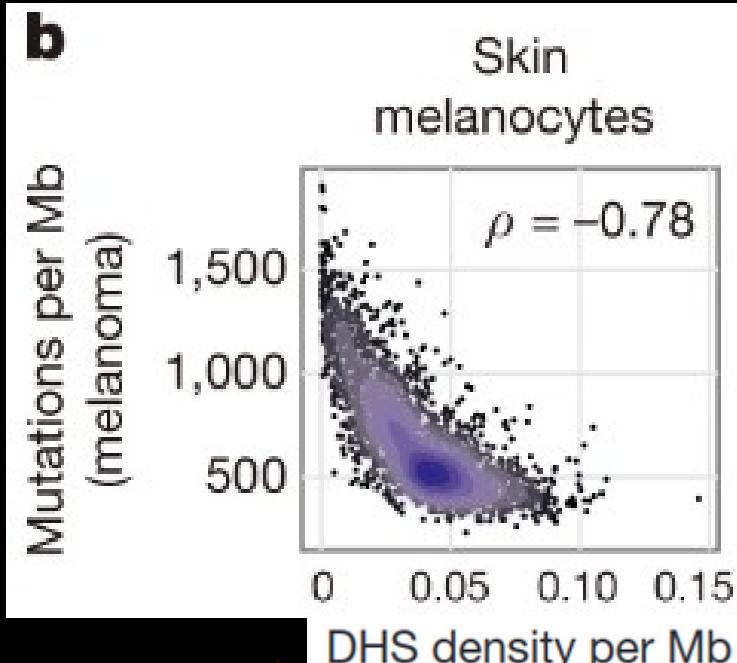
- Our team has previously shown we can predict the cell of origin (COO) of cancer by modeling the relationship between
 - the epigenetics of originating **NORMAL** tissue

- Our team has previously shown we can predict the cell of origin (COO) of cancer by modeling the relationship between
 - the epigenetics of originating NORMAL tissue
 - and the tumor mutational landscape

bSkin
melanocytes

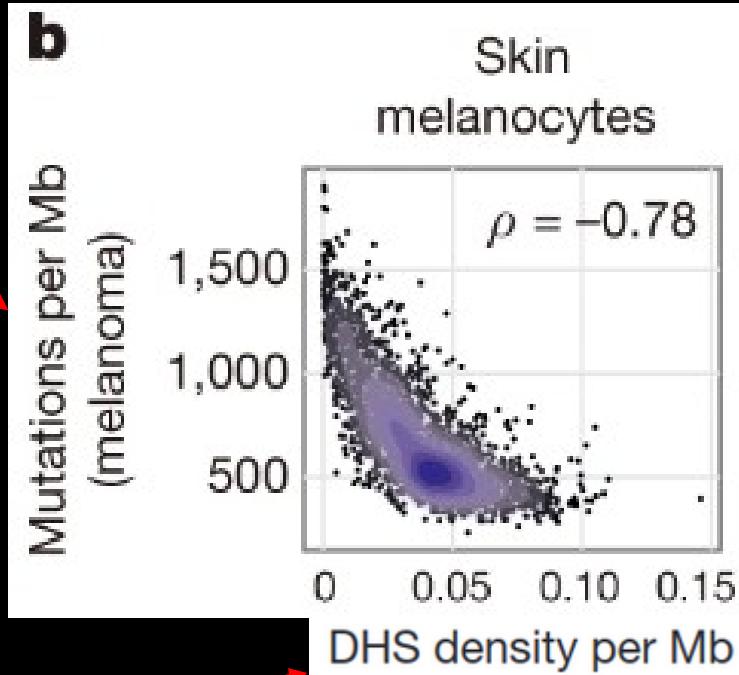
Paz Polak, Rosa Karlic
et al. Nature 2015.

DNase I hypersensitive sites (DHS) sequencing,
Measures chromatin accessibility (here, in
normal skin melanocytes)



Paz Polak, Rosa Karlic
et al. Nature 2015.

CANCER TISSUE
(Melanoma)



NORMAL TISSUE
(Skin Melanocyte)

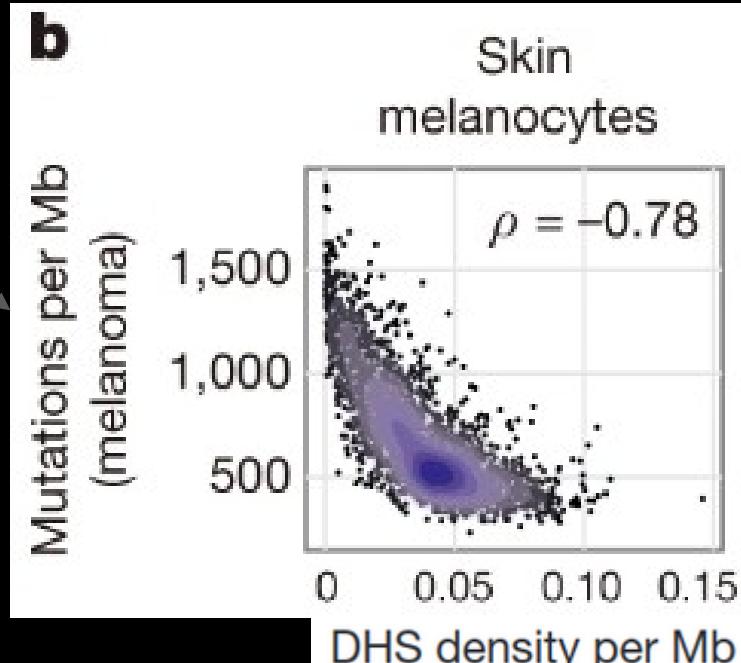


Paz Polak, Rosa Karlic
et al. Nature 2015.

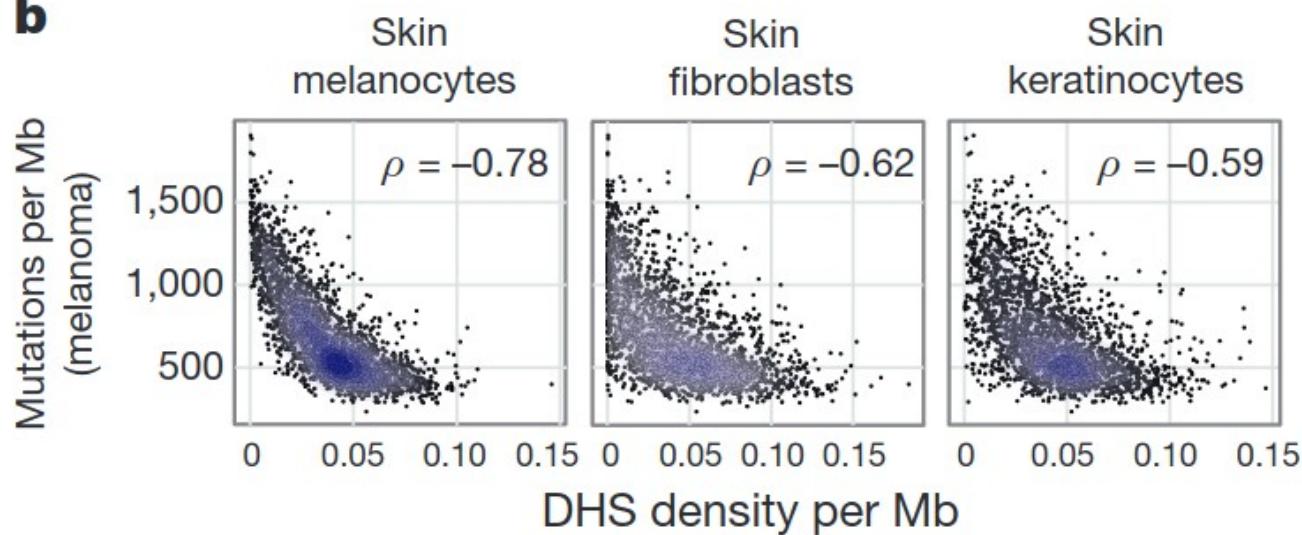
CANCER TISSUE
(Melanoma)

UNMATCHED

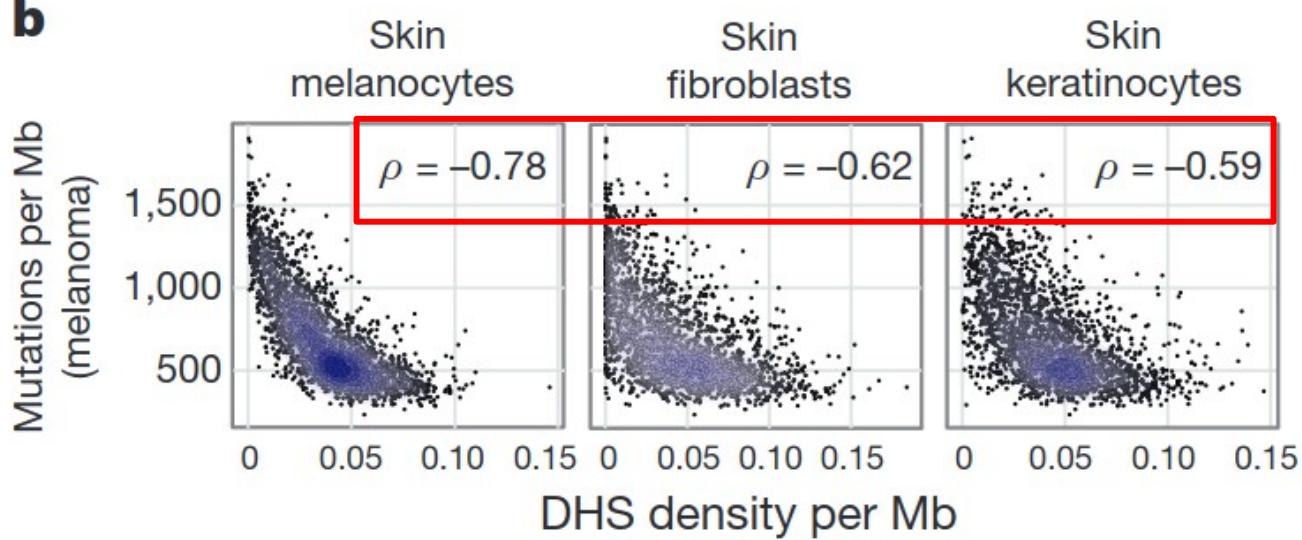
NORMAL TISSUE
(Skin Melanocyte)



Paz Polak, Rosa Karlic
et al. Nature 2015.

b

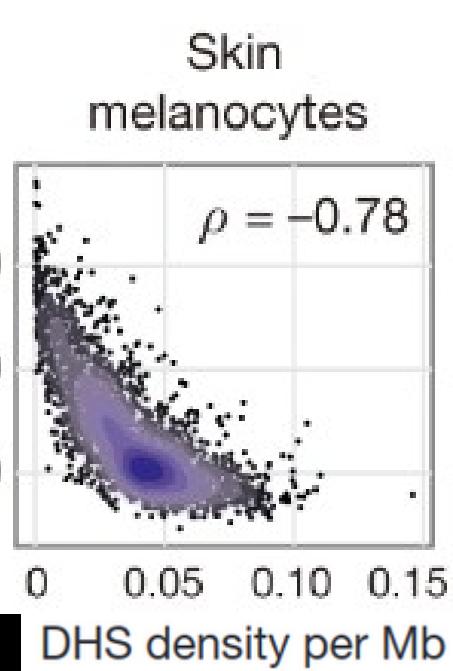
Paz Polak, Rosa Karlic
et al. Nature 2015.

b

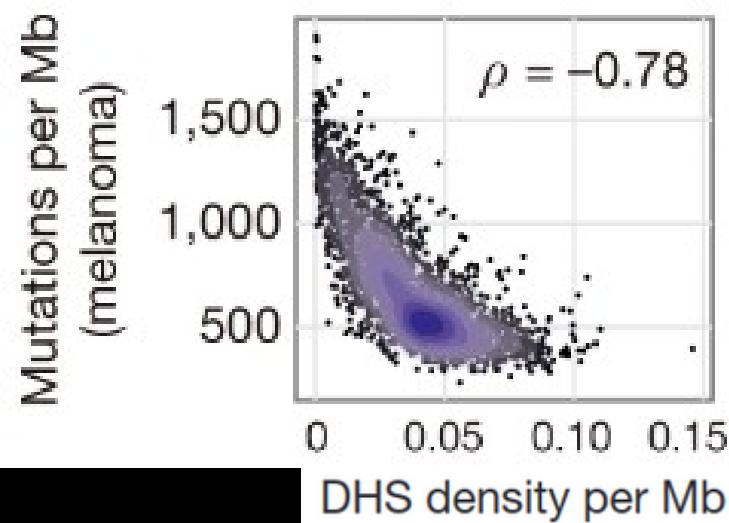
Paz Polak, Rosa Karlic
et al. Nature 2015.

b

Mutations per Mb
(melanoma)

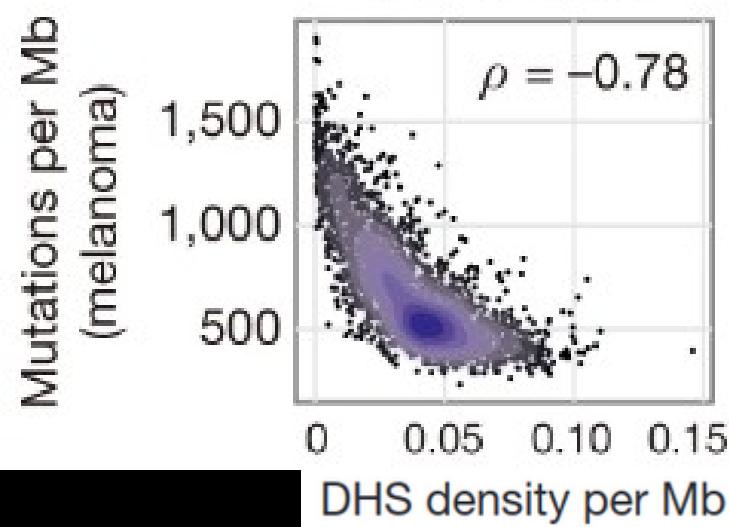


2 questions:

b

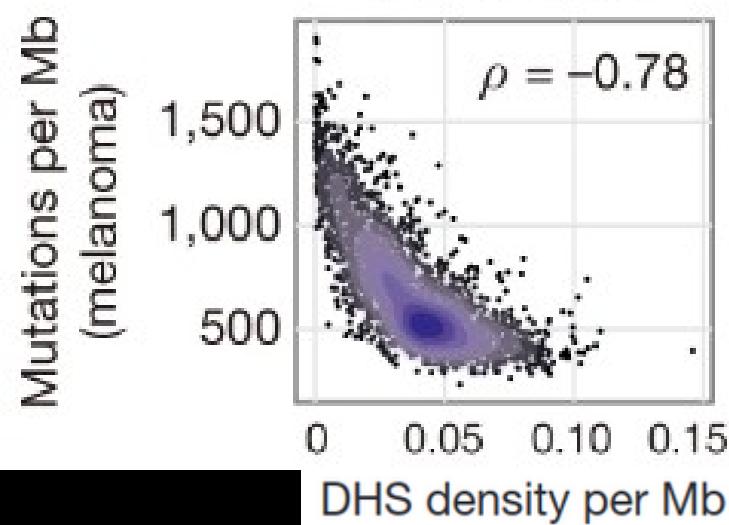
2 questions:

1. Mutational density correlation with NORMAL cell epigenetic state?

b

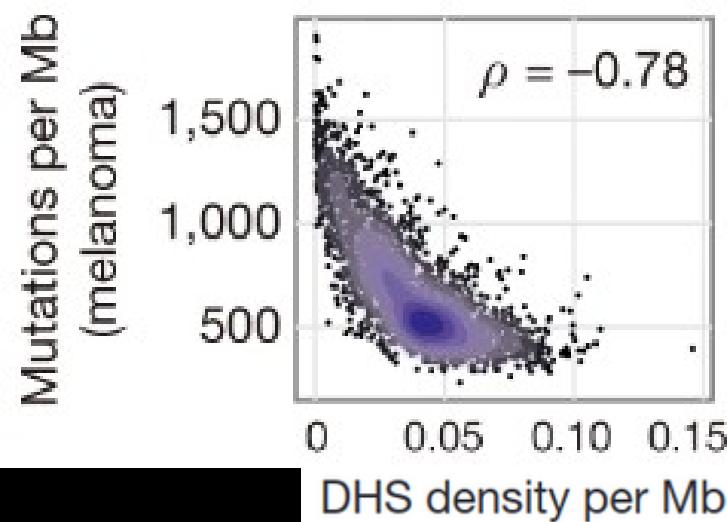
2 questions:

1. Mutational density correlation with NORMAL cell epigenetic state?
 - Mutation landscape = mostly passenger mutations

b

2 questions:

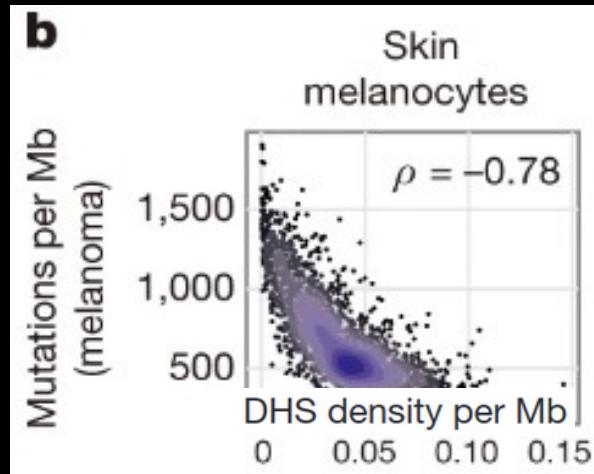
1. Mutational density correlation with NORMAL cell epigenetic state?
 - Mutation landscape = mostly passenger mutations
2. Why anticorrelation? Mechanistic explanation?

b

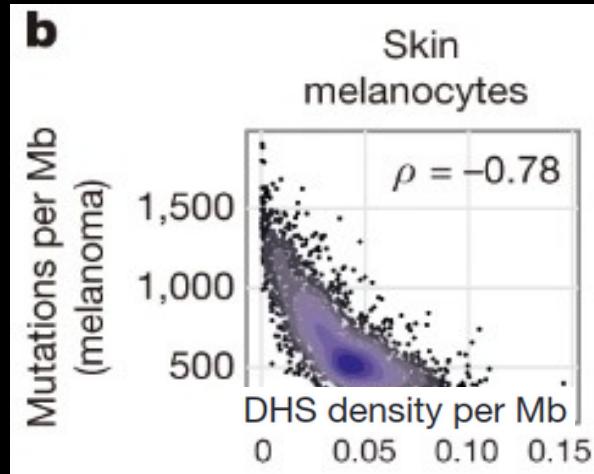
2 questions:

1. Mutational density correlation with NORMAL cell epigenetic state?
 - Mutation landscape = mostly passenger mutations
2. Why anticorrelation? Mechanistic explanation?
 - e.g DNA Repair mechanisms

- How to calculate correlation?



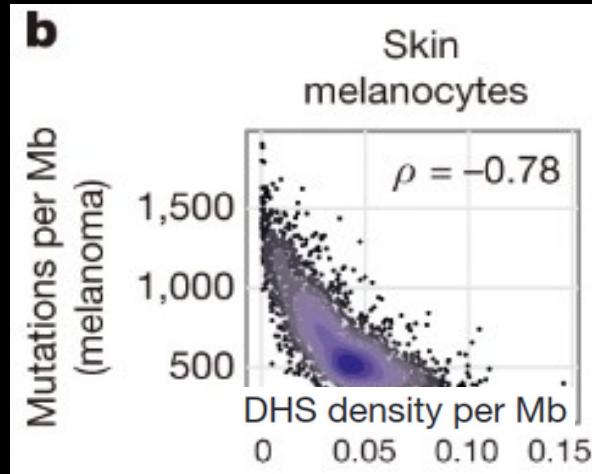
- How to calculate correlation?



chr1: 4010001-5010000
chr1: 5010001-6010000
chr1: 6010001-7010000

chr22: 47000001-48000000
chr22: 48000001-49000000
chr22: 49000001-50000000

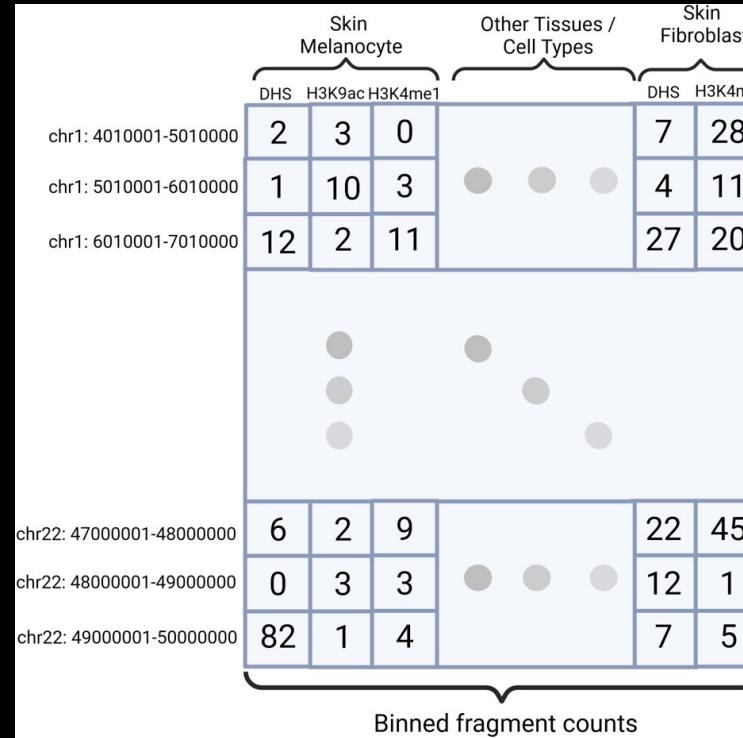
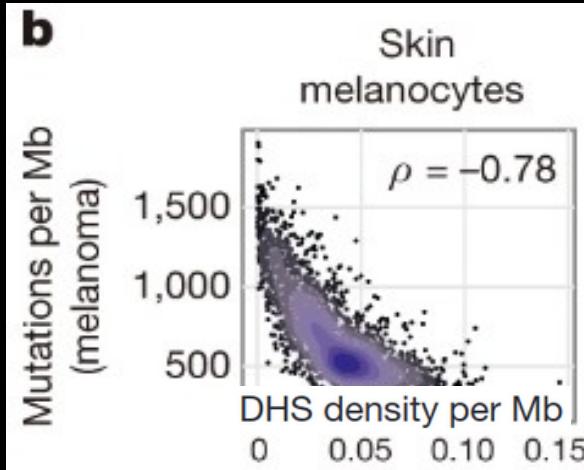
- How to calculate correlation?



		Skin Melanocyte	
		DHS	H3K9ac H3K4me1
chr1: 4010001-5010000		2	3 0
chr1: 5010001-6010000		1	10 3
chr1: 6010001-7010000		12	2 11
chr22: 47000001-48000000		6	2 9
chr22: 48000001-49000000		0	3 3
chr22: 49000001-50000000		82	1 4

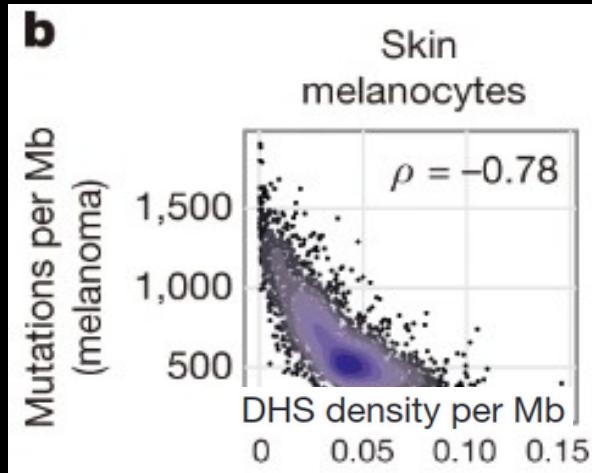
Three grey circles are plotted below the last three rows of the table.

- How to calculate correlation?



- How to calculate correlation?

Binned Epigenetic data matrix



Genomic Regions:

- chr1: 4010001-5010000
- chr1: 5010001-6010000
- chr1: 6010001-7010000
- chr22: 47000001-48000000
- chr22: 48000001-49000000
- chr22: 49000001-50000000

Cell Types / Tissue Groups:

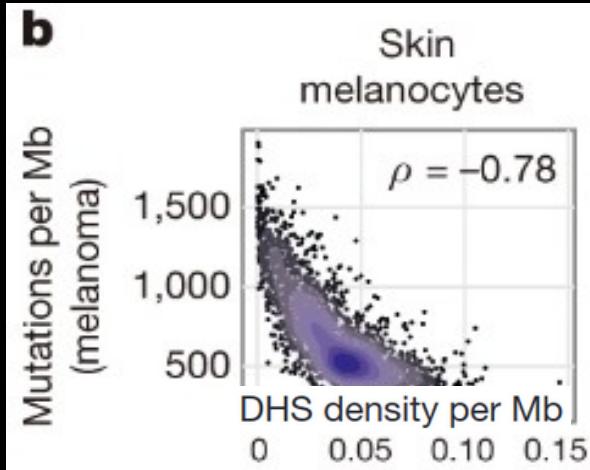
- Skin Melanocyte
- Other Tissues / Cell Types
- Skin Fibroblast

Binned fragment counts (approximate values from heatmap):

Region	DHS	H3K9ac	H3K4me1	Other Tissues / Cell Types	DHS	H3K4me1
chr1: 4010001-5010000	2	3	0		7	28
chr1: 5010001-6010000	1	10	3	● ● ●	4	11
chr1: 6010001-7010000	12	2	11		27	20
chr22: 47000001-48000000	6	2	9	● ● ●	22	45
chr22: 48000001-49000000	0	3	3	● ● ●	12	1
chr22: 49000001-50000000	82	1	4		7	5

Binned fragment counts

- How to calculate correlation?



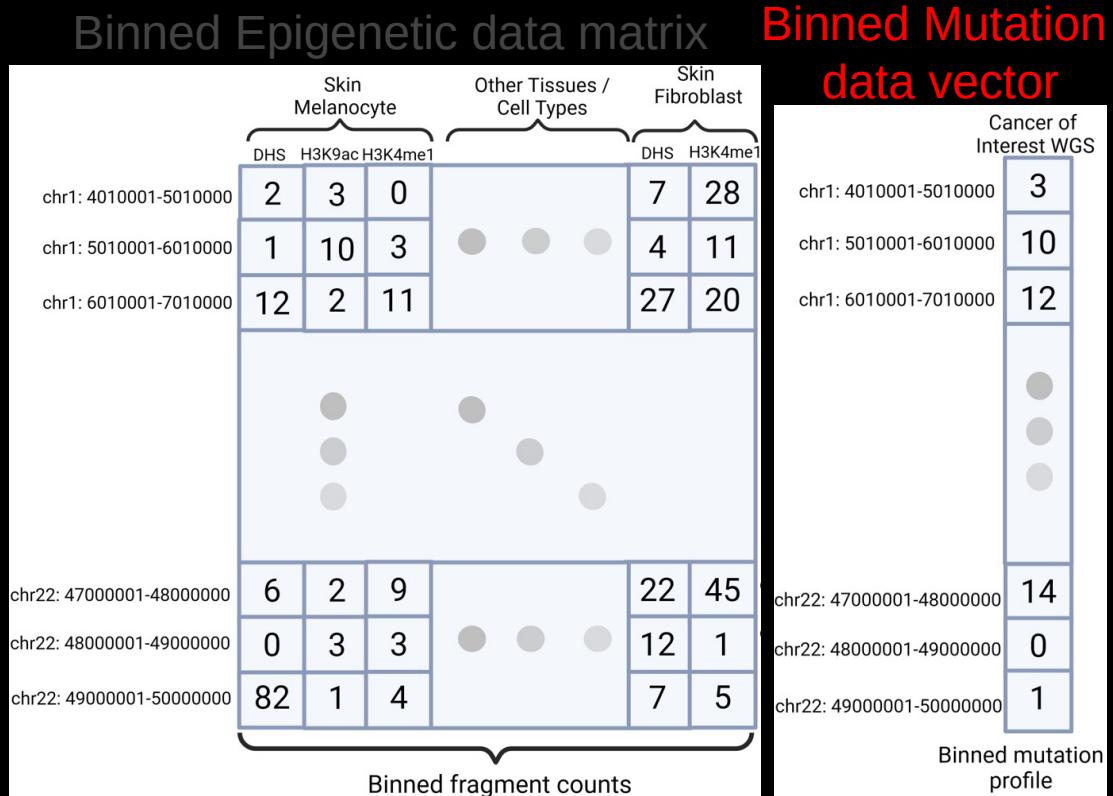
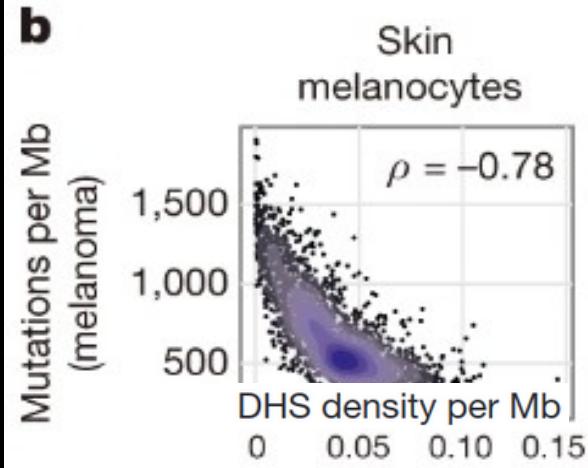
Binned Epigenetic data matrix

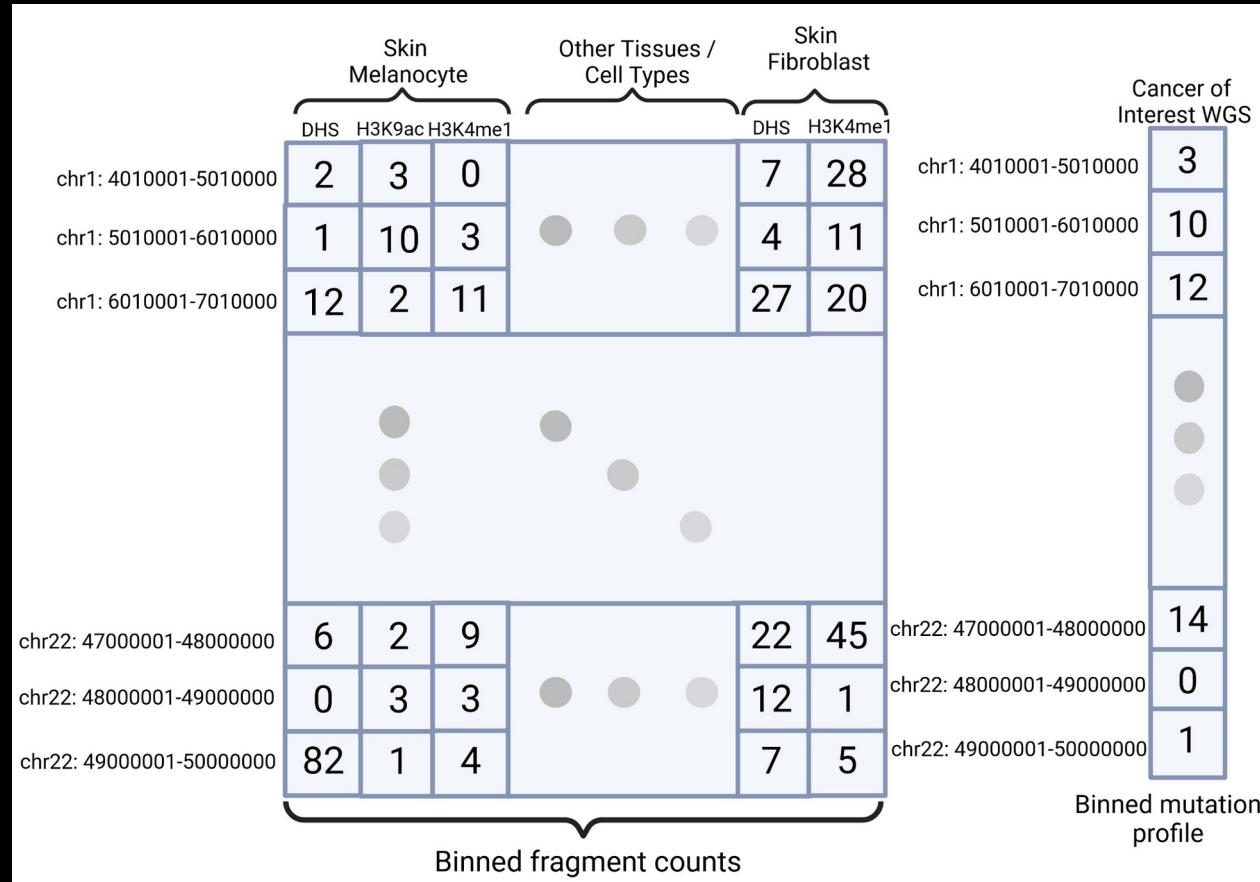
Binned fragment counts

Skin Melanocyte			Other Tissues / Cell Types			Skin Fibroblast		Cancer of Interest WGS
	DHS	H3K9ac	H3K4me1			DHS	H3K4me1	
chr1: 4010001-5010000	2	3	0			7	28	3
chr1: 5010001-6010000	1	10	3	●	●	4	11	10
chr1: 6010001-7010000	12	2	11			27	20	12
chr22: 47000001-48000000	6	2	9			22	45	14
chr22: 48000001-49000000	0	3	3	●	●	12	1	0
chr22: 49000001-50000000	82	1	4			7	5	1

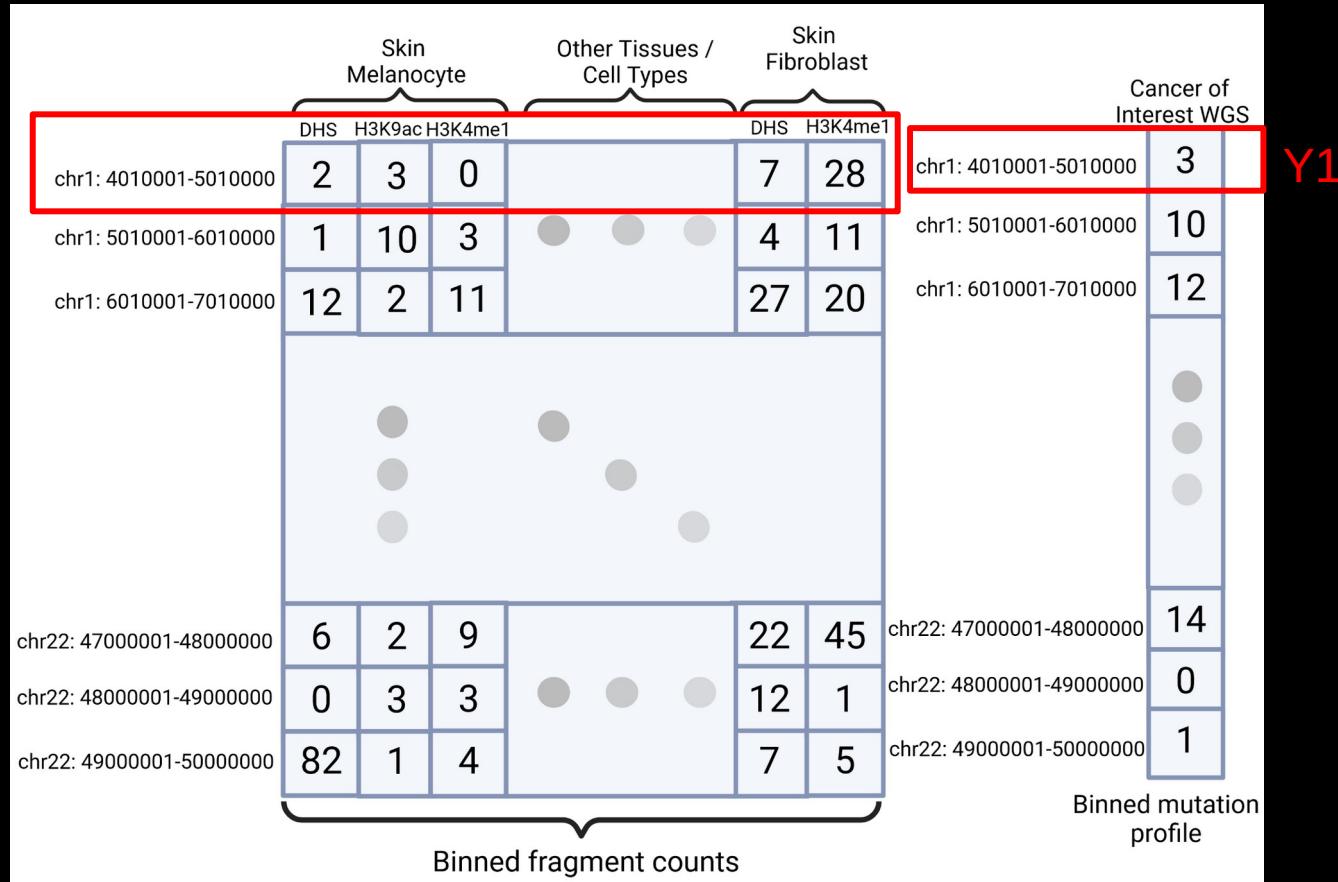
Binned mutation profile

- How to calculate correlation?



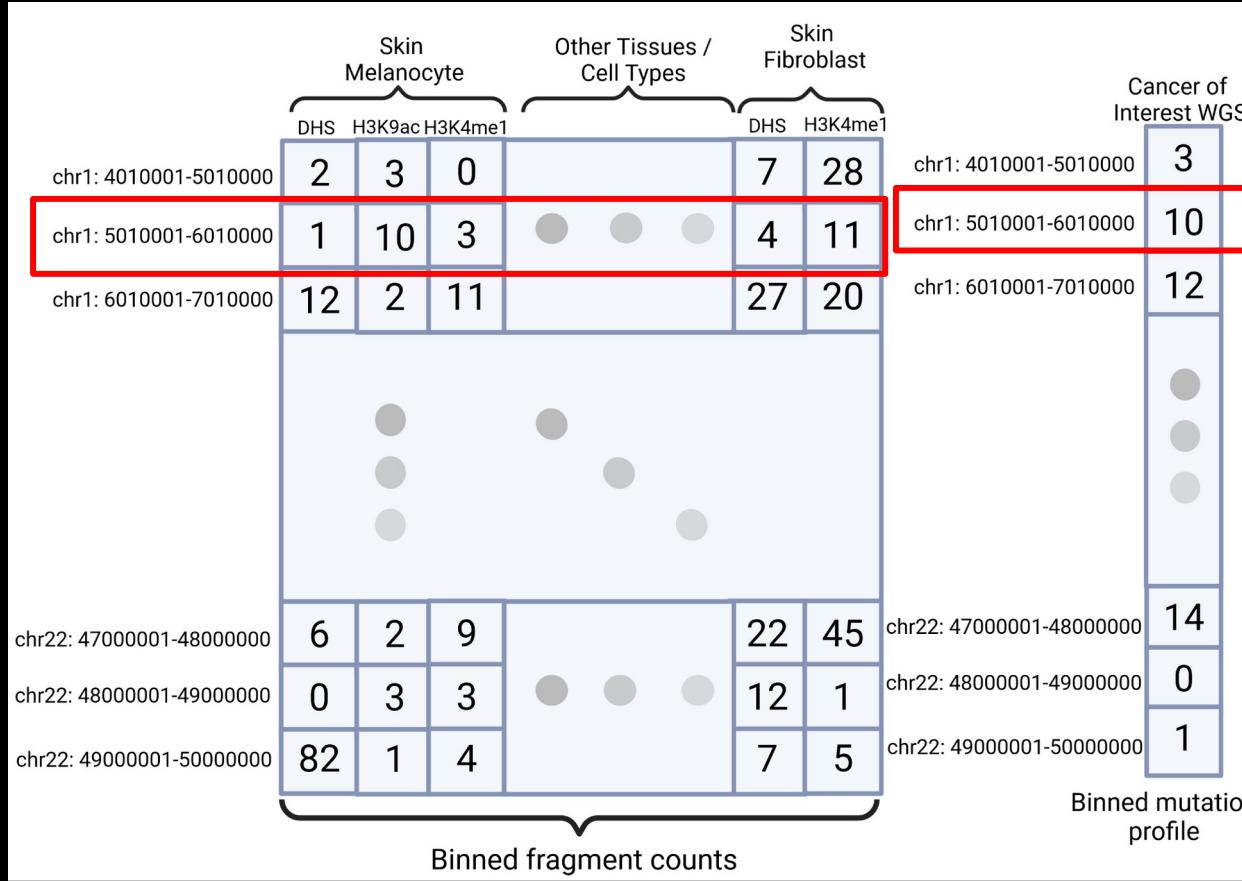


X1



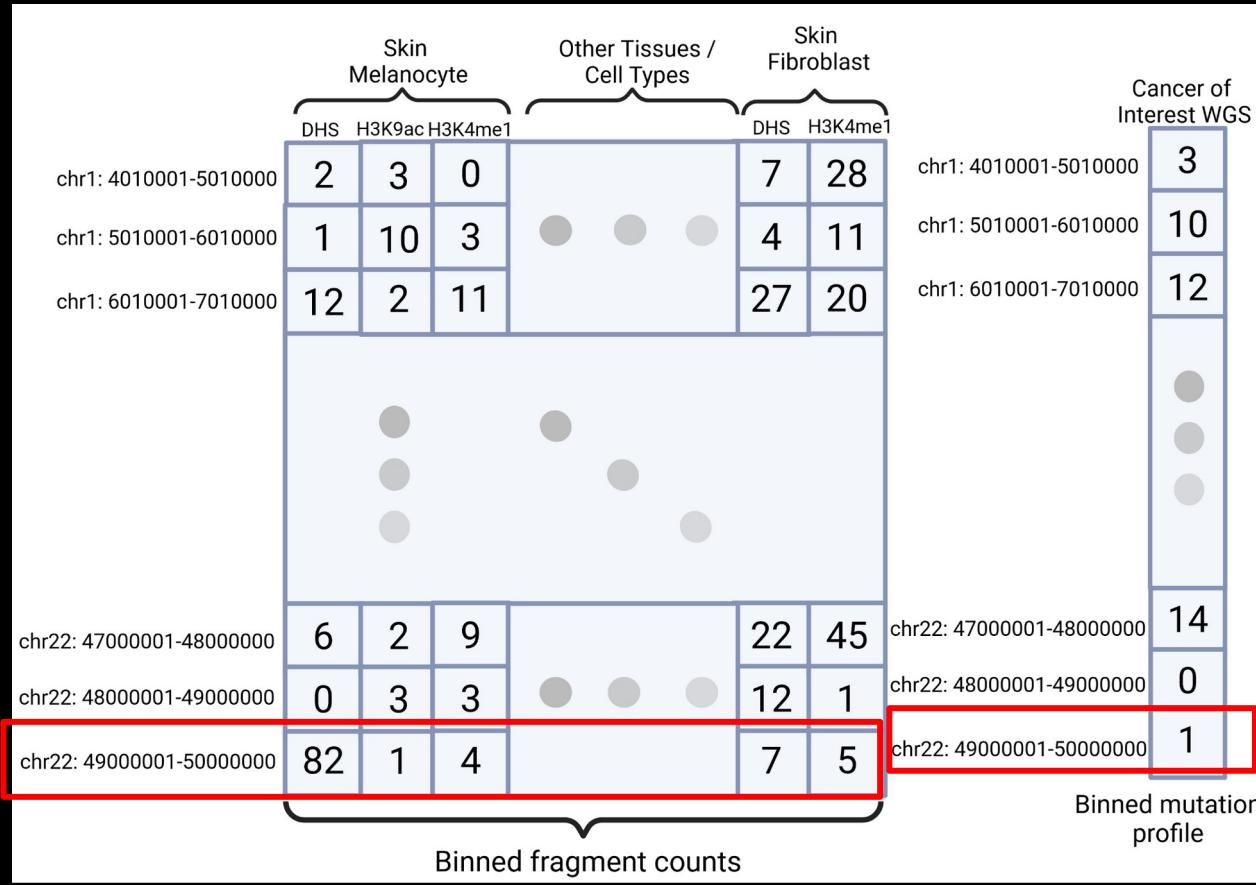
Y1

X2



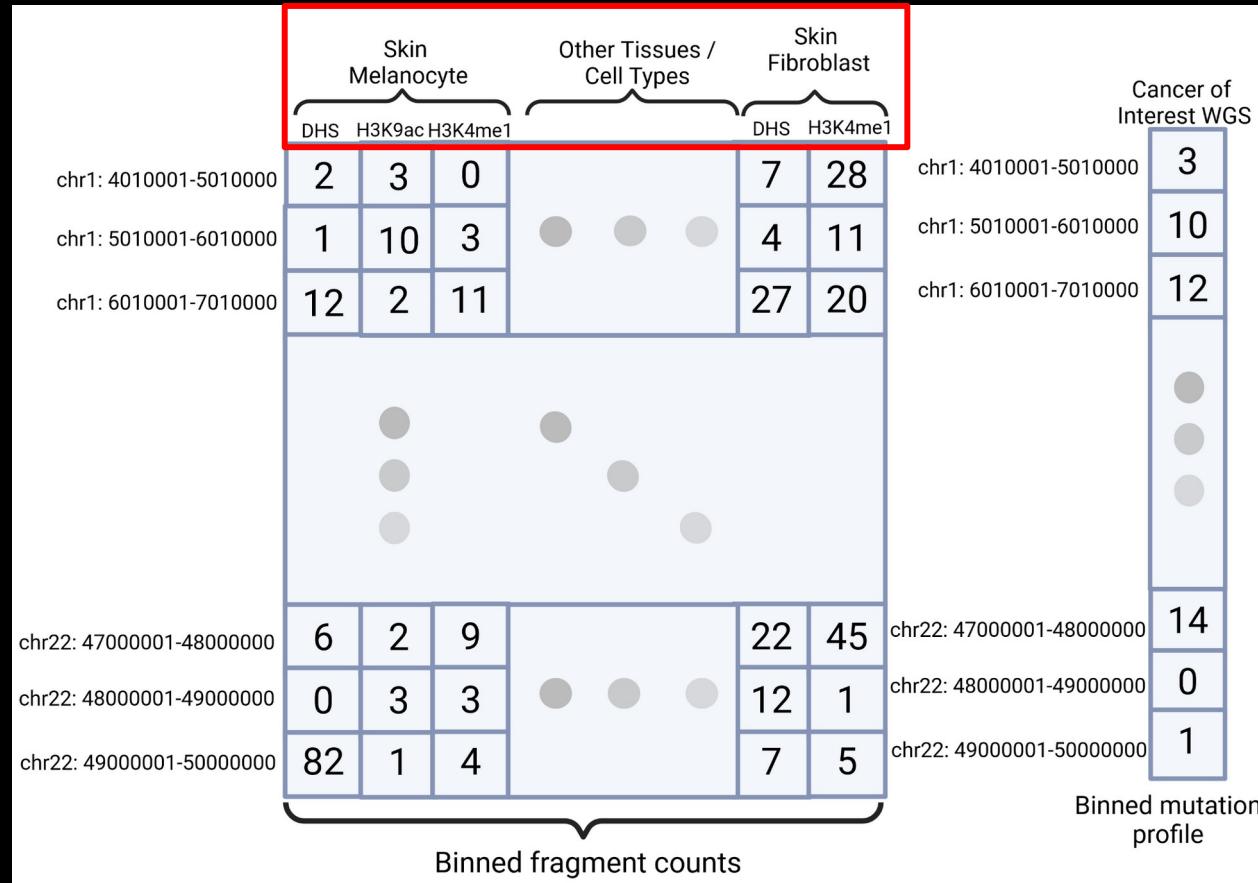
Y2

X2128

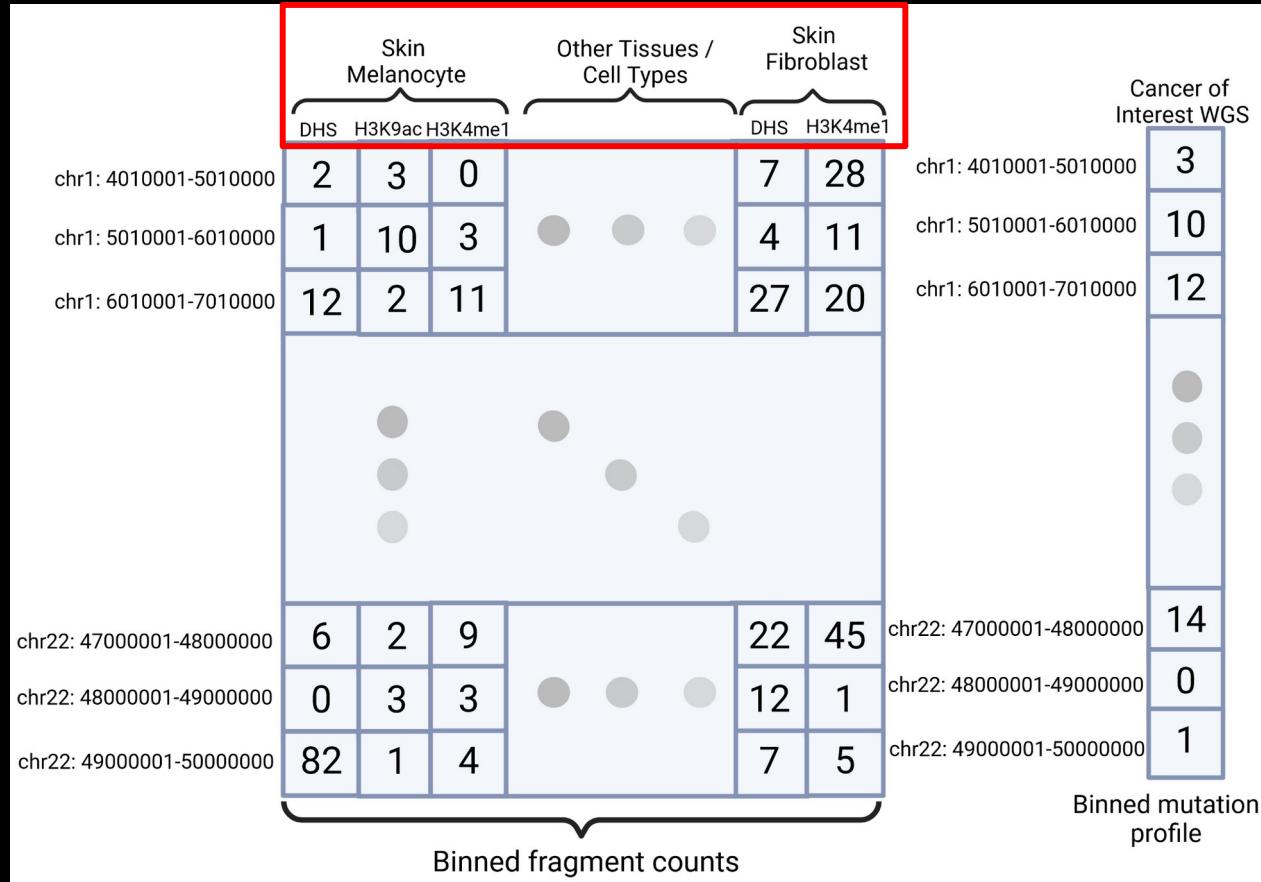


Y2128

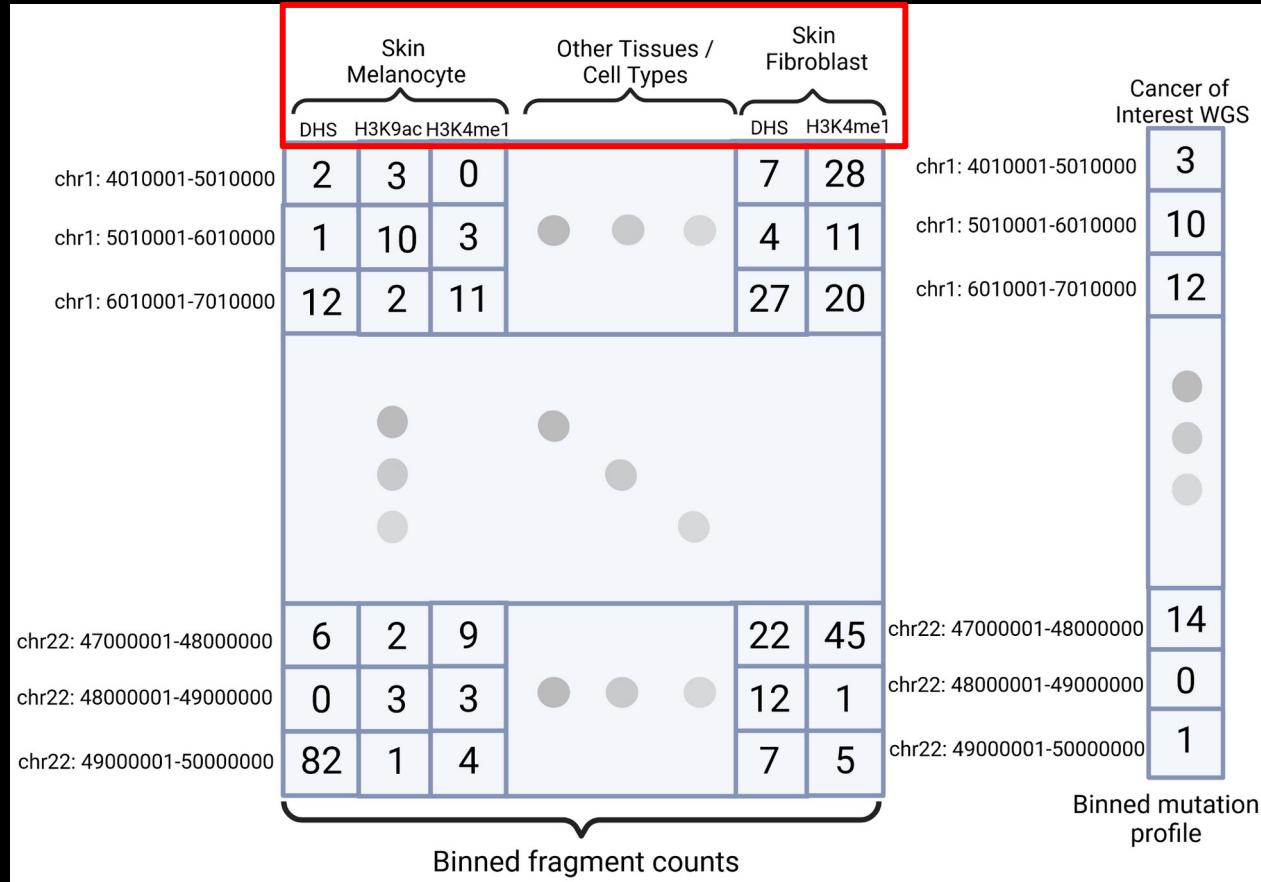
Predictors



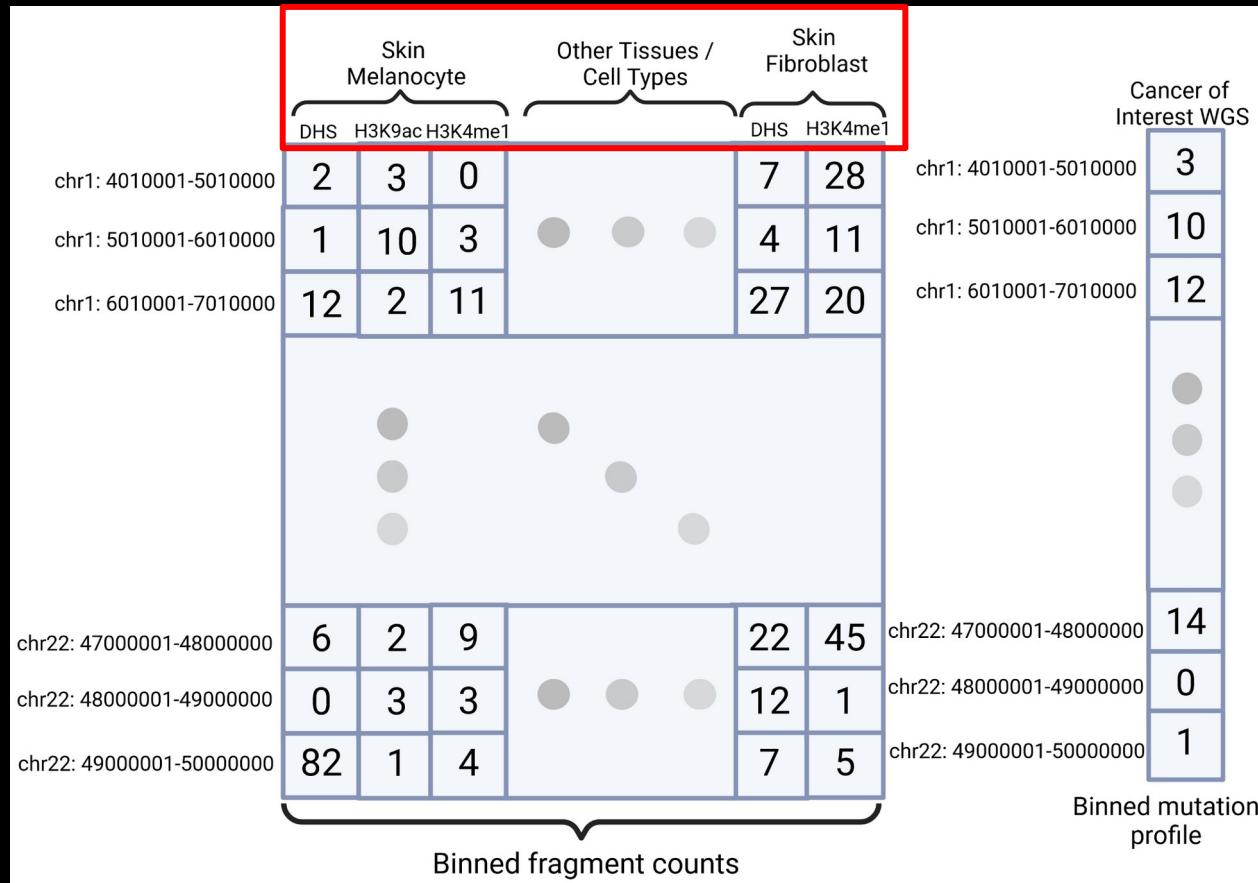
aka independent variables



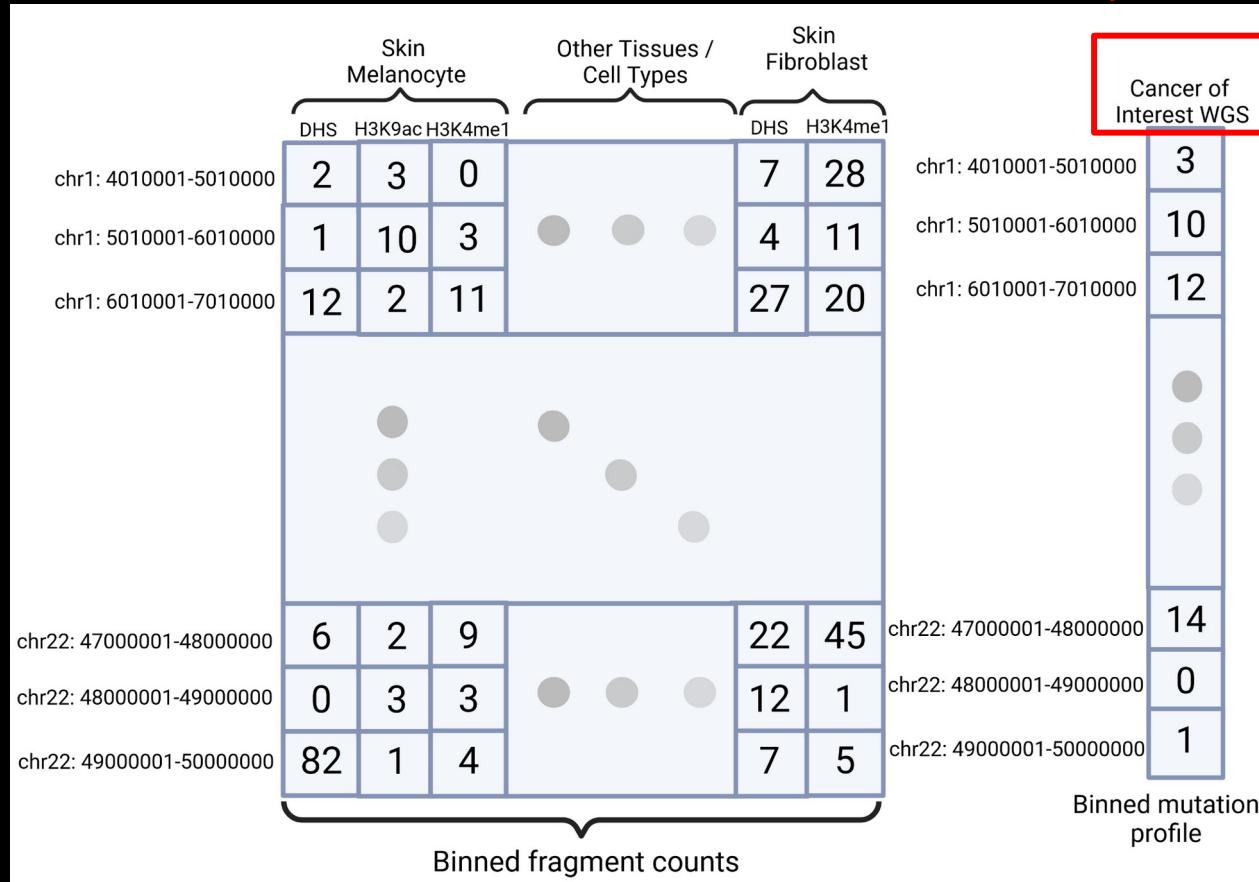
aka features



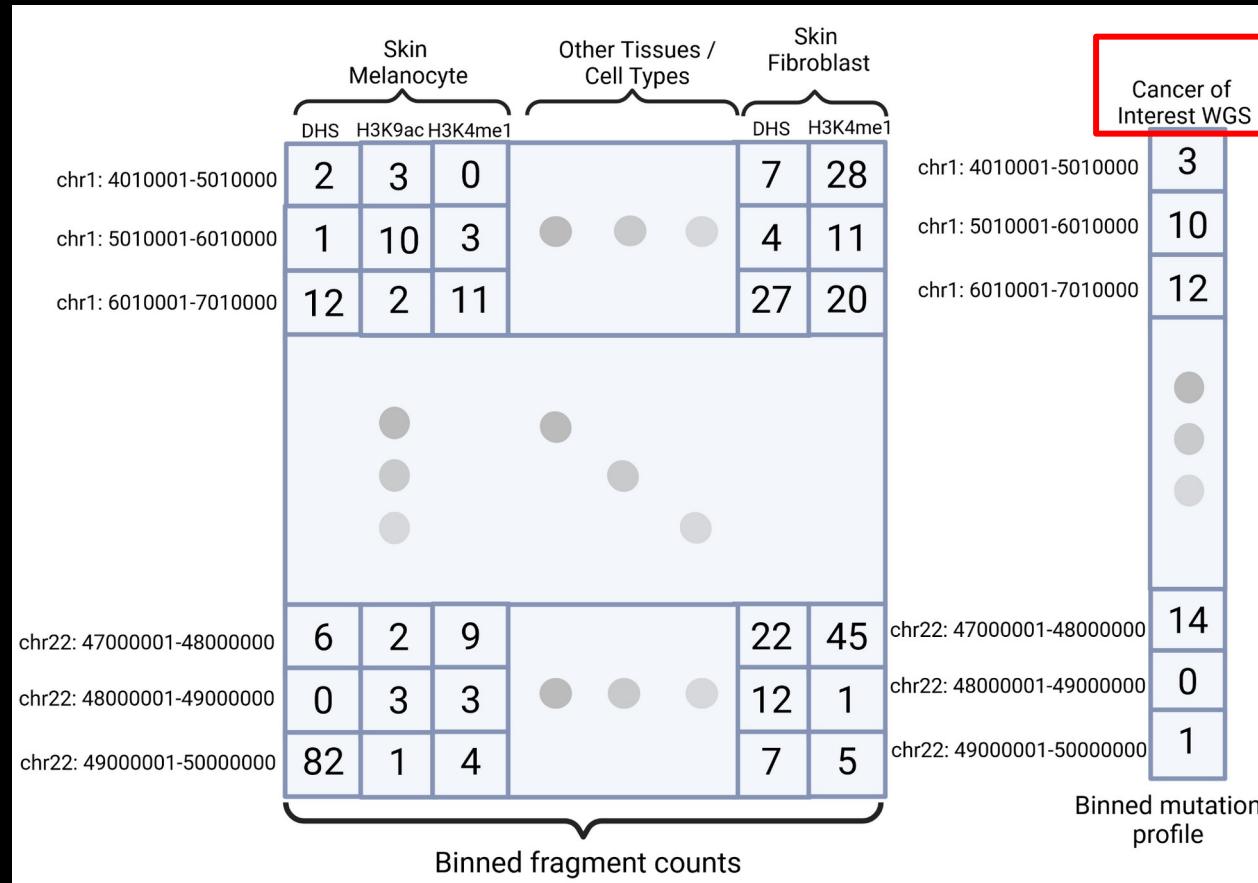
aka the thing you use to predict the other thing



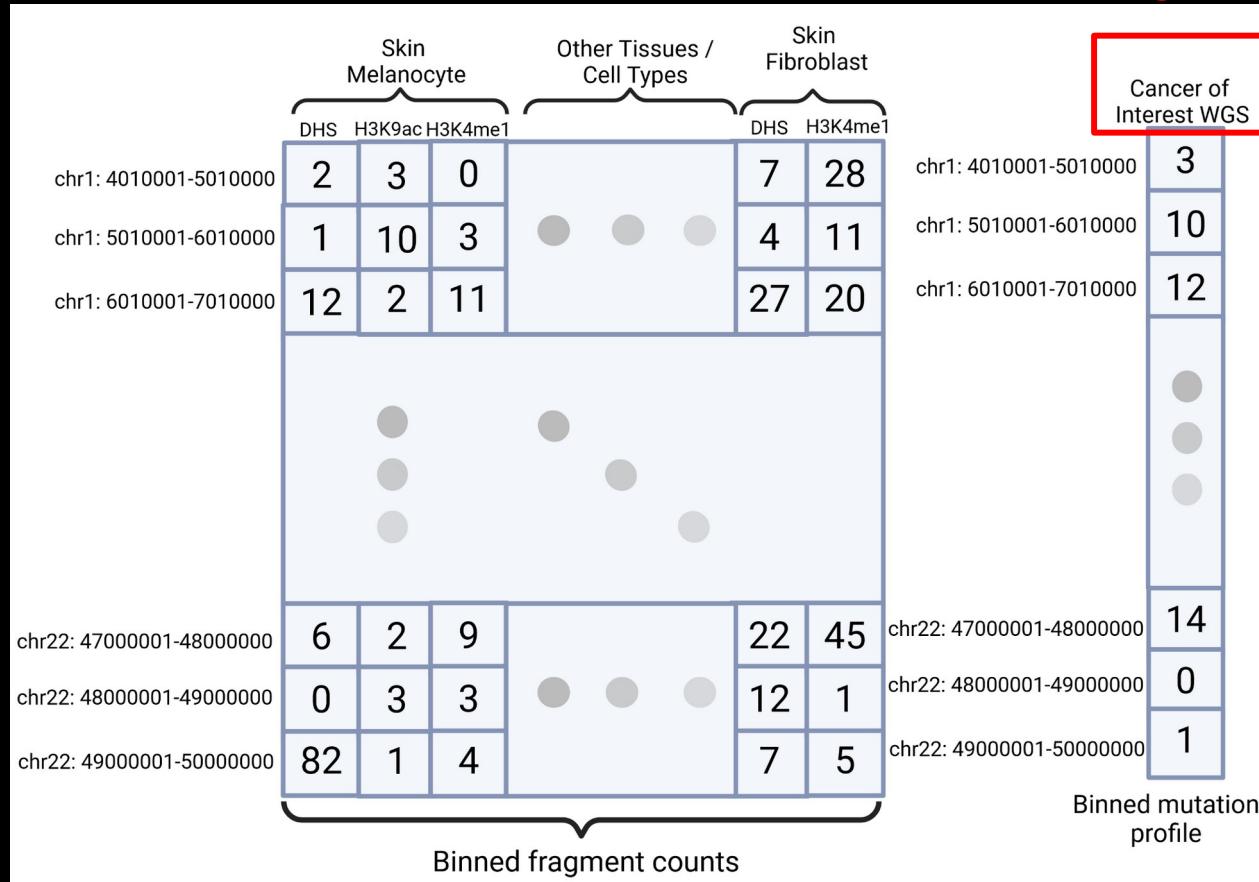
Response variable



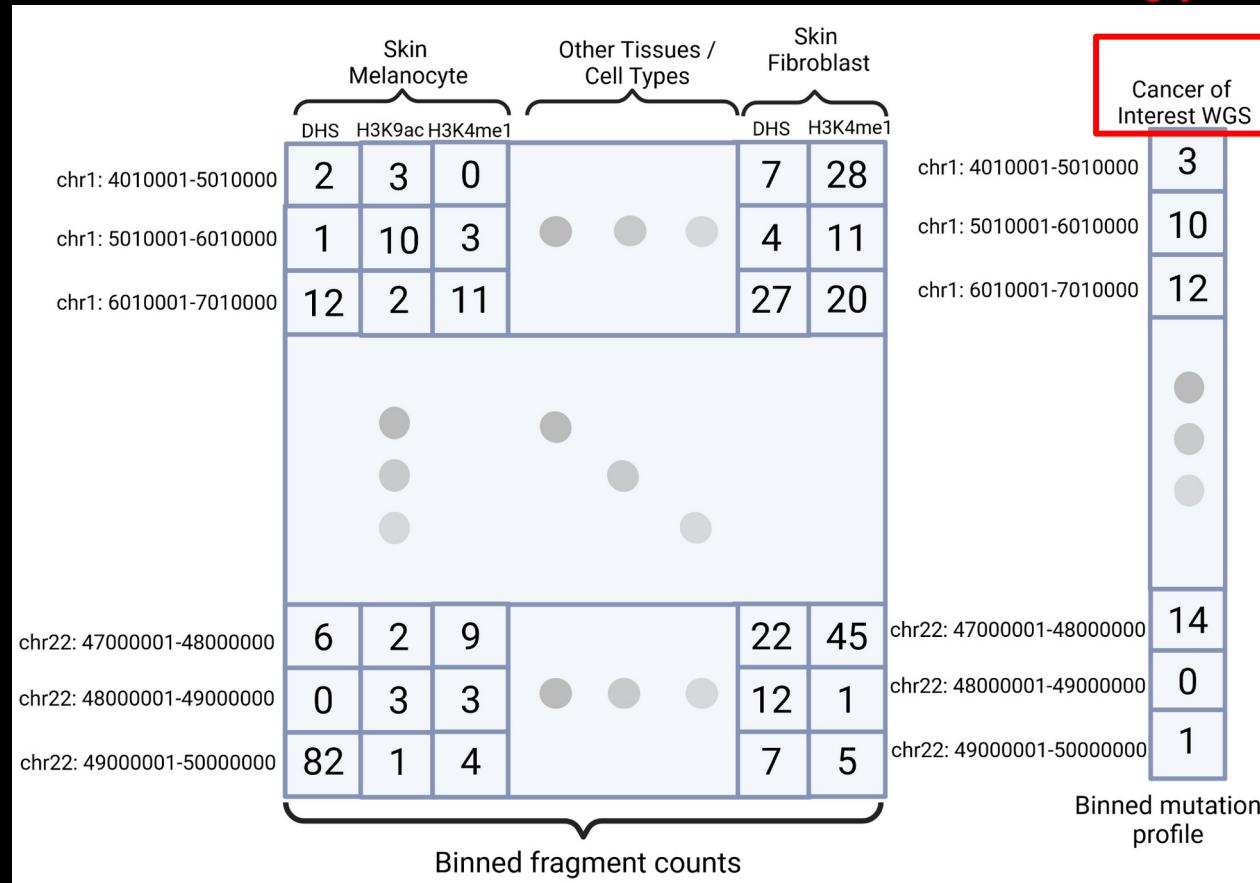
aka dependent variable

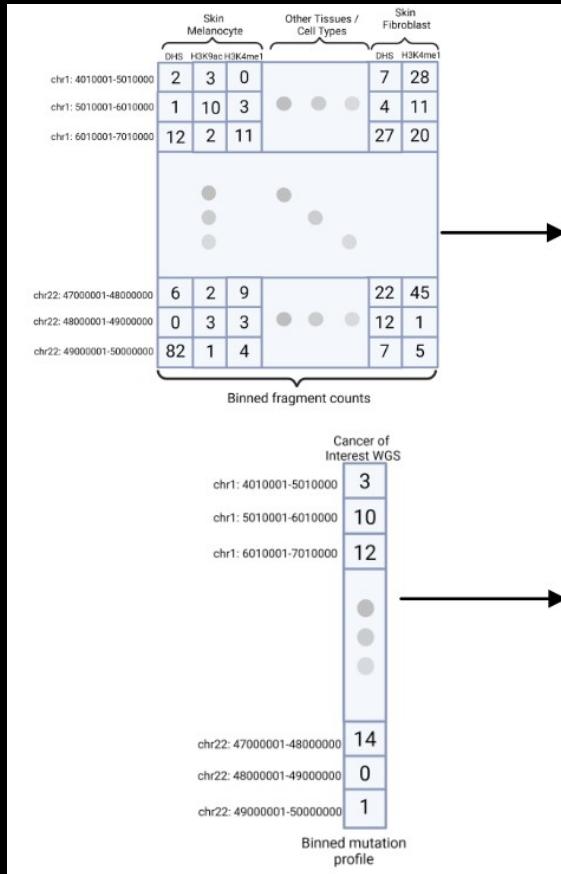


aka target variable

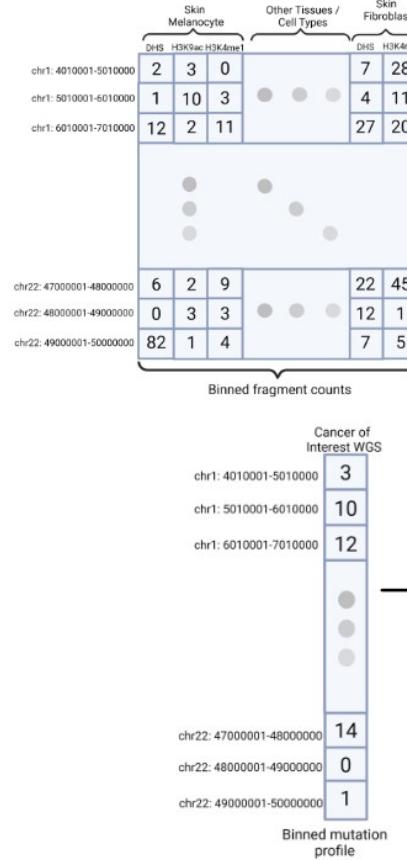


aka the thing you're predicting

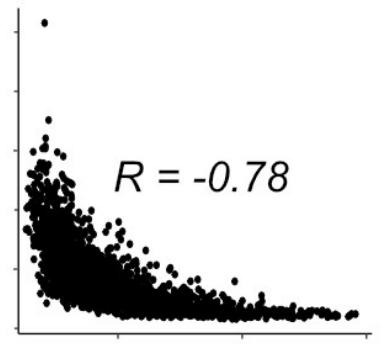




Machine Learning

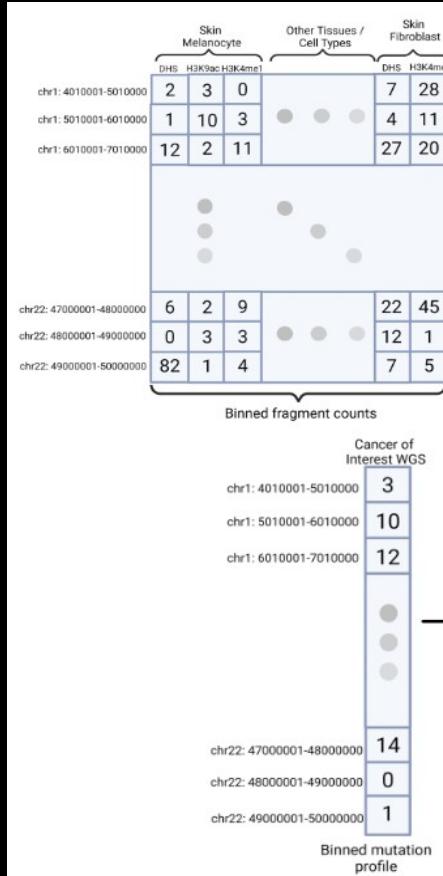


Machine Learning



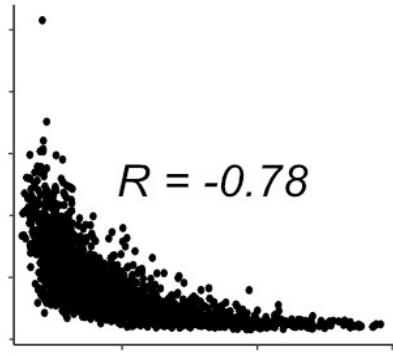
Chromatin accessibility
e.g Skin Melanocyte

Mutational density
e.g Melanoma



→ Chromatin accessibility
e.g Skin Melanocyte

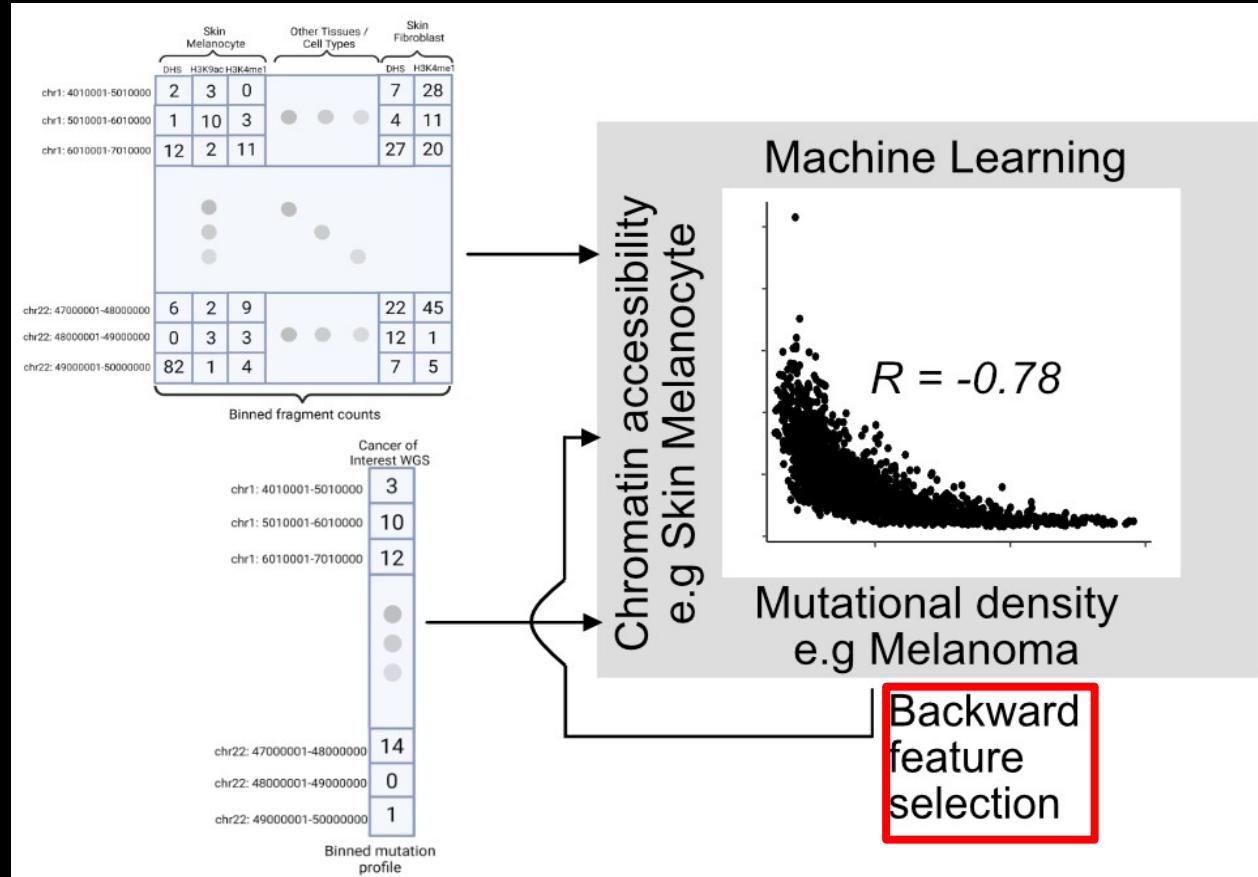
Machine Learning



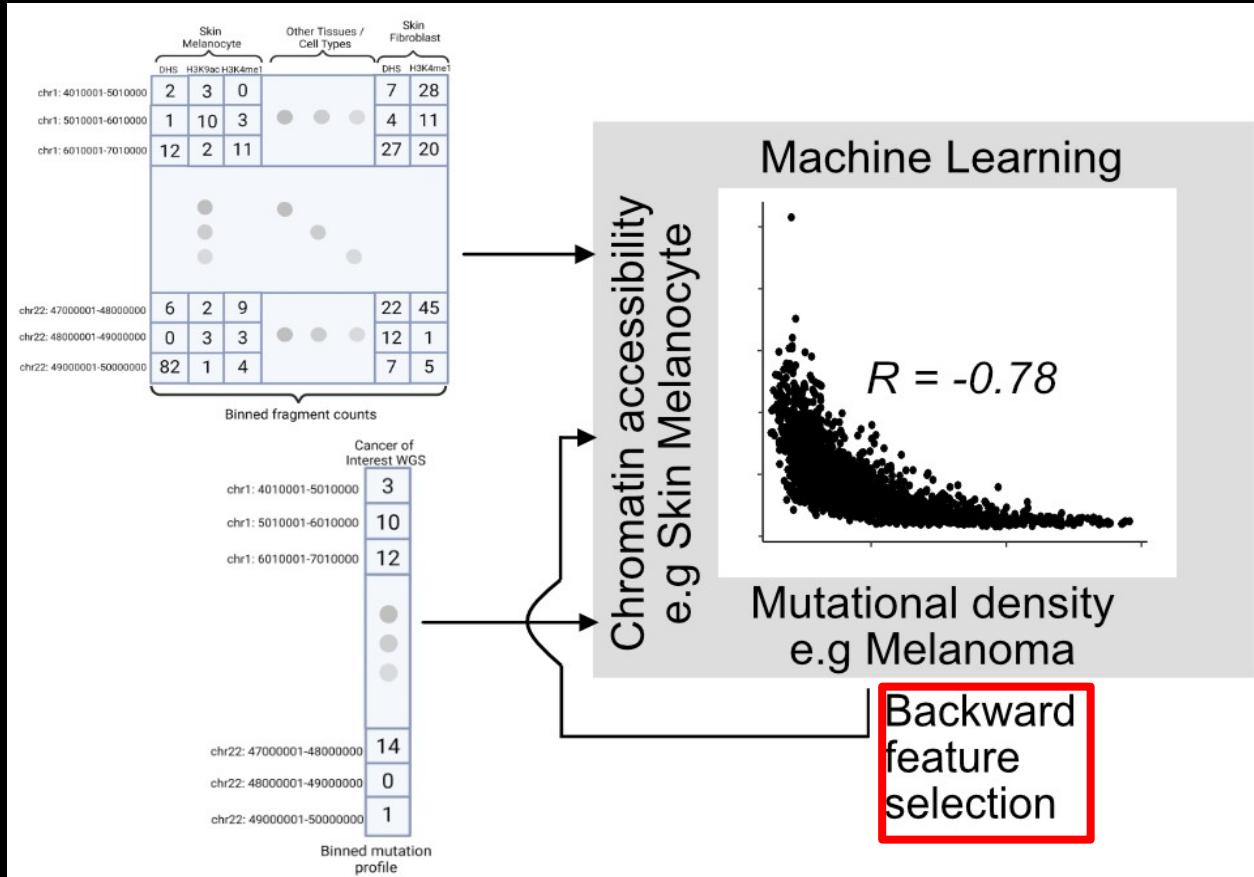
Mutational density
e.g Melanoma

Backward
feature
selection

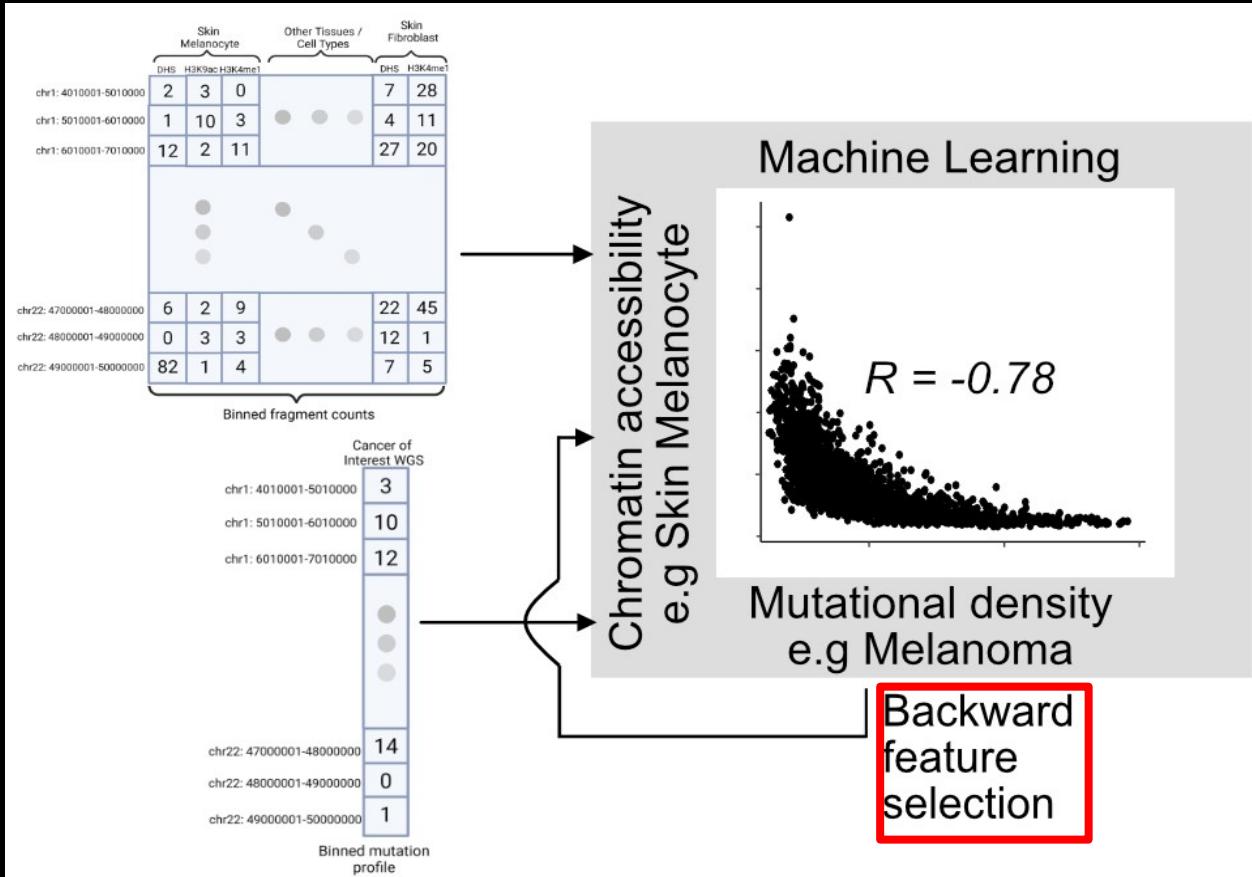
One of many methods
to determine
importance of
predictors (here,
epigenetic predictors)



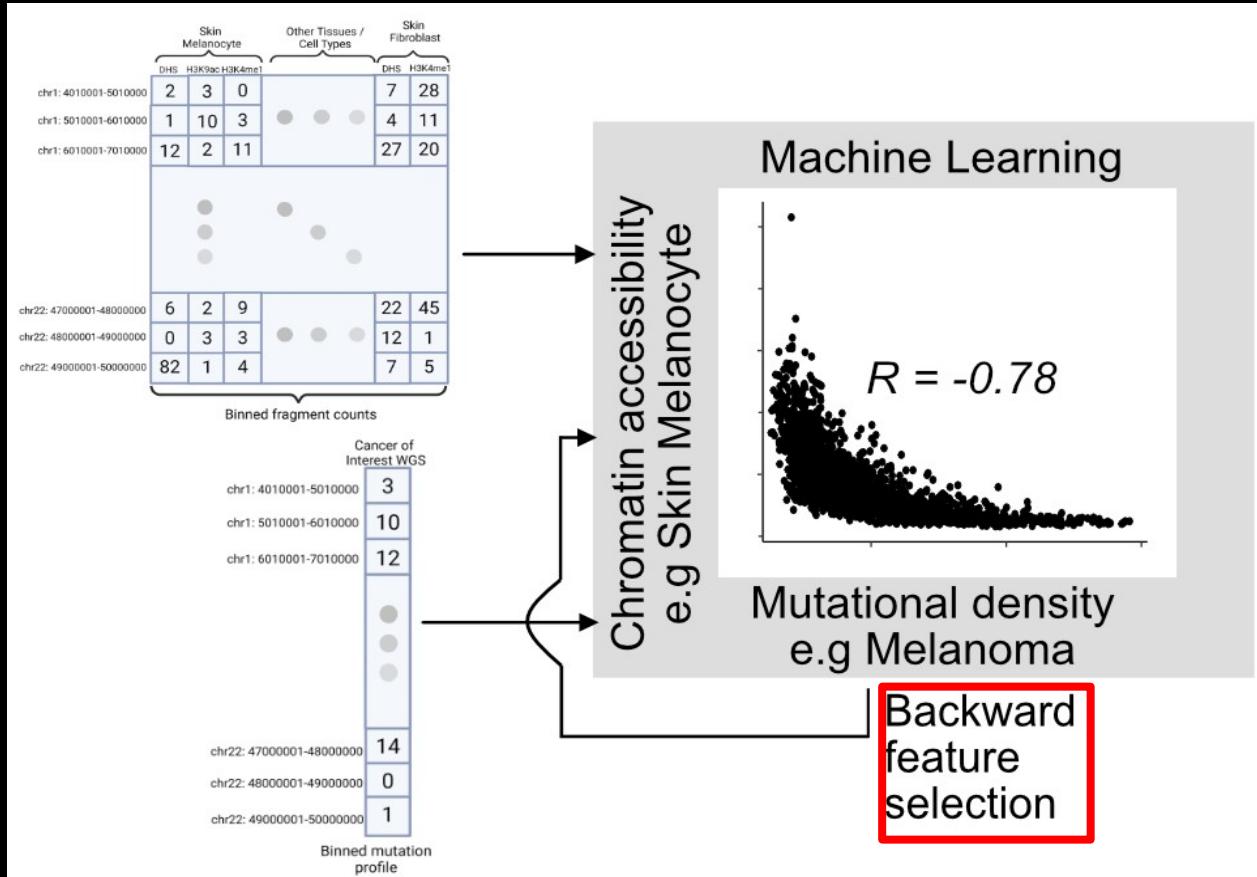
- Say # predictors = 500
- Get top 20
- Build new model on top 20
- Get top 19
- Build model
- Get top 18
- .
- .
- .
- 1 feature left

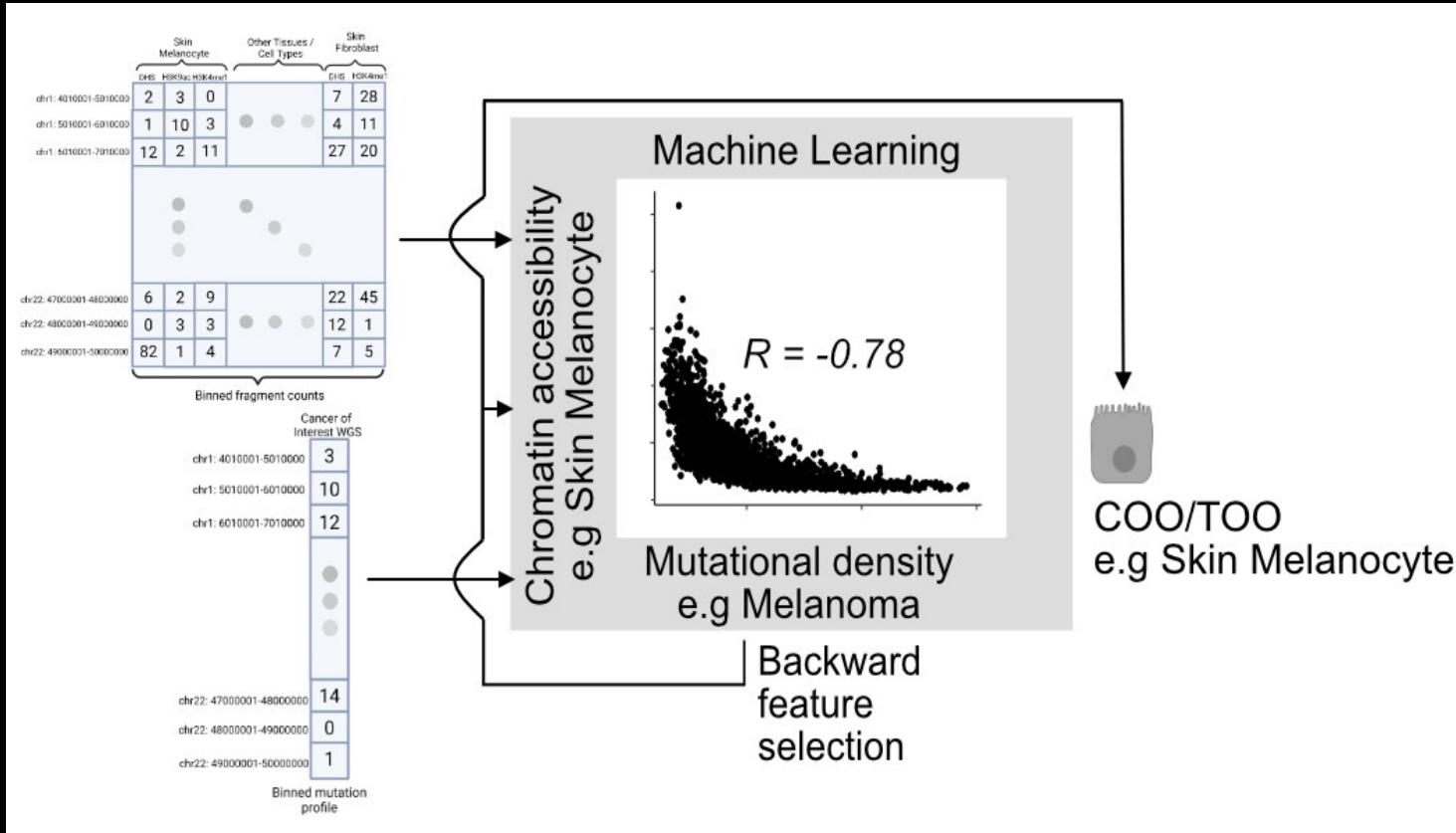


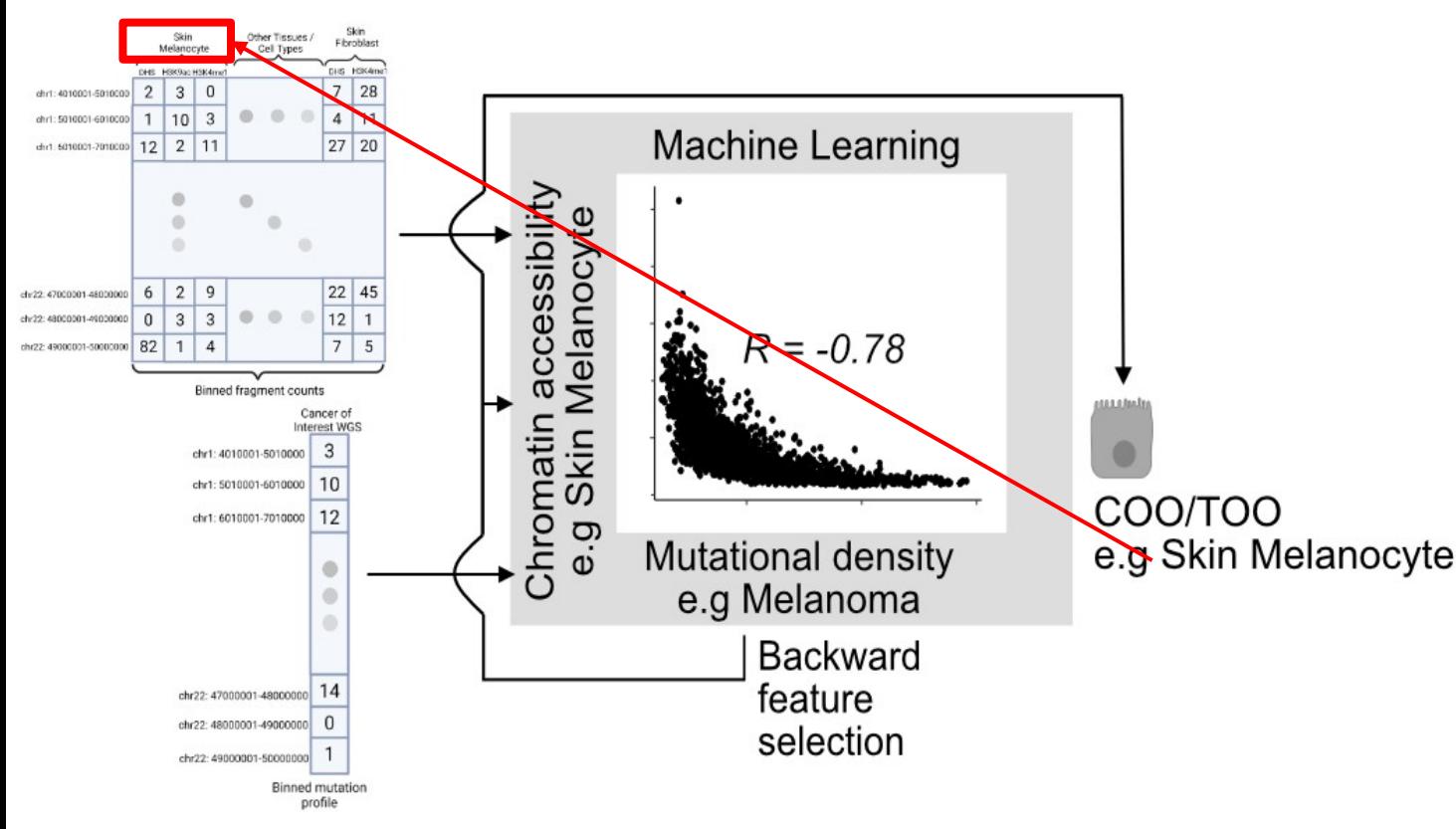
Expect that most
Important feature
comes from the COO



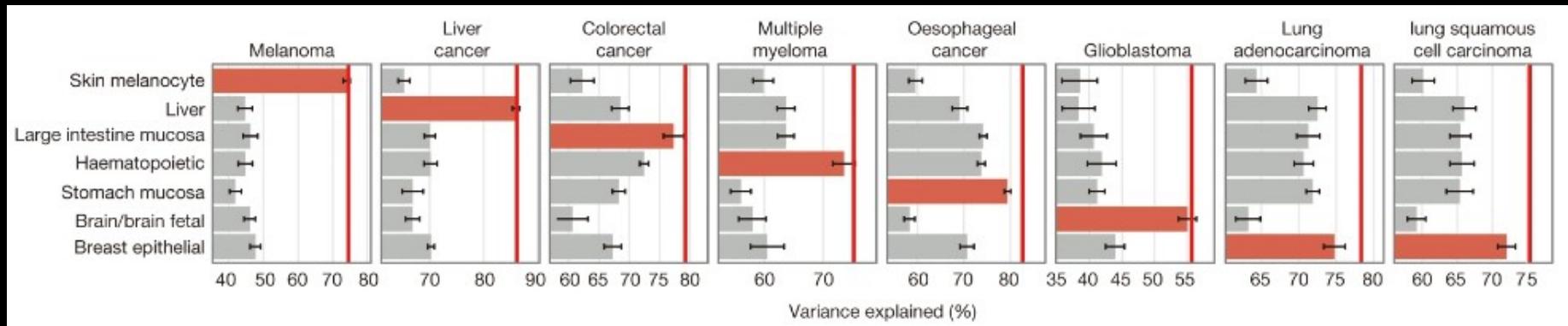
i.e we expect that epigenetics of COO most strongly predict mutational density







- The previous approach has been used (pretty) successfully to predict the COO (mostly, the tissue of origin aka TOO) using bulk epigenetic assays

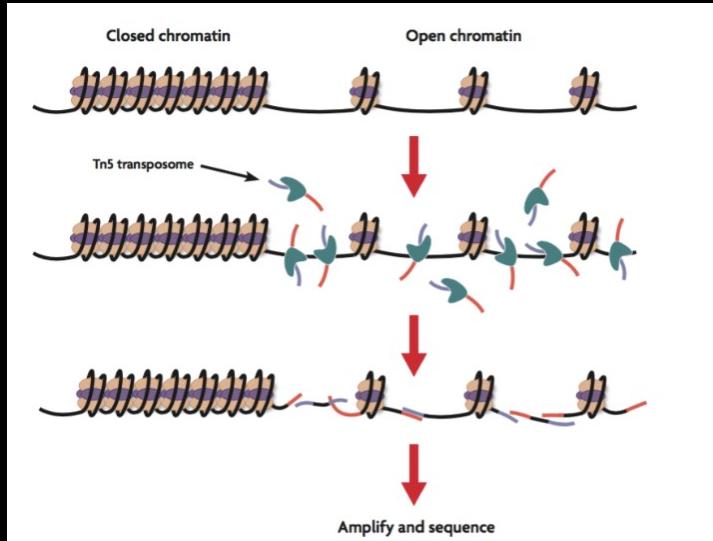


Paz Polak, Rosa Karlic
et al. Nature 2015.

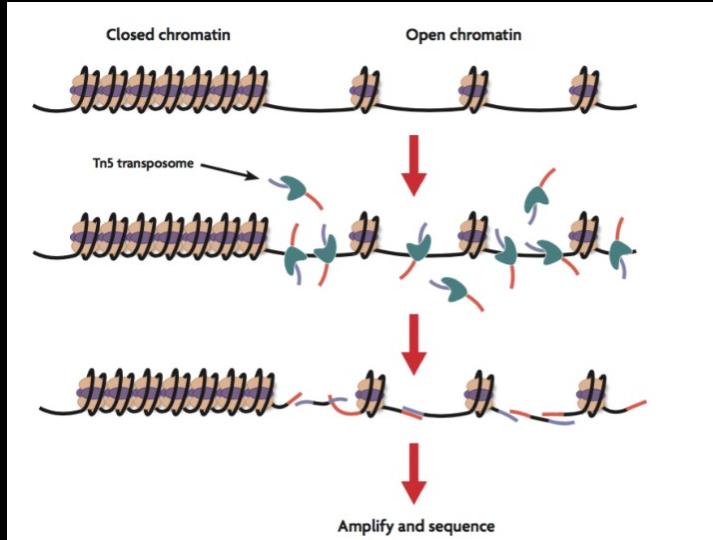
Current Work

- Epigenetic feature of choice: ATAC-seq
 - High throughput sequencing, Chromatin accessibility

- Epigenetic feature of choice: ATAC-seq
 - High throughput sequencing, Chromatin accessibility

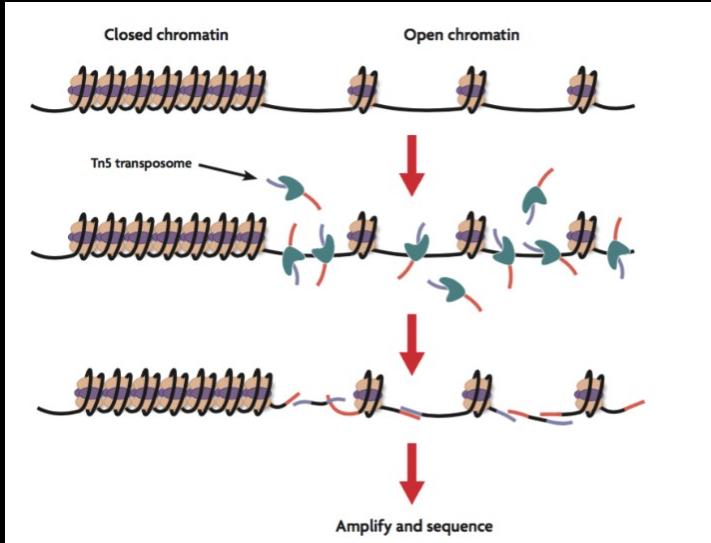


- Epigenetic feature of choice: ATAC-seq
 - High throughput sequencing, Chromatin accessibility



- Actually, **single-cell** ATAC-seq (scATAC)!

- Epigenetic feature of choice: ATAC-seq
 - High throughput sequencing, Chromatin accessibility



- Actually, single-cell ATAC-seq (scATAC)!
 - Analog of scRNA-seq vs RNA-seq

- Why scATAC?
 - Data

- Why scATAC?

- Data
- Data

- Why scATAC?

- Data
- Data
- **Data**

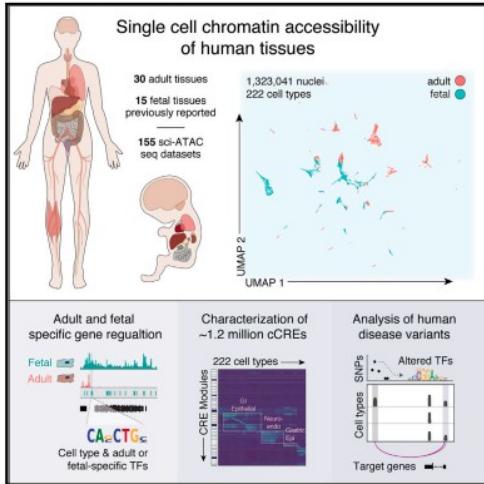
- Why scATAC?

- Data
- Data
- Data
- **DATA!!!!!!!!!!!!!!**

- Why scATAC?
 - Data
 - Data
 - Data
 - DATA!!!!!!!!!!!!!!
 - Untapped TROVE of SINGLE CELL data!

A single-cell atlas of chromatin accessibility in the human genome

Graphical abstract



Authors

Kai Zhang, James D. Hocker,
Michael Miller, ..., Allen Wang,
Sebastian Preissl, Bing Ren

Correspondence
biren@health.ucsd.edu

In brief

A cell-type-resolved map of human *cis*-regulatory elements, derived from single cell analysis of diverse tissue types, facilitates functional interpretation of the noncoding variants associated with complex human traits and diseases.

Highlights

- Integrating > 1.3 million single-cell chromatin profiles from adult/fetal human tissues
- An atlas of ~1.2 million candidate *cis*-regulatory elements across 222 cell types
- Cell-type specificity of fetal and adult candidate *cis*-regulatory elements
- Interpretation of noncoding variants associated with complex traits and diseases

ADULT ATLAS

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

A human cell atlas of fetal chromatin accessibility

Silvia Domcke*, Andrew J. Hill*, Riza M. Daza*, Junyue Cao, Diana R. O'Day, Hannah A. Pliner, Kimberly A. Aldinger, Dmitry Pokholok, Fan Zhang, Jennifer H. Milbank, Michael A. Zager, Ian A. Glass, Frank J. Steemers, Dan Doherty, Cole Trapnell†, Darren A. Cusanovich†, Jay Shendure†

INTRODUCTION: In recent years, the single-cell genomics field has made incredible progress toward disentangling the cellular heterogeneity of human tissues. However, the overwhelming majority of effort has been focused on single-cell gene expression rather than the chromatin landscape that shapes and is shaped by gene expression. Toward advancing our understanding of the regulatory programs that underlie human cell types, we set out to generate single-cell atlases of both chromatin accessibility (this study) and gene expression (Cao *et al.*, this issue) from a broad range of human fetal tissues.

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play key roles in the determination and maintenance of cell fates. Accessible regions are also markedly enriched for genetic variation that contributes to common disease heritability. The vast majority of chromatin accessibility data collected to date lacks single-cell resolution, limiting our ability to infer patterns such as which cell types are most relevant to each common disease. We previously demonstrated single-cell profiling of chromatin accessibility using combinatorial indexing, based on two rounds of *in situ* molecular barcoding. Here, we describe an improved assay that uses three levels of combinatorial

indexing and does not rely on custom reagents. The method, sci-ATAC-seq3, reduces costs and opens the door to the scales necessary for generating a human cell atlas of chromatin accessibility.

RESULTS: We applied sci-ATAC-seq3 to 59 human fetal samples ranging from 89 to 125 days in estimated postconceptual age and representing 15 organs, altogether obtaining high-quality chromatin accessibility profiles from ~800,000 single cells. Gene expression data collected on an overlapping set of tissues were leveraged to annotate cell types. We asked which transcription factor (TF) motifs found in the accessible sites of each cell best explain its cell type affiliation, revealing both known and potentially previously unknown regulators of cell fate specification and/or maintenance. Many TFs could be putatively assigned as activators or repressors depending on whether their expression and the accessibility of their cognate motif were positively or negatively correlated across cell types. Comparing chromatin accessibility from cell types that appear in multiple tissues revealed that whereas blood cell types are highly similar across organs, endothelial cells exhibit organ-specific chromatin accessibility, which appears

to be controlled combinatorially by several TFs with overlapping expression patterns. We leveraged our master set of 1.05 million accessible sites, spanning 532 Mb or 17% of the reference human genome, to score cell type-specific links between candidate enhancers and genes based on coaccessibility, to detect cell type-specific enrichment of heritability for specific common human diseases, and to identify genetic variants affecting chromatin accessibility in *cis*. Comparisons with chromatin accessibility in corresponding adult tissues allowed us to identify fetal-specific cell subtypes and nominate POU2F1 as a potential regulator of excitatory neuron development.

CONCLUSION: Sci-ATAC-seq3 adds to a growing repertoire of single-cell methods that use combinatorial indexing, a technical paradigm whose advantages include exponential scaling and greater range to profile diverse aspects of single-cell biology. We anticipate that the intersection of single-cell chromatin accessibility and gene expression will critically accelerate the field's long-term goal of establishing a deep, predictive understanding of gene regulation. An interactive website facilitates the exploration of these freely available data by tissue, cell type, locus, or motif (descartes.brotmanbaty.org). ■

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: darrenc@email.arizona.edu

(D.A.C.); coletrap@uvu.edu (C.T.); shendure@uw.edu (J.S.)

Cite this article as S. Domcke *et al.*, *Science* **370**, eaba7612 (2020). DOI: [10.1126/science.aba7612](https://doi.org/10.1126/science.aba7612)



READ THE FULL ARTICLE AT

<https://doi.org/10.1126/science.aba7612>

FETAL ATLAS

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

A human cell atlas of fetal chrom

Silvia Domcke*, Andrew J. Hill*, Riza M. Daza*, Junyue Cao, Di Kimberly A. Aldinger, Dmitry Pokholok, Fan Zhang, Jennifer H. Ian A. Glass, Frank J. Steemers, Dan Doherty, Cole Trapnell†, □

INTRODUCTION: In recent years, the single-cell genomics field has made incredible progress toward disentangling the cellular heterogeneity of human tissues. However, the overwhelming majority of effort has been focused on single-cell gene expression rather than the chromatin landscape that shapes and is shaped by gene expression. Toward advancing our understanding of the regulatory programs that underlie human cell types, we set out to generate single-cell atlases of both chromatin accessibility (this study) and gene expression (Cao *et al.*, this issue) from a broad range of human fetal tissues.

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play key roles in the determination and maintenance of cell fates. Accessible regions are also markedly enriched for genetic variation that contributes to common disease heritability. The vast majority of chromatin accessibility data collected to date lacks single-cell resolution, limiting our ability to infer patterns such as which cell types are most relevant to each common disease. We previously demonstrated single-cell profiling of chromatin accessibility using combinatorial indexing, based on two rounds of *in situ* molecular barcoding. Here, we describe an improved assay that uses three levels of combinatorial

indexing gents. Ti costs and for gener accessible

RESULTS: man fetal in estim senting 1 quality c ~800,000 collected leverage which tr in the act its cell ty and pote tors of a nance. M as activa whether ity of the negative paring ct that app whereas across or specific c

RESEARCH ARTICLE SUMMARY

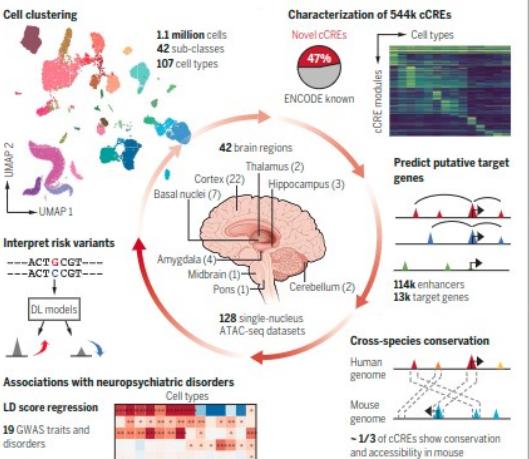
BICCN

A comparative atlas of single-cell chromatin accessibility in the human brain

Yang Eric Li, Sebastian Preissl, Michael Miller, Nicholas D. Johnson, Zihan Wang, Henry Jiao, Chenxu Zhu, Zhaoqing Wang, Yang Xie, Olivier Poirion, Colin Kern, Antonio Pinto-Duarte, Wei Tian, Kimberly Siletti, Nora Emerson, Julia Osteene, Jacinta Lucero, Lin Lin, Qian Yang, Quan Zhu, Nathan Zemke, Sarah Espinoza, Anna Marie Yanny, Julie Nyhus, Nick Dee, Tamara Casper, Nadya Shapovalova, Daniel Hirschstein, Rebecca D. Hodge, Sten Linnarsson, Trygve Bakken, Boaz Levi, C. Dirk Keene, Jingbo Shang, Ed Lein, Allen Wang, M. Margarita Behrens, Joseph R. Ecker, Bing Ren*

INTRODUCTION: Neuropsychiatric disorders and mental illnesses are the leading cause of disease burden in the United States. Tens of thousands of sequence variants in the human genome have been linked to the etiology of these conditions. However, elucidating the role of the identified risk variants remains a challenge because most of them are outside of protein-coding regions and currently lack functional annotation. These

disease risk variants likely exert their influence by perturbing transcriptional regulatory elements, thereby modulating gene expression in cell types pertinent to neuropsychiatric disorders. Recent advancements in single-cell technologies have unveiled a high degree of cellular heterogeneity across the human brain. However, the transcriptional regulatory sequences governing the identity and function of each individual



Single-cell analysis of chromatin accessibility of the human brain. Candidate cis-regulatory elements (cCREs) specific to distinct human brain cell types were identified by single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) and linked to putative target genes through integrative analysis. The usage of cCREs was leveraged to predict brain cell types pertinent to neuropsychiatric traits and disorders and to train machine-learning models to interpret the function of noncoding risk variants. UMAP, Uniform Manifold Approximation and Projection; DL, deep learning; LD, linkage disequilibrium; GWAS, genome wide association study.

brain cell type remain to be delineated, hindering our ability to interpret the noncoding disease risk variants.

RATIONALE: Conventionally, transcriptional regulatory sequences may be determined by evidence of chromatin accessibility that generally accompanies transcription factor binding and chromatin remodeling. However, prior catalogs of transcriptional regulatory elements lack the information about cell-type-specific activities of each element because of the use of bulk tissue samples. Recent technological strides have empowered us to analyze chromatin accessibility at the single-cell level, enabling the creation of cell-type-specific maps of transcriptional regulatory elements for complex organs such as the human brain.

RESULTS: In this study, we present a comprehensive analysis of chromatin accessibility in the human brain at the single-cell level, encompassing a collection of 1.1 million cells from 42 distinct brain regions in three neurotypical adult subjects. We used this chromatin atlas to define 107 distinct brain cell types and uncovered the state of chromatin accessibility at 544,735 putative transcriptional regulatory elements in these cell types. A substantial number of these regulatory elements also exhibited both sequence conservation and chromatin accessibility in mouse brain cells, underlining their functional importance. Through integrative analysis, we have linked many putative transcriptional regulatory elements to potential target genes. We further leveraged this atlas to predict disease relevant cell types for 19 neuropsychiatric traits and disorders. Finally, we developed machine learning models to predict the regulatory function of disease risk variants. We have made this atlas freely available to the public through an interactive web portal CATLAS (www.catlas.org).

CONCLUSION: The single-cell chromatin atlas of the human brain represents a valuable resource for the neuroscience community. It offers insights into the gene-regulatory programs shaping the diversity of brain cell types and aids in interpreting the functional roles of disease risk variants located outside of protein-coding regions. This atlas, in combination with other molecular and anatomical data, promises to advance our understanding of brain function and neuropathology, ultimately offering avenues for more effective approaches to addressing neuropsychiatric disorders. ■

The list of author affiliations is available in the full article online.

*Corresponding author. Email: biren@health.usc.edu
Cite this article as Y. E. Li *et al.*, Science 382, eaf7044 (2023). DOI: 10.1126/science.adf7044

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adf7044>

BRAIN ATLAS

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

A human cell atlas of fetal chrom

Silvia Domcke*, Andrew J. Hill*, Riza M. Daza*, Junyue Cao, Kimberly A. Aldinger, Dmitry Pokholok, Fan Zhang, Jennifer Ian A. Glass, Frank J. Steemers, Dan Doherty, Cole Trapnell†, indexin gents. † costs an for gene accessit

INTRODUCTION: In recent years, the single-cell genomics field has made incredible progress toward disentangling the cellular heterogeneity of human tissues. However, the overwhelming majority of effort has been focused on single-cell gene expression rather than the chromatin landscape that shapes and is shaped by gene expression. Toward advancing our understanding of the regulatory programs that underlie human cell types, we set out to generate single-cell atlases of both chromatin accessibility (this study) and gene expression (Cao *et al.*, this issue) from a broad range of human fetal tissues.

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play key roles in the determination and maintenance of cell fates. Accessible regions are also markedly enriched for genetic variation that contributes to common disease heritability. The vast majority of chromatin accessibility data collected to date lacks single-cell resolution, limiting our ability to infer patterns such as which cell types are most relevant to each common disease. We previously demonstrated single-cell profiling of chromatin accessibility using combinatorial indexing, based on two rounds of *in situ* molecular barcoding. Here, we describe an improved assay that uses three levels of combinatorial



ARTICLE

<https://doi.org/10.1038/s41467-021-27660-3>

OPEN

Single-cell chromatin accessibility landscape in kidney identifies additional cell-of-origin in heterogeneous papillary renal cell carcinoma

Qi Wang^{1,8}, Yang Zhang^{1,2,8}, Bolei Zhang³, Yao Fu⁴, Xiaozhi Zhao⁵, Jing Zhang¹, Ke Zuo², Yuexian Xing², Song Jiang^{5,2}, Zhaohui Qin^{1,6}, Erguang Li¹, Hongqian Guo^{1,5✉}, Zhihong Liu^{1,2✉} & Jingping Yang^{1,2,7✉}

KIDNEY

Papillary renal cell carcinoma (pRCC) is the most heterogeneous renal cell carcinoma. Patient survival varies and no effective therapies for advanced pRCC exist. Histological and molecular characterization studies have highlighted the heterogeneity of pRCC tumours. Recent studies identified the proximal tubule (PT) cell as a cell-of-origin for pRCC. However, it remains elusive whether other pRCC subtypes have different cell-of-origin. Here, by obtaining genome-wide chromatin accessibility profiles of normal human kidney cells using single-cell transposase-accessible chromatin-sequencing and comparing the profiles with pRCC samples, we discover that besides PT cells, pRCC can also originate from kidney collecting duct principal cells. We show pRCCs with different cell-of-origin exhibit different molecular characteristics and clinical behaviors. Further, metabolic reprogramming appears to mediate the progression of pRCC to the advanced state. Here, our results suggest that determining cell-of-origin and monitoring origin-dependent metabolism could potentially be useful for early diagnosis and treatment of pRCC.

Silvia Domcke^{1*}, Andrew J. Hi
Kimberly A. Aldinger, Dmitry
Ian A. Glass, Frank J. Steeme

INTRODUCTION: In recent years, the genomics field has made significant progress toward disentangling the complex landscape of human tissues. However, the majority of effort has been focused on cell gene expression rather than the regulatory landscape that shapes gene expression. Toward advancing our understanding of the regulatory programs that govern human cell types, we set out to create cell atlases of both chromatin accessibility (from a broad range of human

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play a critical role in the determination and differentiation fates. Accessible regions are enriched for genetic variation and common disease heritability. Chromatin accessibility is currently limited by the resolution of single-cell sequencing methods. The ability to infer patterns such as those most relevant to each cell type has previously been demonstrated through the use of chromatin accessibility profiles and indexing, based on two-dimensional barcoding. Here, we describe a new assay that uses three-level

A deep lung cell atlas reveals cytokine-mediated lineage switching of a rare cell progenitor of the human airway epithelium

for updates

Avinash Waghray^{*1,3,7,12}, Isha Monga^{*4}, Brian Lin^{*1,3,6,7}, Viral Shah^{*1,3,5}, Michal Slyper^{*2,10}, Bruno Giotti^{*4}, Jiajie Xu^{1,3}, Julia Waldman², Danielle Dionne², Lan T. Nguyen², Wendy Lou², Peiwen Cai⁴, Eric Park⁴, Christoph Muus^{2,8}, Jiawei Sun^{1,3}, Manalee V. Surve^{1,3}, Lujia Cha Cha Yang⁹, Orit Rozenblatt-Rosen², Toni M Dolerey², Srinivas Vinod Saladi^{7,11}, Alexander M Tsankov⁴⁺, Aviv Regev^{2,10+}, Jayaraj Rajagopal^{1,2,3,5,7+}

1 Center for Regenerative Medicine, Massachusetts General Hospital, Boston, USA

2 Klarman Cell Observatory, Broad Institute of Harvard and MIT, Cambridge, MA, USA

3 Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA.

4 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA

5 Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston, USA

6 Developmental, Molecular and Chemical Biology, Tufts University School of Medicine, Boston, USA

7 Broad Institute of MIT and Harvard, Cambridge, MA, USA

8 John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

9 Harvard University, Boston, USA

10 Genentech, South San Francisco, CA, USA.

11 Department of Otolaryngology Head and Neck Surgery, Massachusetts Eye and Ear Infirmary, Harvard Medical School.

12 Currently at Immunology and Inflammation Research Therapeutic Area, Sanofi, Cambridge, MA

* co-first authors

+ co-senior authors; to whom correspondence should be addressed: alexander.tsankov@mssm.edu (A.T.), aviv.regev.sc@gmail.com (A.R.), jrajagopal@mgh.harvard.edu (J.R.).

LUNG ATLAS

C

RESEARCH

A

RESEARCH ART

h

HUMAN GENOMICS

Gr

A human cell at

Hi

Silvia Domcke*, Andrew J. Hi
Kimberly A. Aldinger, Dmitry
Ian A. Glass, Frank J. Steeme

Hi

INTRODUCTION: In recent years, the genomics field has made significant progress toward disentangling the complex regulatory landscapes of human tissues. However, the majority of effort has been focused on cell gene expression rather than the landscape that shapes it and its expression. Toward advancing our understanding of the regulatory programs that govern human cell types, we set out to create cell atlases of both chromatin accessibility and gene expression (from a broad range of human tissues).

Hi

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play a critical role in the determination and differentiation of cell fates. Accessible regions are enriched for genetic variation associated with common disease heritability. Chromatin accessibility is a key feature of chromatin accessibility maps, which lack single-cell resolution. This limits the ability to infer patterns such as those most relevant to each cell type. Previously demonstrated methods of chromatin accessibility mapping involve indexing, based on two regular barcoding. Here, we describe a new assay that uses three-level

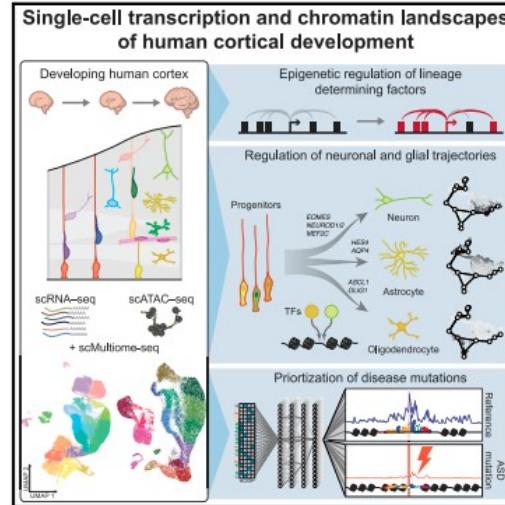
A deep investigation of progenitors

Avinash V.
Giotti*, Ji
Eric Park
Rozenblatt
2,10+, Jayar1 Center for
2 Klarman
3 Harvard
4 Department
New York
5 Department
Massachu
6 Developm
USA
7 Broad Inst
8 John A. F
MA, USA
9 Harvard
10 Genente
11 Departm
Medical Sc
12 Current
* co-first au
+ co-senior au
aviv.regev.

Cell

Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution

Graphical abstract



HIGHLIGHTS

- Single-cell RNA and chromatin profiling charts human corticogenesis
- Distinct TFs underlie neurogenesis and gliogenesis regulatory programs
- Lineage-determining TFs adopt an active chromatin state early in differentiation
- Neural networks prioritize noncoding *de novo* mutations in autism spectrum disorder

Authors

Alexandro E. Trevino, Fabian Müller,
Jimena Andersen, ..., Anshul Kundaje,
Sergiu P. Pasca, William J. Greenleaf

uno
Lai⁴,
Drit
Regev

Correspondence

spasca@stanford.edu (S.P.P.),
wjg@stanford.edu (W.J.G.)

In brief

A single-cell atlas of gene expression and chromatin accessibility of human developing cortex during mid-gestation reveals lineage-determining transcription factors for human corticogenesis and identifies prioritized mutations for autism spectrum disorder.

: for updates

Xing²,
1,2,7✉

FETAL BRAIN

vard

.T.),

C

RESEARCH

A
h

RESEARCH ART

Gr

HUMAN GENOMICS

A human cell at

Silvia Domcke^{1*}, Andrew J. Hi
Kimberly A. Aldinger, Dmitry
Ian A. Glass, Frank J. Steeme

INTRODUCTION: In recent years, the genomics field has made significant progress toward disentangling the complex landscape of human tissues. However, the majority of effort has been focused on cell gene expression rather than the landscape that shapes it and its expression. Toward advancing the regulatory program of human cell types, we set out cell atlases of both chromatin accessibility (from a broad range of human

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play a critical role in the determination and differentiation fates. Accessible regions are enriched for genetic variation and common disease heritability. Chromatin accessibility is a key feature that allows single-cell resolution to infer patterns such as those most relevant to each cell type. Previously demonstrated studies of chromatin accessibility use indexing, based on two regular barcoding. Here, we developed an assay that uses three level

A deep I

progenit

Avinash V
Giotti^{*4}, Ji
Eric Park¹
Rozenblatt-^{2,10+}, Jayar

1 Center fo
2 Klarman
3 Harvard S
4 Departme
New York
5 Departme
Massachu
6 Developm
USA
7 Broad Ins
8 John A. F
MA, USA
9 Harvard I
10 Genente
11 Departm
Medical Sc
12 Currentl

* co-first au
+ co-senior
aviv.regev.



OPEN

Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer

Winston R. Becker^{1,2,10}, Stephanie A. Nevins^{1,10}, Derek C. Chen¹, Roxanne Chiu¹, Aaron M. Horning¹, Tuhin K. Guha¹, Rozelle Laquindanum², Meredith Mills³, Hassan Chaib¹, Uri Ladabaum⁴, Teri Longacre⁵, Jeanne Shen^{1,5}, Edward D. Esplin¹, Anshul Kundaje^{1,6}, James M. Ford^{1,3}, Christina Curtis^{1,3,7}, Michael P. Snyder^{1,10} and William J. Greenleaf^{1,8,9}

To chart cell composition and cell state changes that occur during the transformation of healthy colon to precancerous adenomas to colorectal cancer (CRC), we generated single-cell chromatin accessibility profiles and single-cell transcriptomes from 1,000 to 10,000 cells per sample for 48 polyps, 27 normal tissues and 6 CRCs collected from patients with or without germline APC mutations. A large fraction of polyp and CRC cells exhibit a stem-like phenotype, and we define a continuum of epigenetic and transcriptional changes occurring in these stem-like cells as they progress from homeostasis to CRC. Advanced polyps contain increasing numbers of stem-like cells, regulatory T cells and a subtype of pre-cancer-associated fibroblasts. In the cancerous tissue, we observe T cell exhaustion, RUNX1-regulated cancer-associated fibroblasts and increasing accessibility associated with HNF4A motifs in epithelia. DNA methylation changes in sporadic CRC are strongly anti-correlated with accessibility changes along this continuum, further identifying regulatory markers for molecular staging of polyps.

The identification of genes and pathways that drive formation of invasive cancers has been the central focus of a number of large-scale genomic efforts^{1–3}. These efforts have catalogued the diversity and commonality of many genetic and transcriptional changes that accompany malignancy in diverse cancer types. However, most studies have focused on bulk profiling of advanced stage tumors and have largely ignored premalignant lesions. As a result, a detailed understanding of the progression of phenotypic changes that occur during the transition from normal to precancerous to cancerous state, as well as the molecular drivers of this transformation, remain underexplored.

CRC is an ideal system to study the continuum of phenotypic states along malignant transformation as it follows a stereotyped progression from normal to atypical to carcinoma that includes the formation of precancerous polyps⁴, which can subsequently give rise to CRCs. A number of the changes associated with these transitions are nearly universal to all CRC malignancies, as typified by the adenoma-to-carcinoma sequence^{5–8}. For example, an estimated 80–90% of colorectal tumors are initiated by loss of APC, resulting in β-catenin stabilization and increased WNT signaling⁹ leading to intestinal hyperplasia¹⁰. Subsequent mutations in other cancer driver genes such as KRAS, TP53 and SMAD4 result in the transformation to carcinoma.

Because APC mutations are almost universally the initiating event for polyps and CRCs, patients with familial adenomatous polyposis (FAP), who have germline mutations in APC, are a suitable

population in which to study the natural progression of polyps. These patients typically develop hundreds of polyps by early adulthood^{11,12}, and therefore an individual patient can provide numerous polyps of varied molecular ages and stages of progression, all arising in the same germline background.

To chart the regulatory and transcriptomic changes that occur on the phenotypic continuum from healthy colon to invasive carcinoma, as part of the Human Tumor Atlas Network¹³, we profiled single-nucleus transcriptomes (single-nucleus RNA sequencing (RNA-seq) (snRNA-seq)) and epigenomes (single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq) (scATAC-seq)) of healthy colon, polyps and CRCs. Many polyps were obtained from patients with FAP who underwent surgical resections, allowing both analysis of polyps with diverse sizes and locations of origin, and collection of neighboring unaffected colon tissue. From these single-cell datasets, we first catalog immune, stromal and epithelial cell types. We find large shifts in fibroblast subpopulations that occur along the transition from normal colon to CRC. We identify a subpopulation of exhausted T cells present only in CRC tissue. We observe a much larger fraction of cells exhibiting a stem-like state (both transcriptionally and epigenetically) within polyps and CRCs. We find that polyps populate an epigenetic and transcriptional continuum from normal colon to CRC characterized by sequential opening and closing of chromatin and upregulation and downregulation of genes associated with the cancer state. We identify regulatory elements and transcription factors

uno
Lai⁴,
Drit
Regev

In and
ation
ption
nd
utism

vard
. T.),

: for updates

COLON

INTRODUCTION: In recent years, the genomics field has made significant progress toward disentangling the complex landscape of human tissues. However, the majority of effort has been focused on cell gene expression rather than the broader landscape that shapes gene expression. Toward advancing our understanding of the regulatory programs governing human cell types, we set out to create cell atlases of both chromatin accessibility and gene expression (from a broad range of human cell types).

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play a critical role in the determination and differentiation of various cell fates. Accessible regions are enriched for genetic variation and common disease heritability. Chromatin accessibility is a key feature that lacks single-cell resolution, making it difficult to infer patterns such as those most relevant to each cell type. Previously demonstrated methods of chromatin accessibility indexing, based on two regular barcoding, have been shown to be less accurate than other assays that use three-level

A deep

LETTERS

<https://doi.org/10.1038/s41587-019-0332-7>

progeni

Avinash
Giotti^{*,†}, J.
Eric Parl
Rozenblat
2,10⁺, Jayal

1 Center for
2 Klarman
3 Harvard
4 Department
New Yo
5 Departm
Massach
6 Develop
USA
7 Broad In
8 John A.
MA, US
9 Harvard
10 Genent
11 Departm
Medical S
12 Curren

* co-first a
+ co-senior
aviv.regev

Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia

Jeffrey M. Granja^{1,2,3,13}, Sandy Klemm^{1,3,13*}, Lisa M. McGinnis^{3,4,13*}, Arwa S. Kathiria³, Anja Mezger^{3,5}, M. Ryan Corces^{1,4}, Benjamin Parks^{1,3,6}, Eric Gars⁴, Michaela Liedtke⁷, Grace X. Y. Zheng^{1,8}, Howard Y. Chang^{1,3,9,10}, Ravindra Majeti⁷ and William J. Greenleaf^{1,3,11,12*}

Identifying the causes of human diseases requires deconvolution of abnormal molecular phenotypes spanning DNA accessibility, gene expression and protein abundance^{1–3}. We present a single-cell framework that integrates highly multiplexed protein quantification, transcriptome profiling and analysis of chromatin accessibility. Using this approach, we establish a normal epigenetic baseline for healthy blood development, which we then use to deconvolve aberrant molecular features within blood from patients with mixed-phenotypic acute leukemia^{4–6}. Despite widespread epigenetic heterogeneity within the patient cohort, we observe common malignant signatures across patients as well as patient-specific regulatory features that are shared across phenotypic compartments of individual patients. Integrative analysis of transcriptomic and chromatin-accessibility maps identified 91,601 putative peak-to-gene linkages and transcription factors that regulate leukemia-specific genes, such as *RUNX1*-linked regulatory elements proximal to the marker gene *CD69*. These results demonstrate how integrative, multiomic analysis of single cells within the framework of normal development can reveal both distinct and shared molecular mechanisms of disease from patient samples.

To validate pathologic features within neoplastic cells, we first aimed to establish molecular features of normal development for comparison. As mixed-phenotype acute leukemias (MPALs) present with features of multiple hematopoietic lineages, we first constructed independent immunophenotypic, transcriptomic and epigenetic maps of normal blood development using droplet-based cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)⁷ (combined single-cell antibody-derived tag and RNA sequencing) and single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq; single-cell chromatin-accessibility profiling)⁸ on bone marrow and peripheral blood mononuclear cells (BMMCs and PBMCs, respectively; Fig. 1a). For CITE-seq analyses, we simultaneously generated 10x Genomics 3' single-cell RNA sequencing (scRNA-seq) and antibody-derived tag sequencing (scADT-seq; Supplementary Table 3) libraries from 35,882 BMMCs ($n=12,602$), CD34⁺-enriched BMMCs ($n=8,176$) and PBMCs ($n=14,804$). On average, 1,273 informative genes (2,370 unique transcript molecules) were detected per cell and replicates

were highly correlated (Supplementary Fig. 1a–c). We then selected a feature set of transcripts to mitigate batch effects and linearly projected retained transcript counts into a lower-dimensional space using latent semantic indexing^{9,10} (LSI; Methods). Cells were clustered using Seurat's shared nearest neighbor (SNN) approach¹¹, annotated using a manually curated marker gene list and visualized using uniform manifold approximation and projection (UMAP)¹² (Fig. 1b and Supplementary Fig. 1f).

We next established an epigenetic map of normal hematopoiesis by measuring chromatin accessibility across 35,038 single BMMCs ($n=16,510$), CD34⁺ BMMCs ($n=10,160$) and PBMCs ($n=8,368$) using droplet scATAC-seq (10x Genomics)⁸. These cells exhibited a canonical fragment-size distribution with clearly resolved submono- and multicellular modes, a high signal-to-noise ratio at transcription site (TSSs), an average of 11,597 uniquely accessible fragments per cell on average, a majority (61%) of Tn5 insertions aligning within peaks and high reproducibility across replicates (Supplementary Fig. 2a–h). Using LSI, Seurat's SNN clustering and UMAP, we generated a chromatin-accessibility map of hematopoiesis that complements the transcriptional map of hematopoiesis (Fig. 1c and Supplementary Fig. 2i).

To validate the proposed transcriptomic and epigenetic single-cell maps of hematopoiesis, we directly visualized lineage-restricted cell-surface marker and transcription-factor (TF) enrichment across each map. As anticipated, both scADT and scRNA-seq measurements of surface makers demonstrate *CD3D* enrichment across bone marrow and peripheral T cells; *CD14* enrichment within the monocytic lineage; broad up regulation of *CD19* across the B cell lineage; and *CDA* enrichment within cytotoxic T lymphocytes¹³ (Fig. 1d). Estimates of gene activity on the basis of correlated variation in promoter and distal-peak accessibility (Cicerio¹⁴) broadly recapitulates this pattern, confirming that lineage specification is consistently reflected across the phenotypic, transcriptional and epigenetic maps of hematopoietic development (Fig. 1d). We then visualized our scADT-seq data of BMMCs and PBMCs using UMAP and found that we could broadly recapitulate our transcriptomic hematopoietic map (Supplementary Fig. 1g,h). To further support these cell-type identifications and developmental mappings, we show concordance between three separate single-cell measurements, including direct transcript measurements from the

uno
Lai⁴,
Drit
Regev

vard

T.),

for updates

Xing²,
1,2,7[✉]

BONEMARROW + PBMC

*Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. ¹Biophysics Program, Stanford University School of Medicine, Stanford, CA, USA. ²Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. ³Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. ⁴Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. ⁵Department of Computer Science, Stanford University School of Engineering, Stanford, CA, USA. ⁶Department of Medicine, Division of Hematology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ⁷OncoGenomics, Blaeruparken, CA, USA. ⁸Department of Dermatology, Stanford University School of Medicine, Stanford, CA, USA. ⁹Department of Cell Biology, Yale University School of Medicine, New Haven, CT, USA. ¹⁰Department of Cell Biology, University of Pennsylvania, Philadelphia, PA, USA. ¹¹Department of Cell Biology, University of Michigan, Ann Arbor, MI, USA. ¹²Department of Cell Biology, University of Texas Southwestern Medical School, Dallas, TX, USA. ¹³These authors contributed equally to this work. [✉]Correspondence to: William J. Greenleaf (william.greenleaf@stanford.edu). [†]Present address: Department of Cell Biology, University of Michigan, Ann Arbor, MI, USA.

Silvia Domcke¹, Andrew J. Hi
Kimberly A. Aldinger, Dmitry
Ian A. Glass, Frank J. Steemer

INTRODUCTION: In recent years, the genomics field has made significant progress toward disentangling the complex architecture of human tissues. However, the majority of effort has been focused on cell gene expression rather than the landscape that shapes it and its expression. Toward advancing our understanding of the regulatory programs that govern human cell types, we set out to create cell atlases of both chromatin accessibility and gene expression (from a broad range of human tissues).

RATIONALE: Regions of accessible chromatin in our genome, such as enhancers, play a critical role in the determination and differentiation of cell fates. Accessible regions are enriched for genetic variation and common disease heritability. Chromatin accessibility, however, lacks single-cell resolution, which makes it difficult to infer patterns such as those most relevant to each cell type. Previously demonstrated methods of chromatin accessibility analysis involve indexing, based on two regular barcoding. Here, we describe a new assay that uses three-level

A deep

LE
<https://doi.org/10.1016/j.cell.2023.08.012>

Cell

Resource

progeni

Sin
proAvinash
Giotti^{*4}, J
Eric Parl
Rozenblat
2,10⁺, Jayan1 Center for
2 Klarman
3 Harvard
4 Department
New York
5 Department
Massachus
6 Develop
USA
7 Broad Inst
8 John A.
MA, US
9 Harvard
10 Genent
11 Depart
Medical S
12 Current
* co-first a
+ co-senior
aviv.regev@Jeffre
M. Ry
HowaIdentifi
tional ac
sibility, a
single pro
tein or chro
mone, whic
h is norma
within a k
emia⁴. The
pathways
across t
that are pa
patient sp
in acc
linkage s
specific p
proxim
how in
the fra
distinct
patientTo id
aimed t
compari
ment w
structured
epigenetic
cellular
ing (CT
RNA se
chromati
accessibi
monoch
CITE-se
single-o
tag sequ
35,882 I
and PBM
unique |

A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates

for updates

Authors

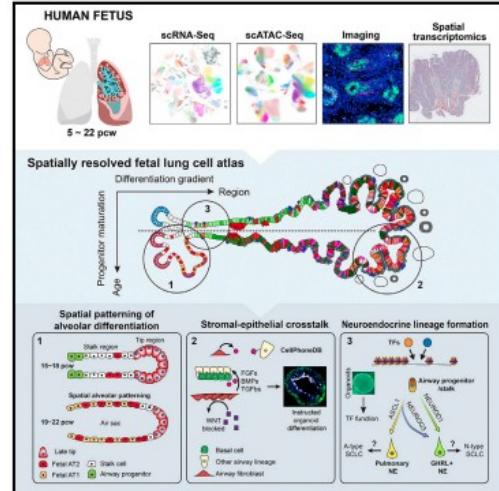
Peng He, Kyungtae Lim, Dawei Sun, ...,
John C. Marioni, Kerstin B. Meyer,
Emma L. Rawlins

Correspondence

e.rawlins@gurdon.cam.ac.uk

In brief

Xing²,
1,2,7
Multiomic analysis of human fetal lungs from 5–22 post-conception weeks unveils cell-lineage trajectories across different cell types during development and will provide fresh insights into lung disease progression in adults.



Highlights

- Spatiotemporal atlas of human lung development identifies 144 cell types/states
- Tracking the developmental origins of multiple cell types, including new progenitors
- Functional diversity of fibroblasts in distinct anatomical signaling niches
- Experimental validation of TFs controlling neuroendocrine cell heterogeneity

¹Center for Medicine Universit

²Depart Stanford

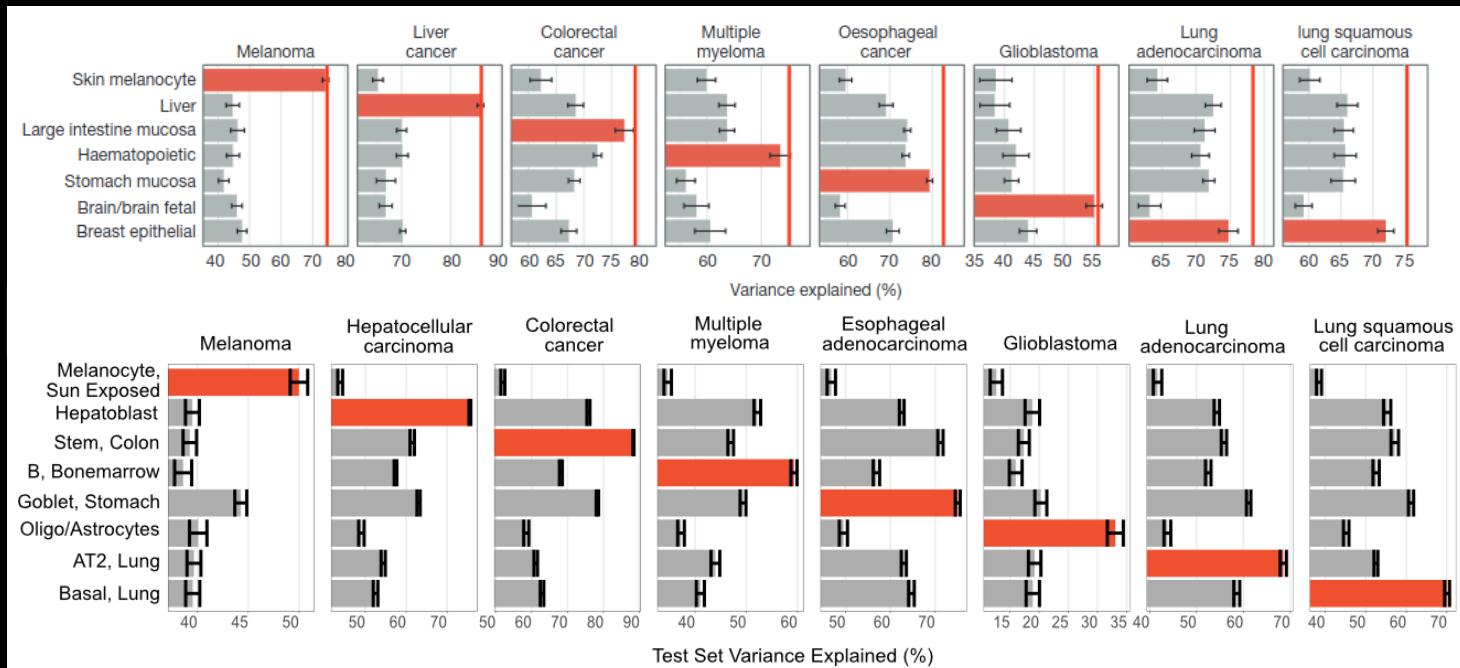
FETAL LUNG ATLAS

- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

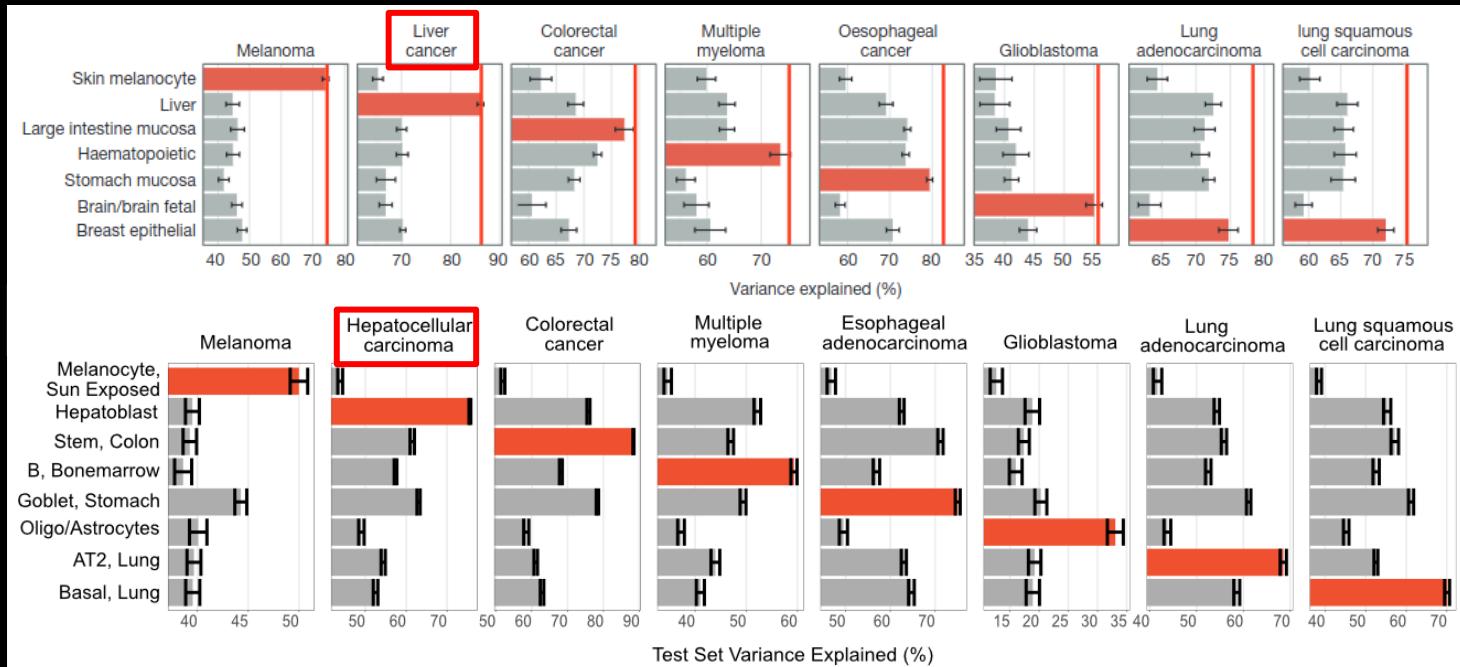
scATAC-seq
features



- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

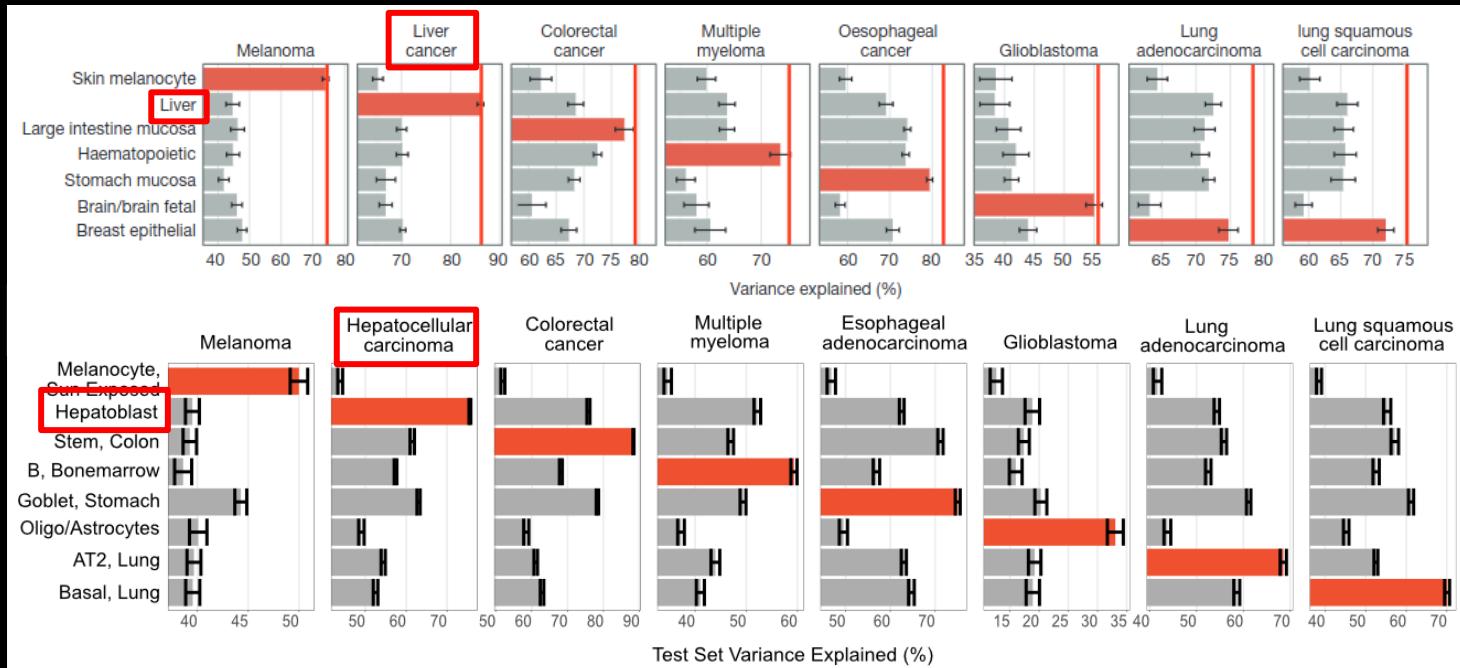
scATAC-seq
features



- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

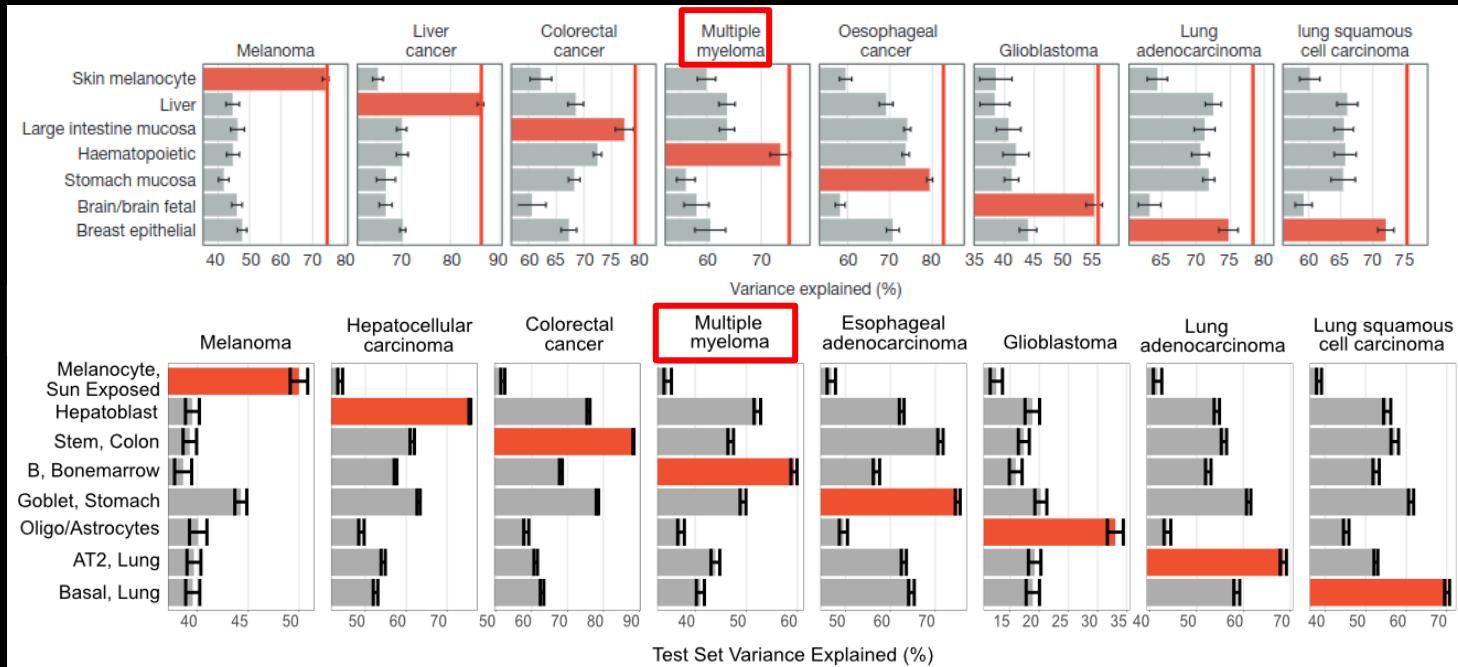
scATAC-seq
features



- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

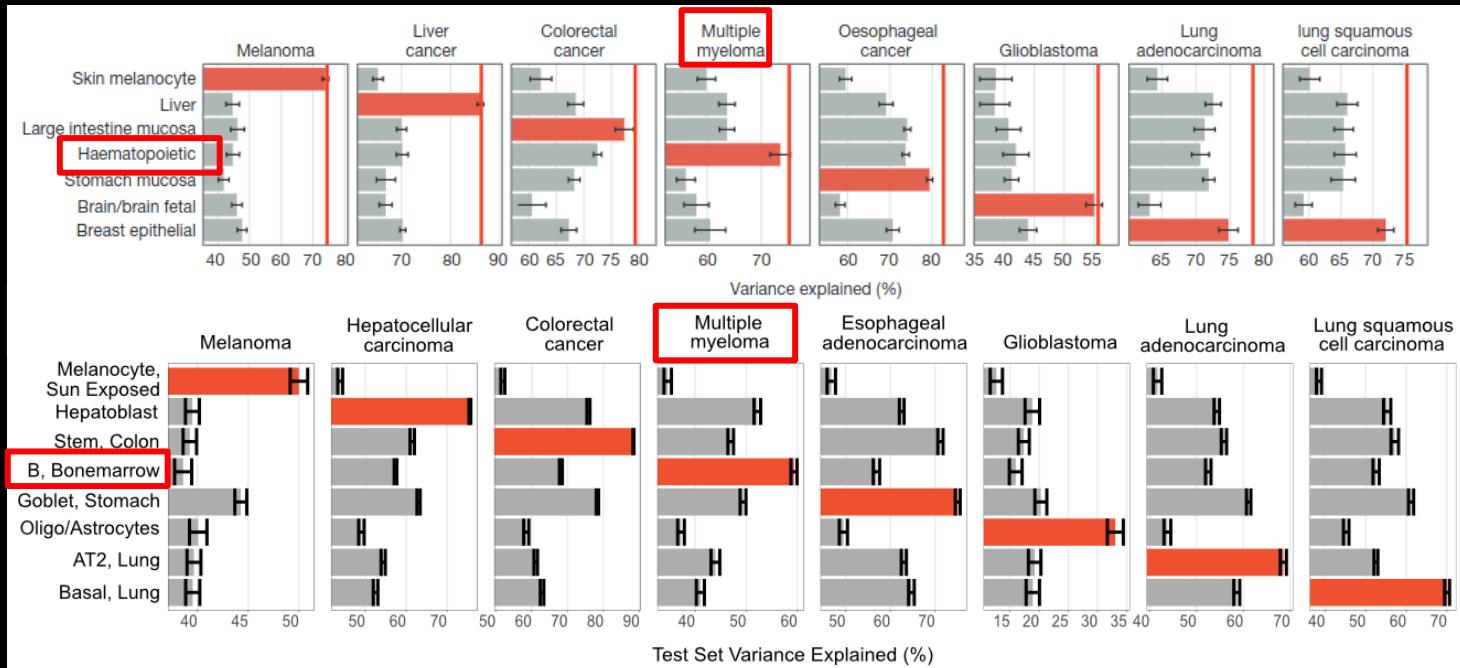
scATAC-seq
features



- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

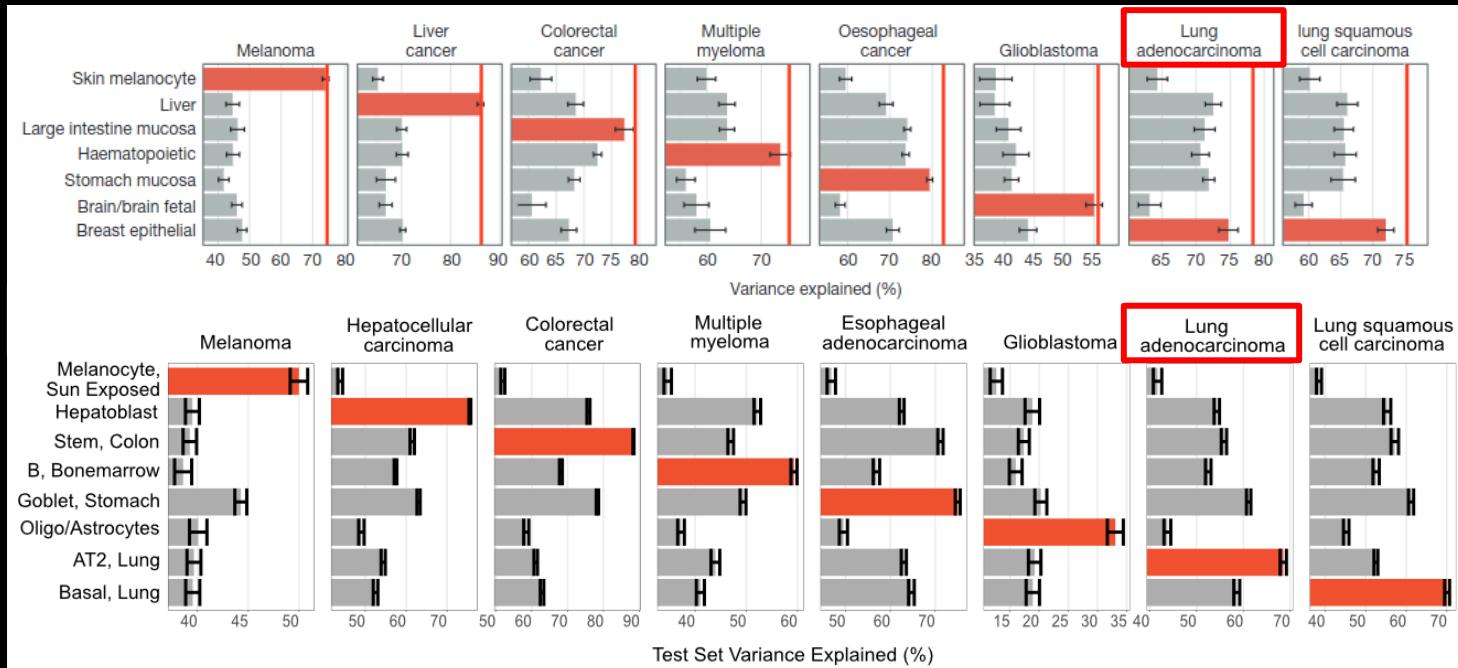
scATAC-seq
features



- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

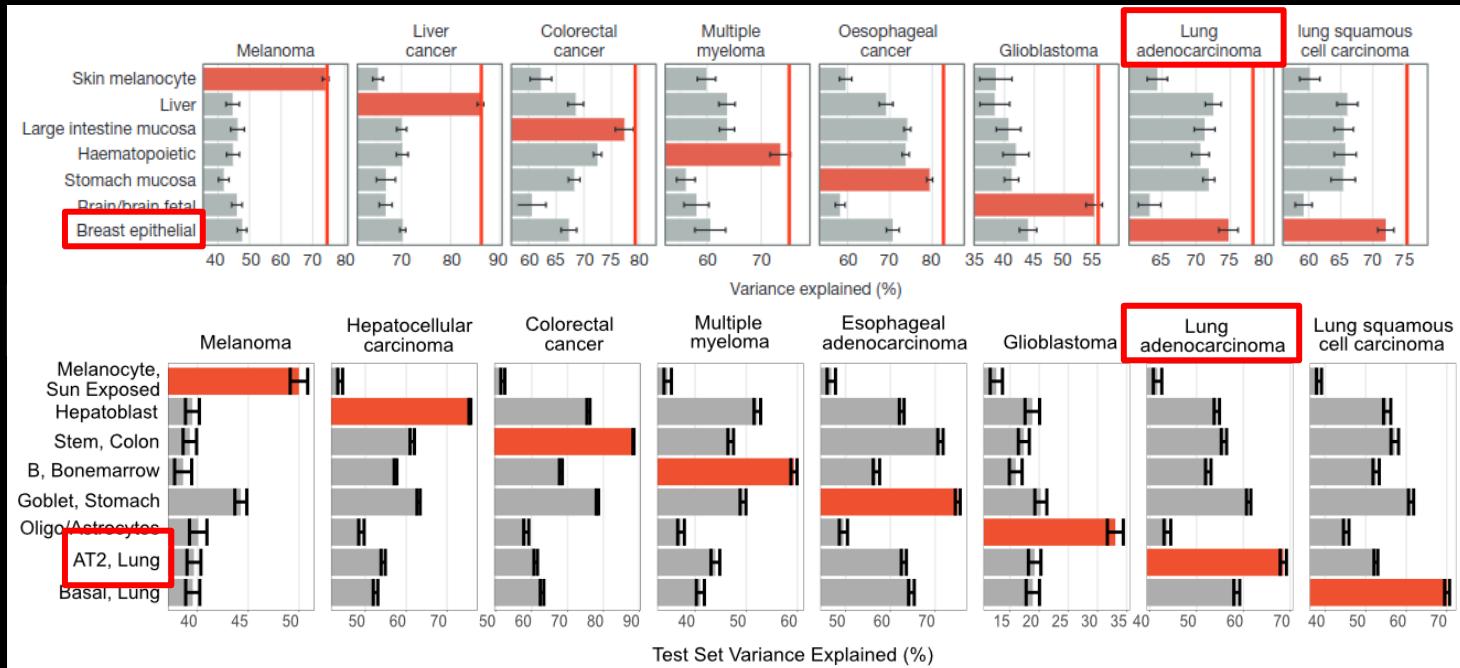
scATAC-seq
features



- scATAC reproduce results from bulk chromatin modification data while providing higher specificity

bulk
chromatin
features
(Polak, Karlic, et al.
Nature 2015)

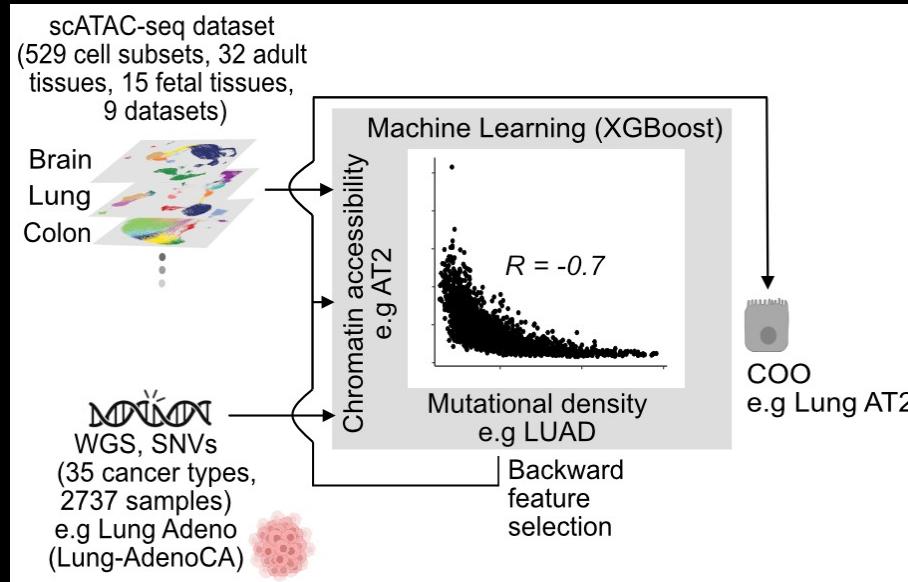
scATAC-seq
features



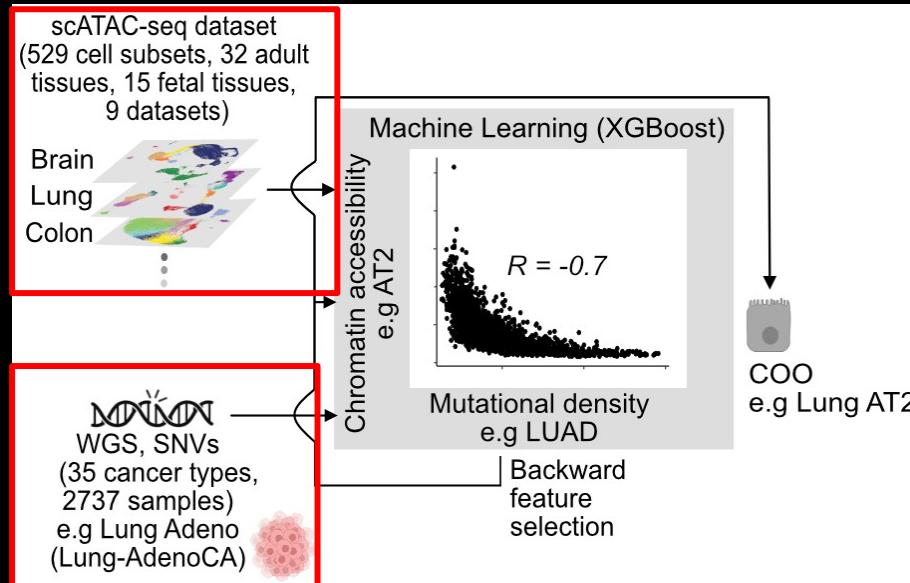
- In total, we have predictions for 35 cancer types / subtypes
- Not the first work to extend to other cancer types, but first to do it using single cell data, **enabling unprecedented specificity in predictions**

- In total, we have predictions for 35 cancer types / subtypes
- Not the first work to extend to other cancer types, but first to do it using single cell data, **enabling unprecedented specificity in predictions**
- **Assembled one of the most extensive scATAC datasets to date**

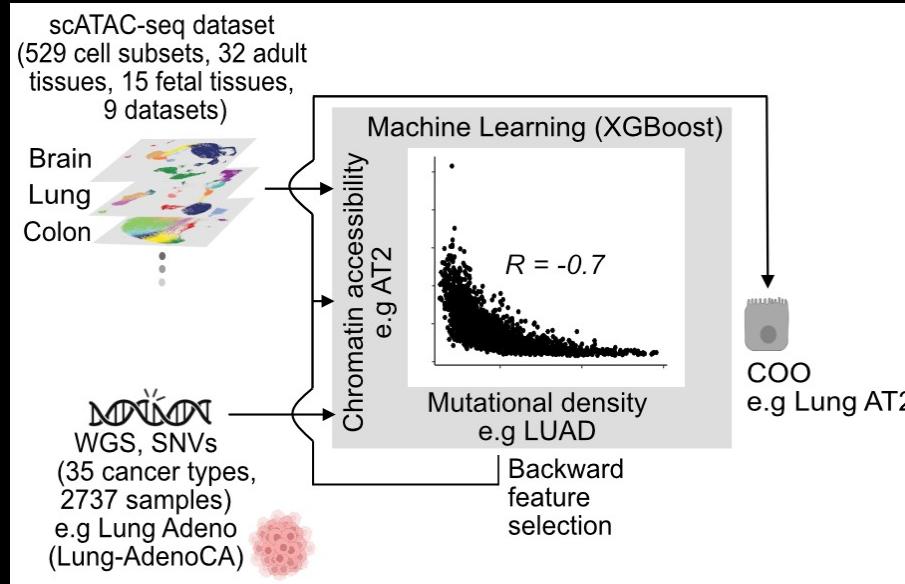
- In total, we have predictions for 35 cancer types / subtypes
- Not the first work to extend to other cancer types, but first to do it using single cell data, **enabling unprecedented specificity in predictions**
- **Assembled one of the most extensive scATAC datasets to date**



- In total, we have predictions for 35 cancer types / subtypes
- Not the first work to extend to other cancer types, but first to do it using single cell data, **enabling unprecedented specificity in predictions**
- **Assembled one of the most extensive scATAC datasets to date**

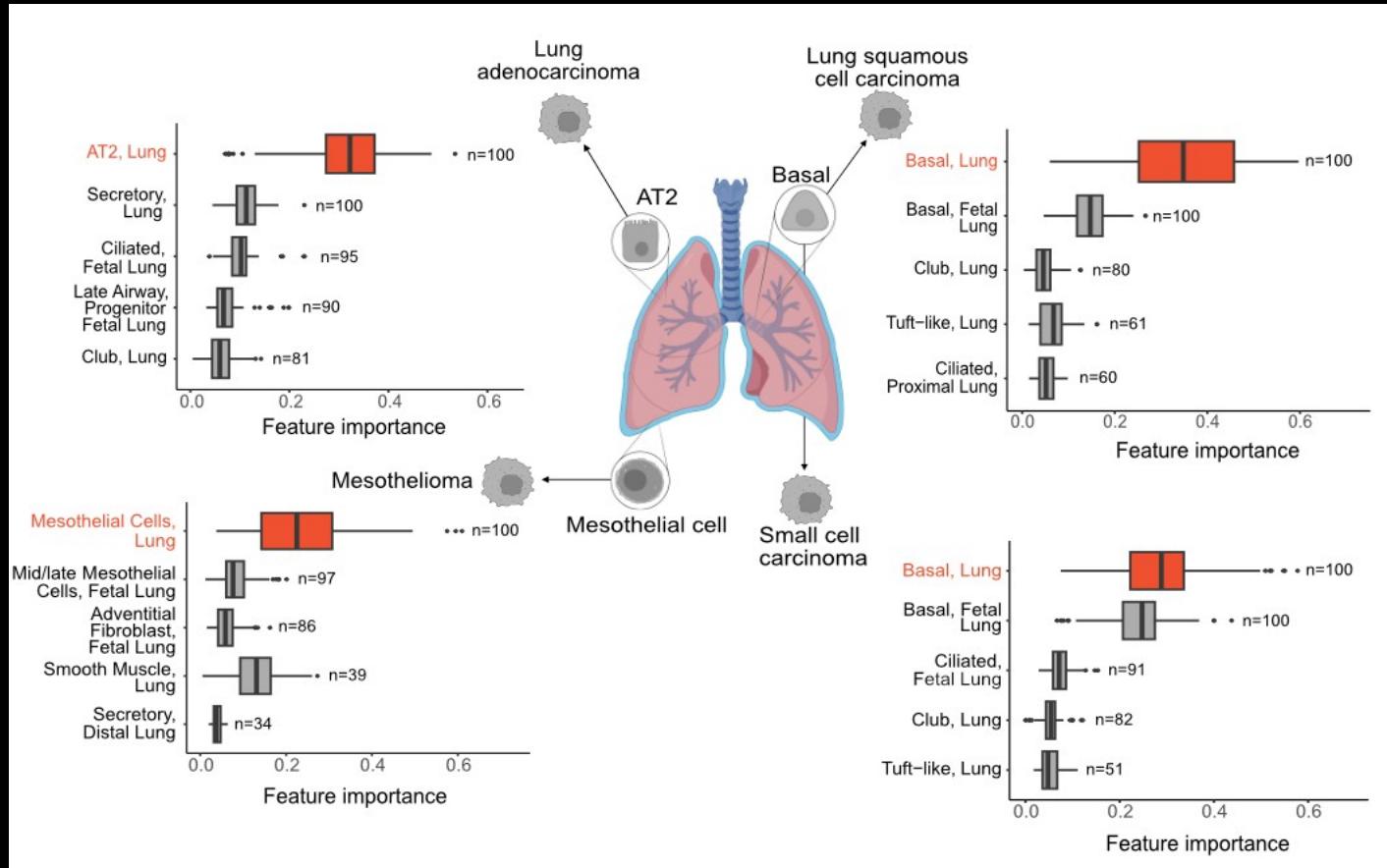


- In total, we have predictions for 35 cancer types / subtypes
- Not the first work to extend to other cancer types, but first to do it using single cell data, **enabling unprecedented specificity in predictions**
- **Assembled one of the most extensive scATAC datasets to date**

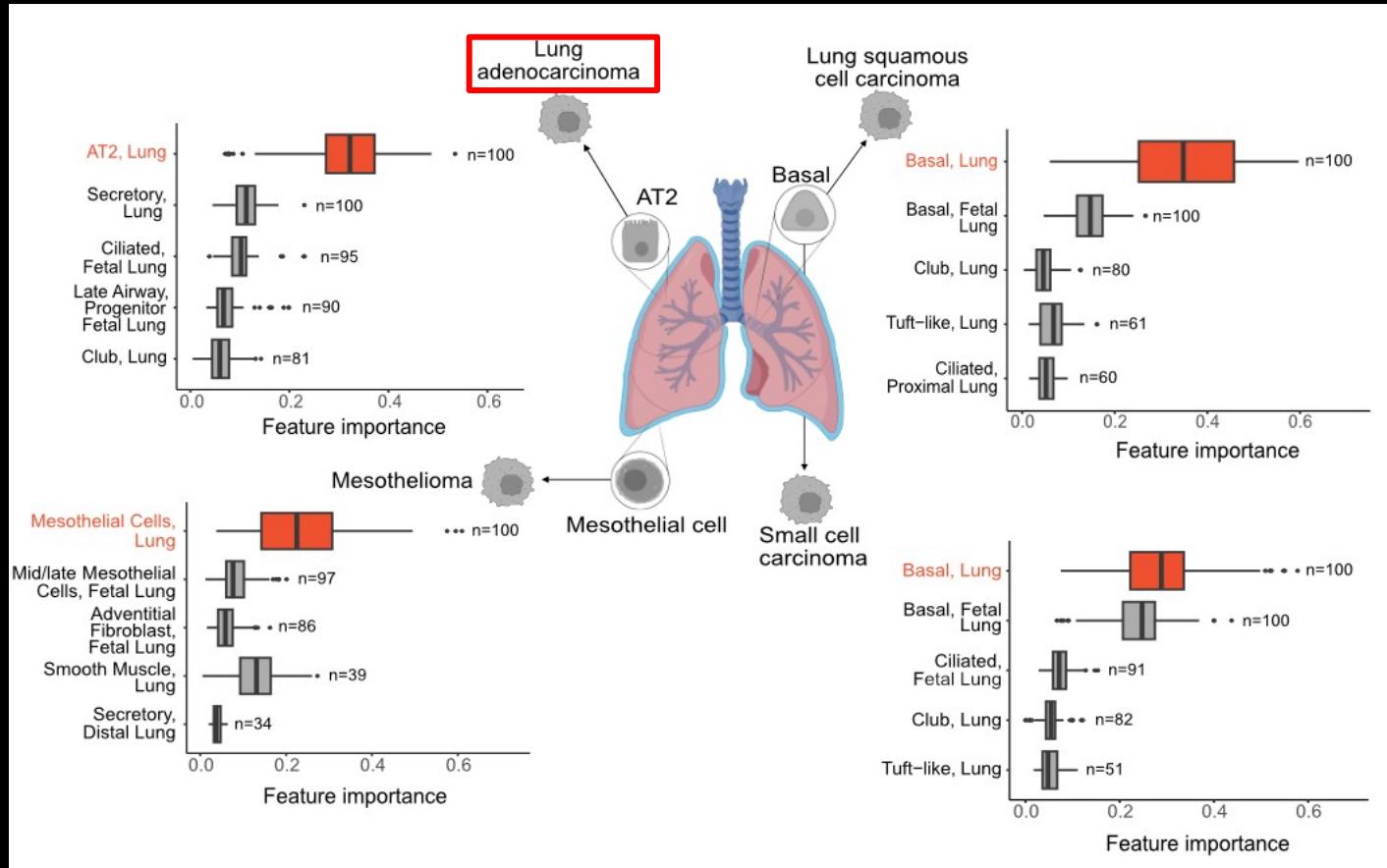


- Some results will be highlighted next

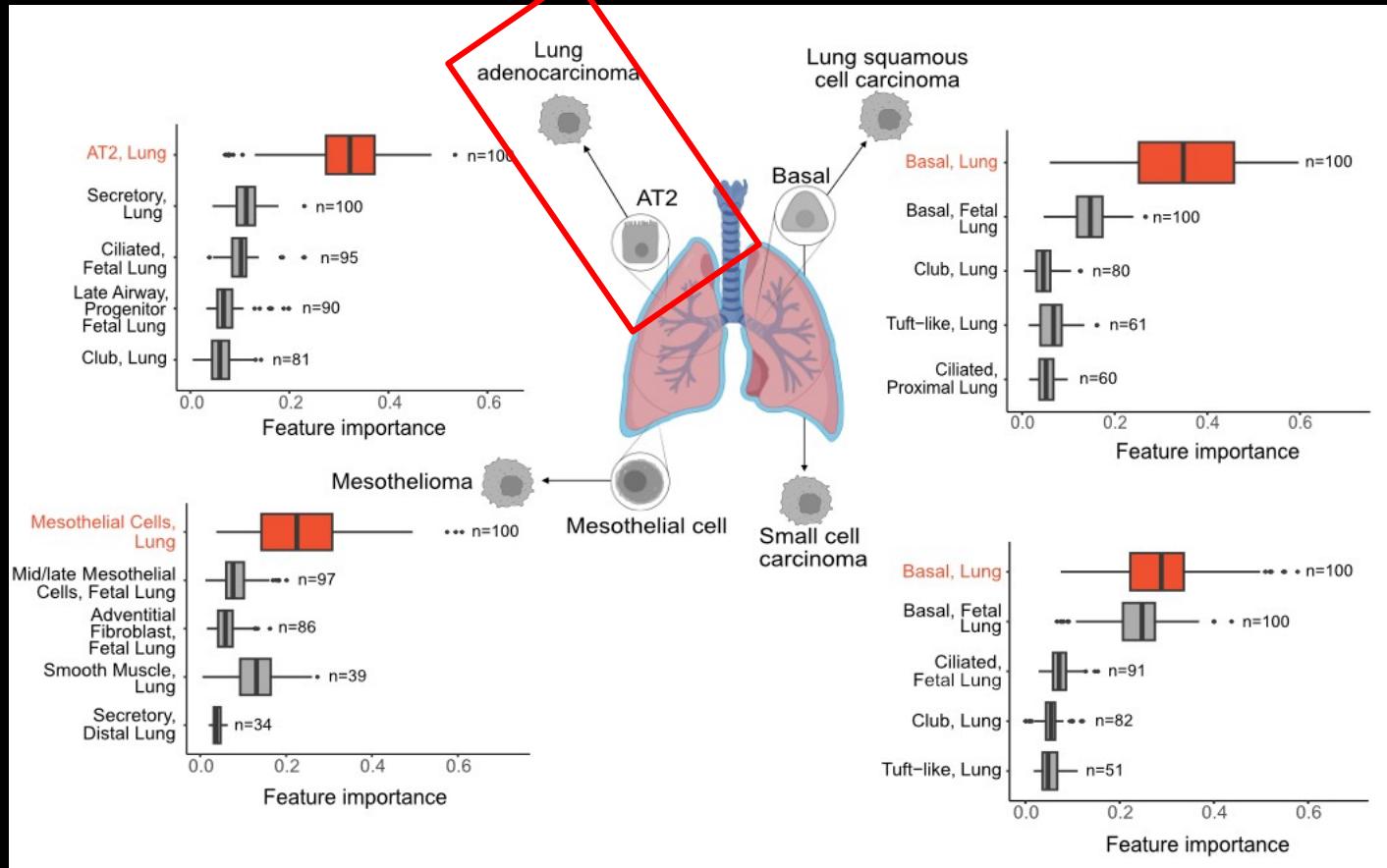
- Lung-associated malignancies have different COOs



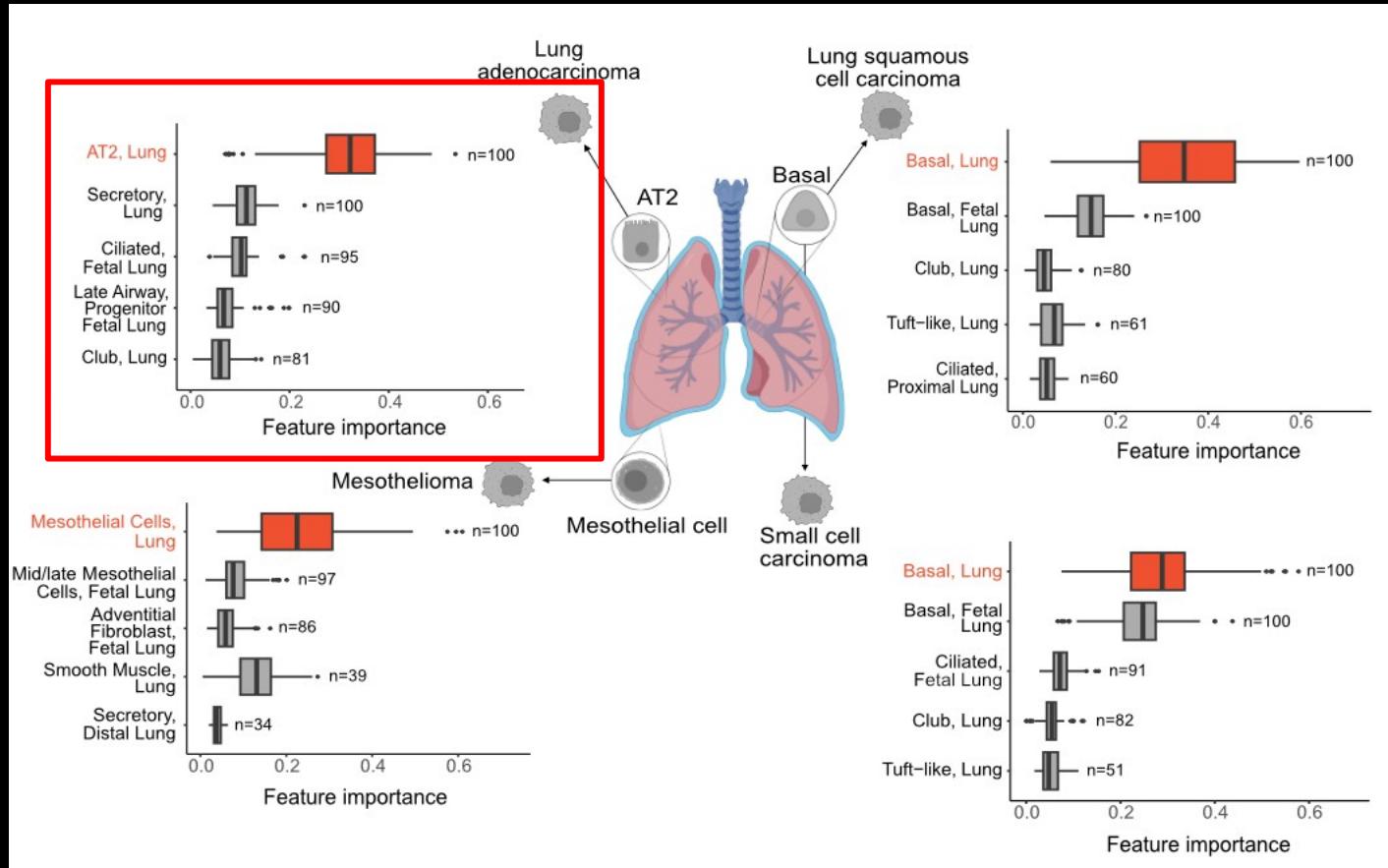
- Lung-associated malignancies have different COOs



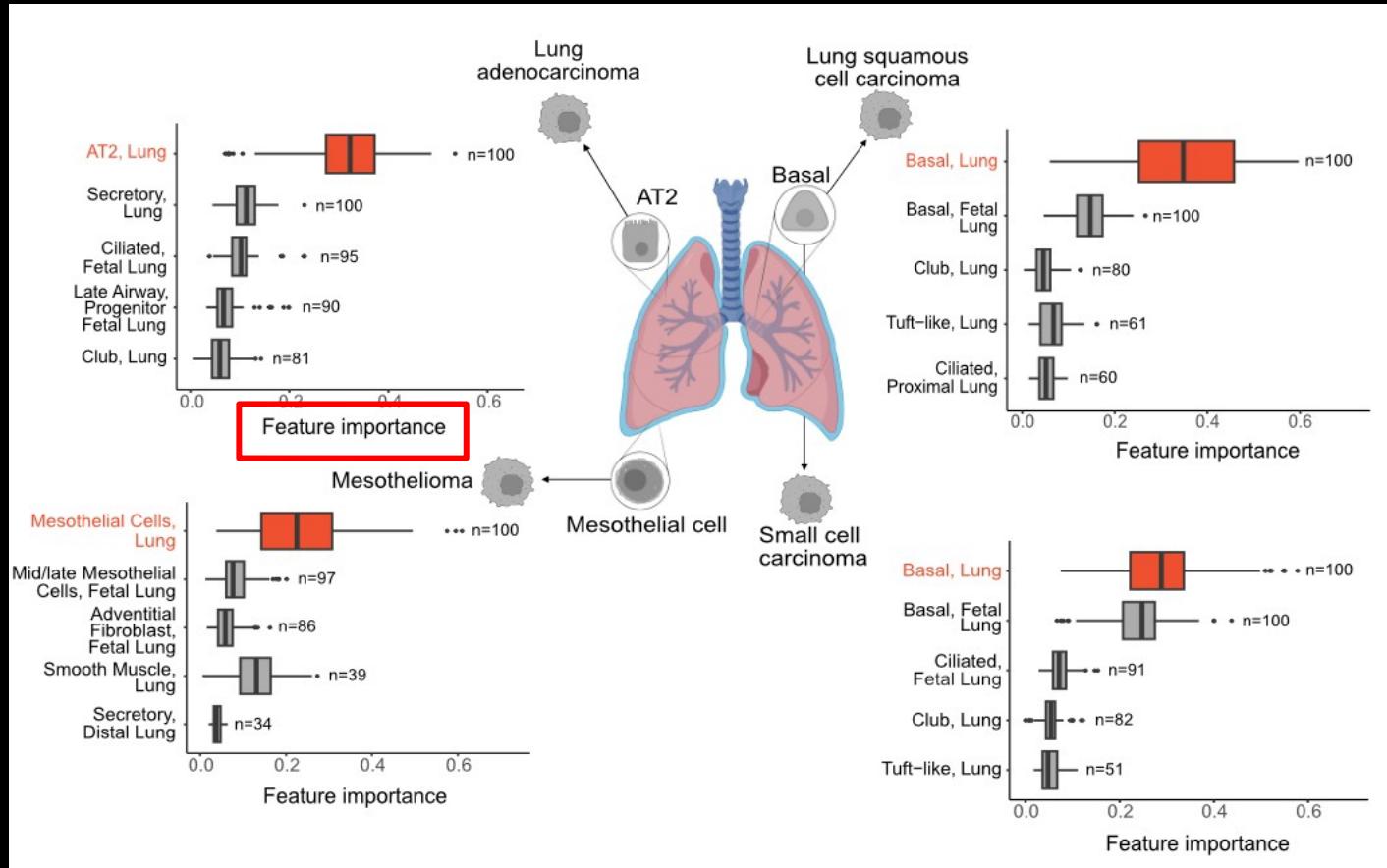
- Lung-associated malignancies have different COOs



- Lung-associated malignancies have different COOs

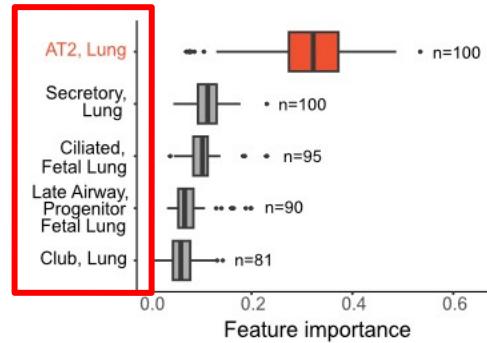


- Lung-associated malignancies have different COOs

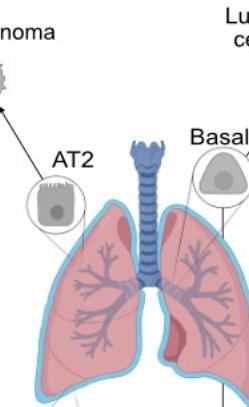


- Lung-associated malignancies have different COOs

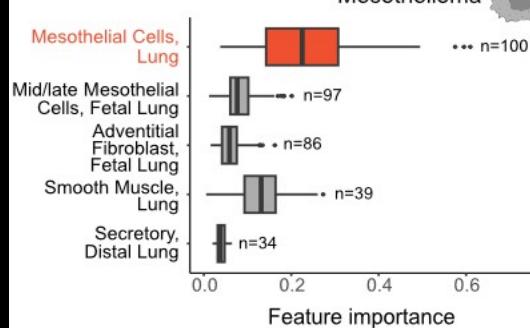
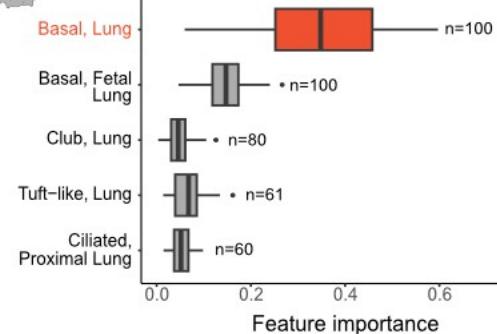
Top 5 features at iteration 15 of backward feature selection across 100 runs of the model



Lung adenocarcinoma

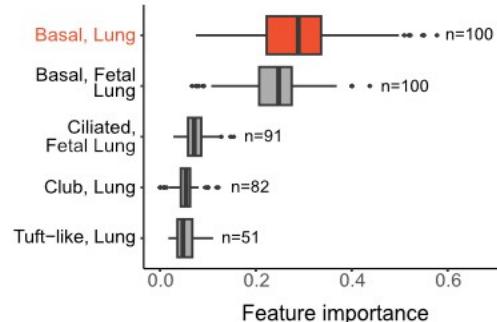


Lung squamous cell carcinoma

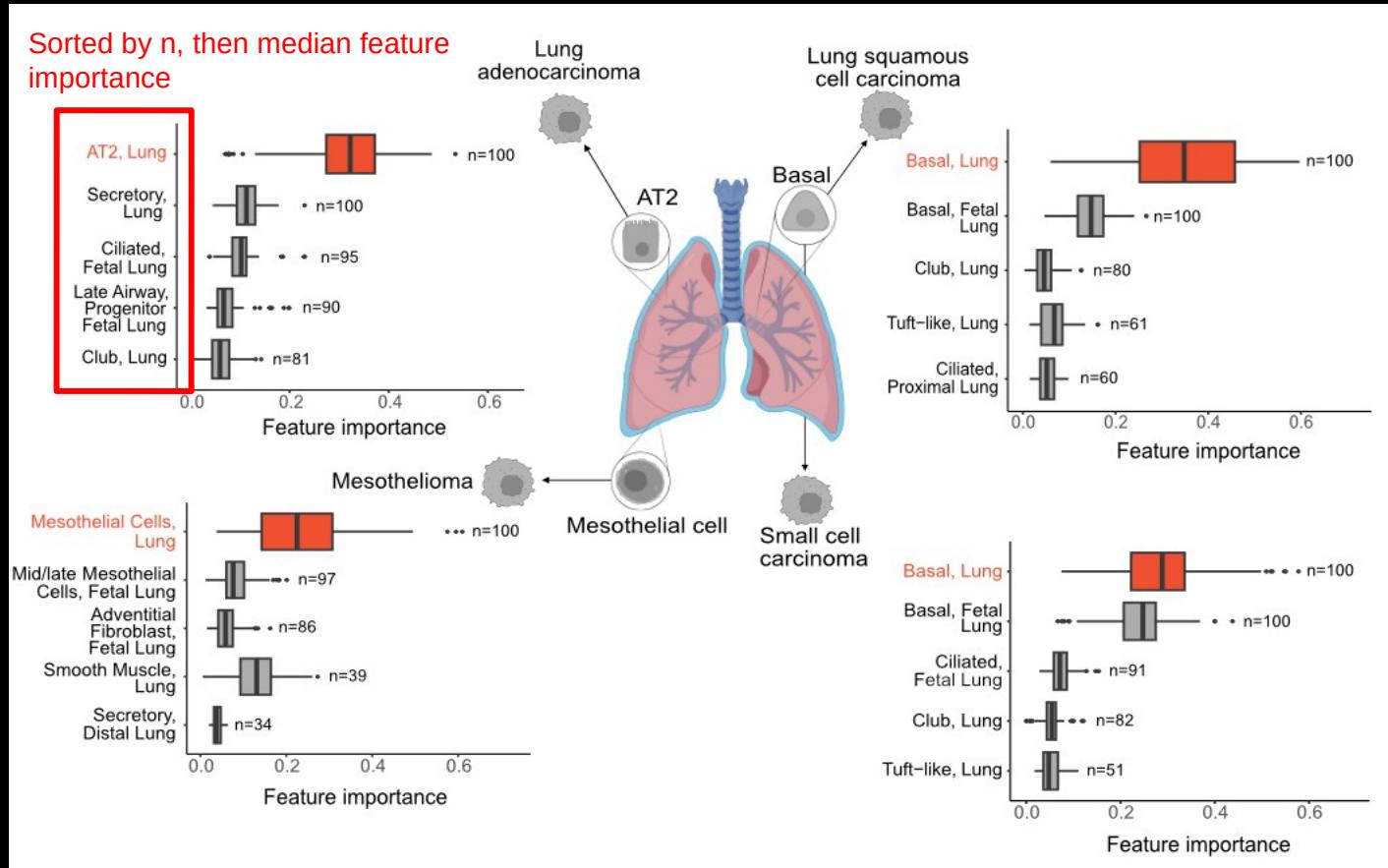


Mesothelioma

Mesothelial cell
Small cell carcinoma

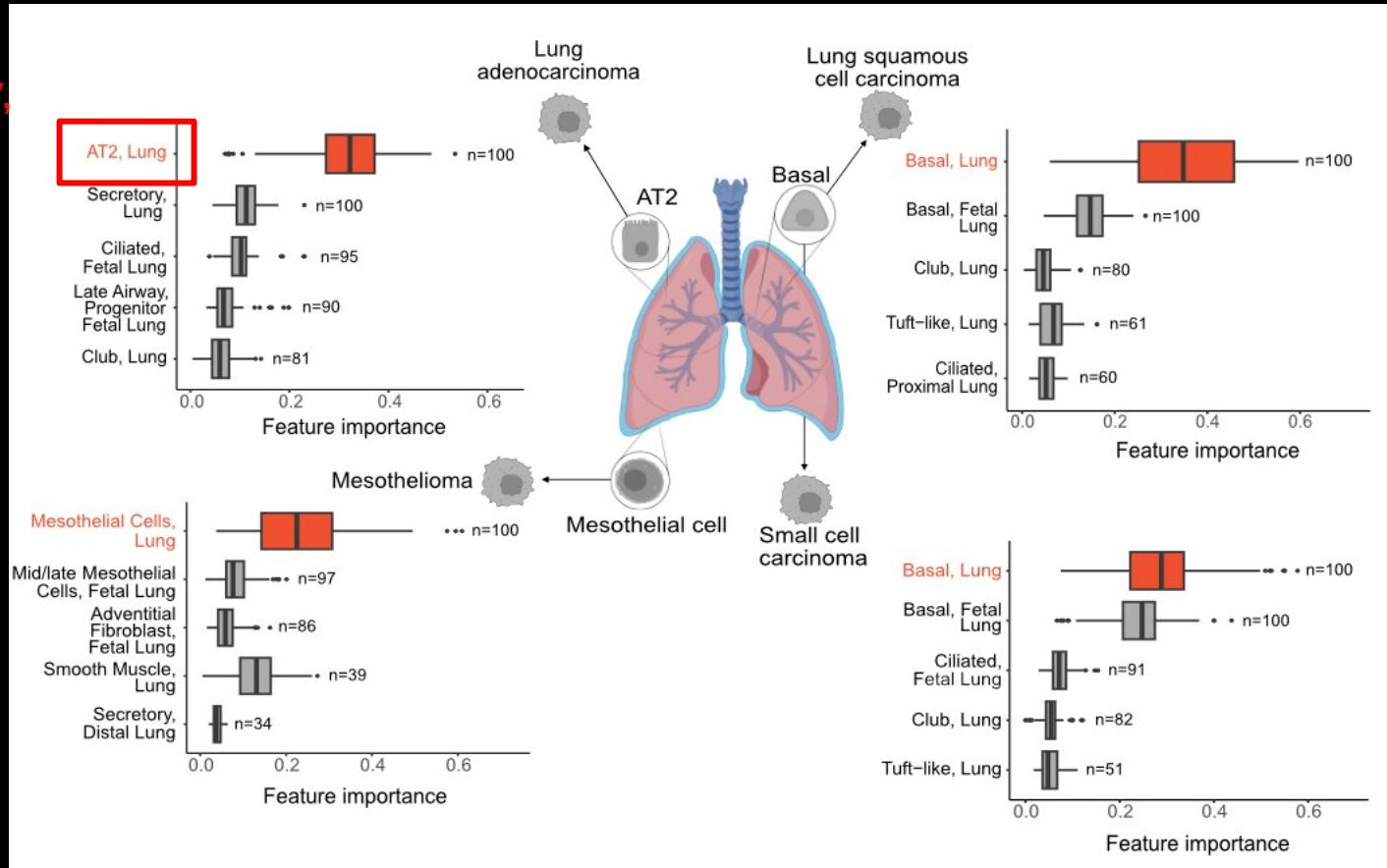


- Lung-associated malignancies have different COOs

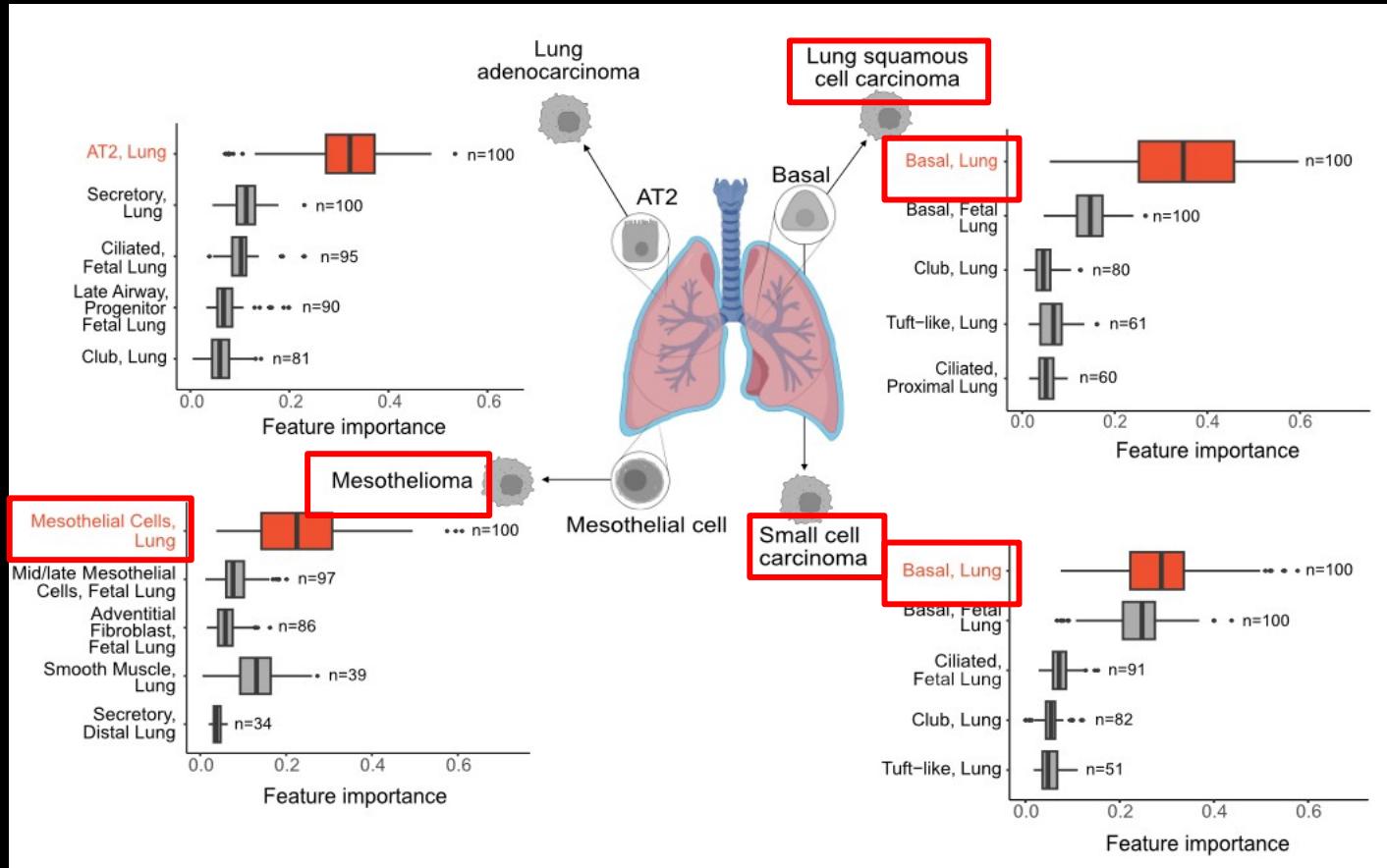


- Lung-associated malignancies have different COOs

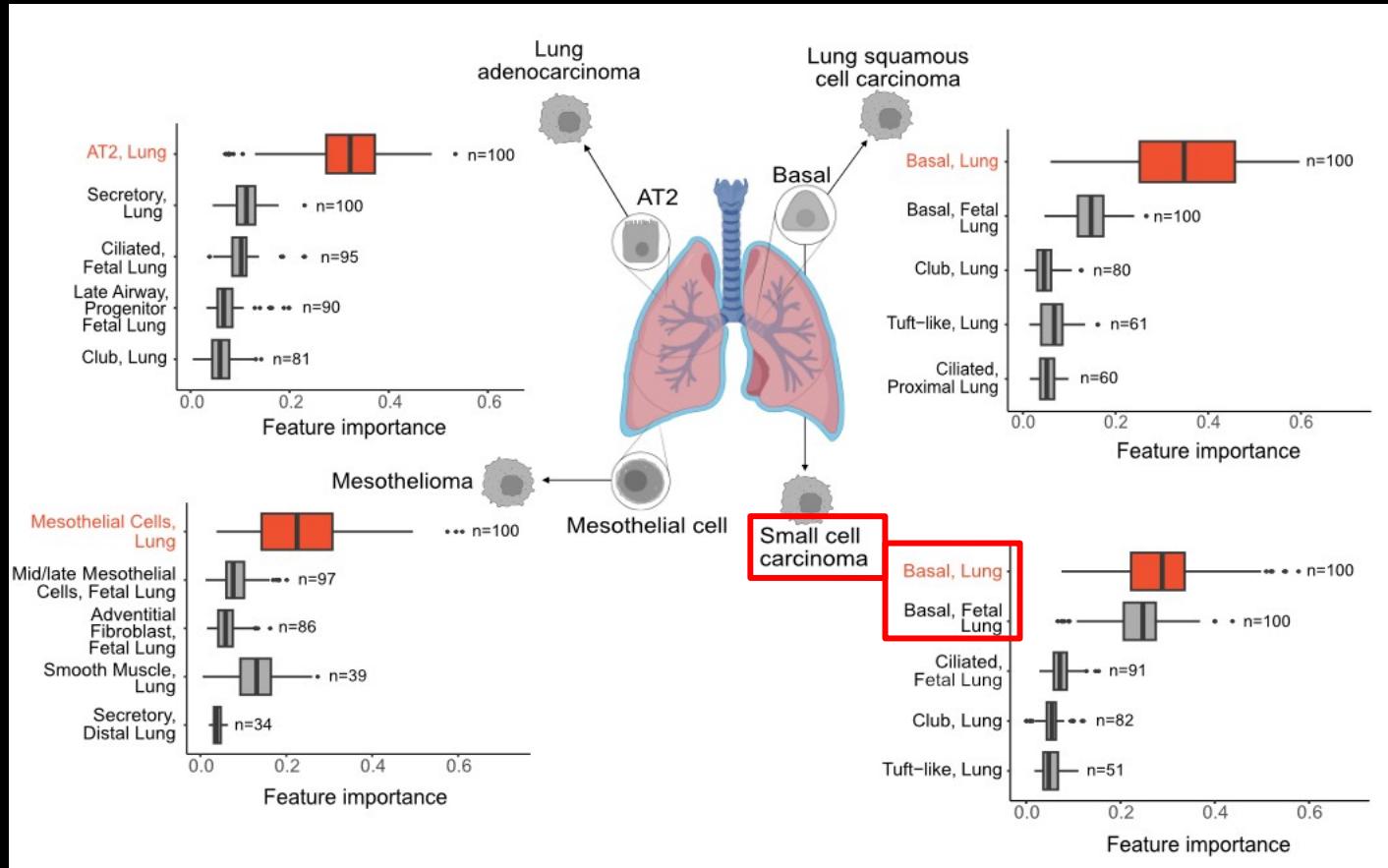
Presumably,
the COO



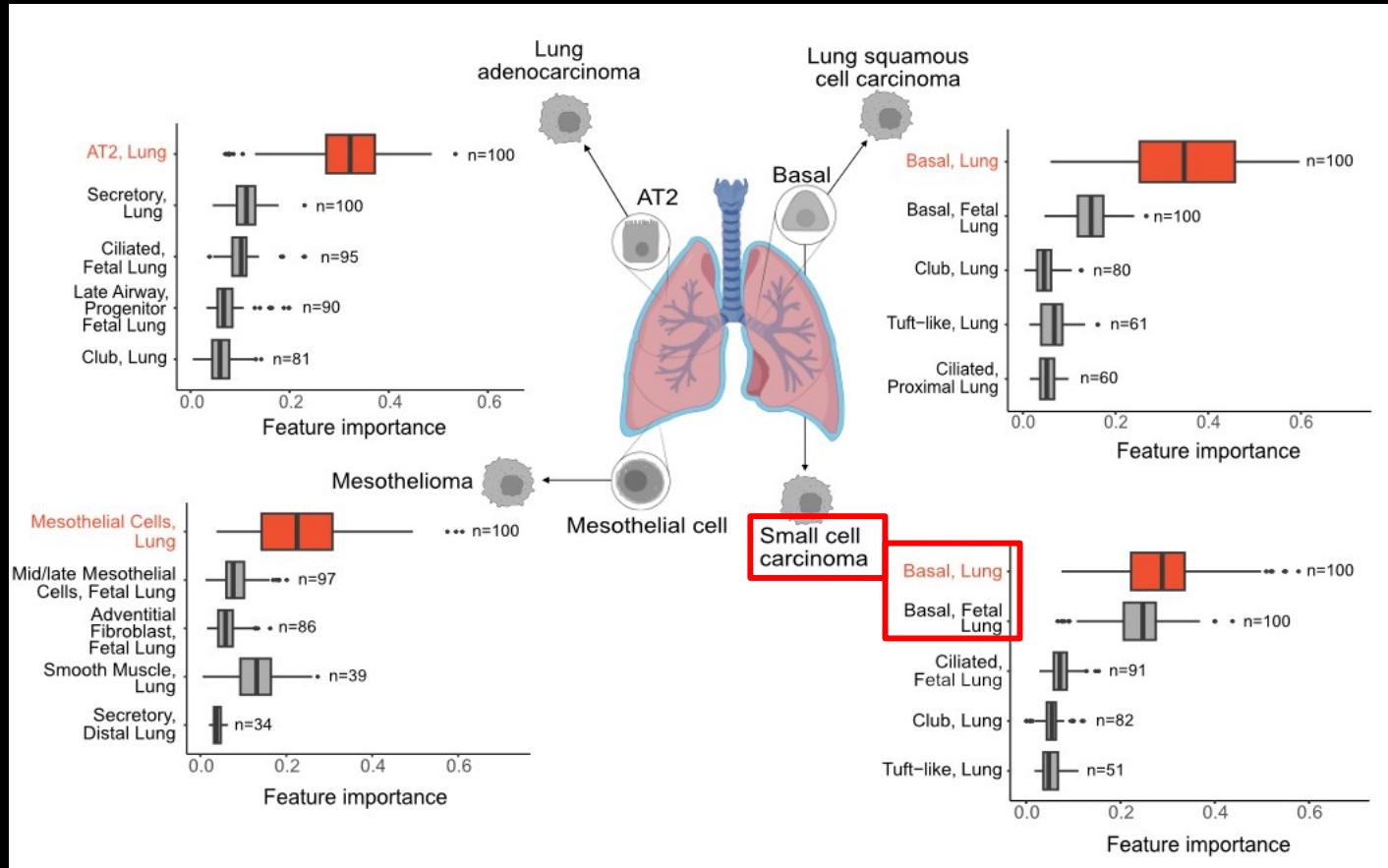
- Lung-associated malignancies have different COOs



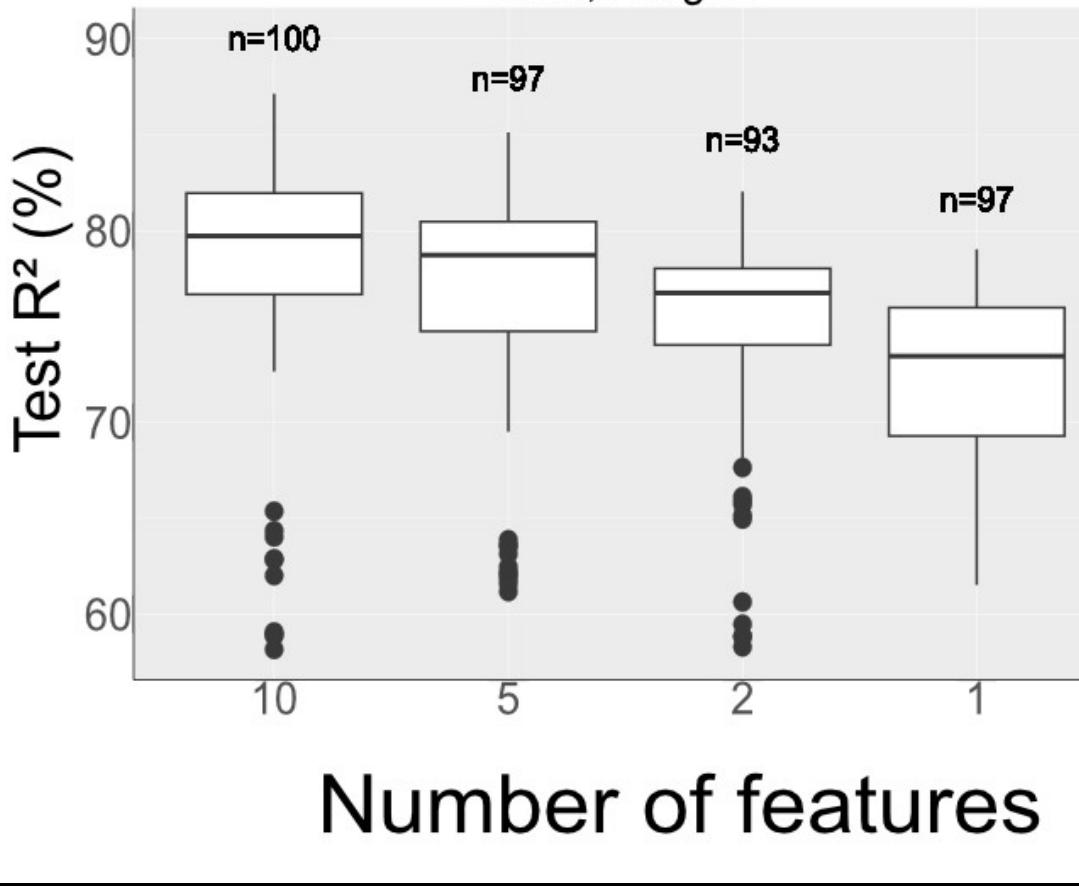
- Lung-associated malignancies have different COOs



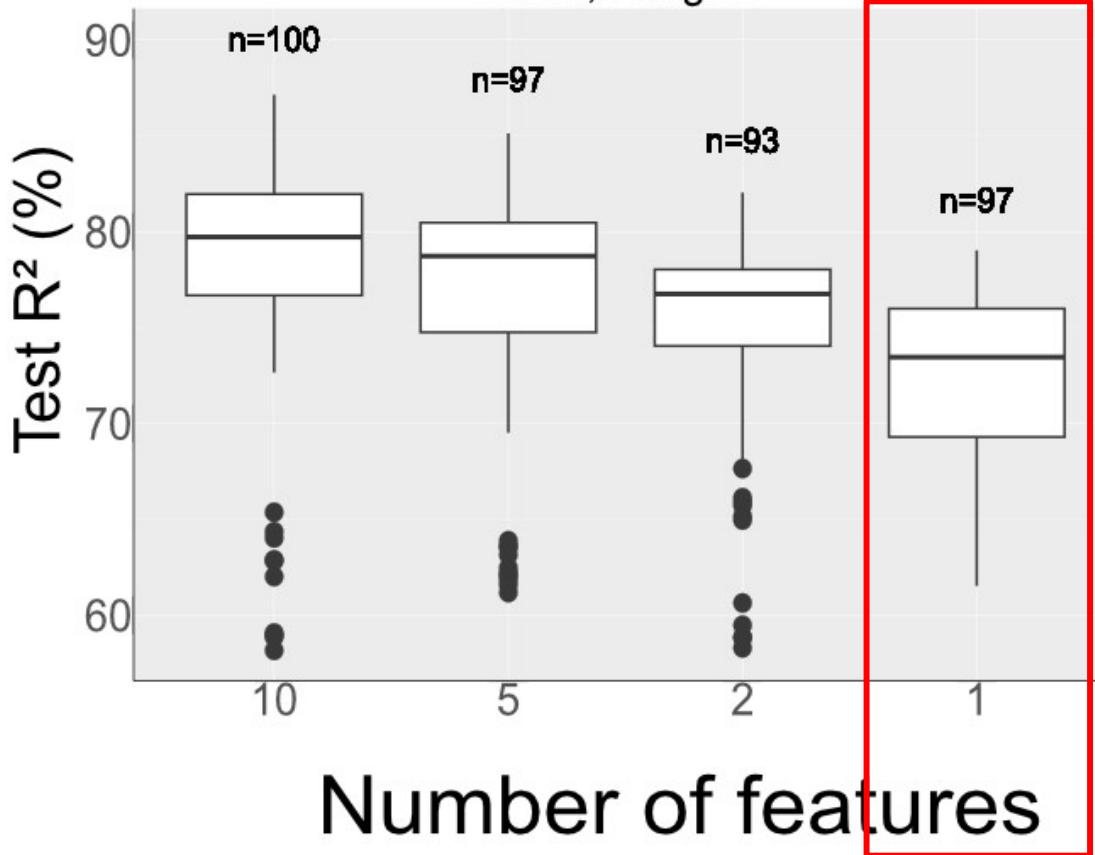
- Lung-associated malignancies have different COOs



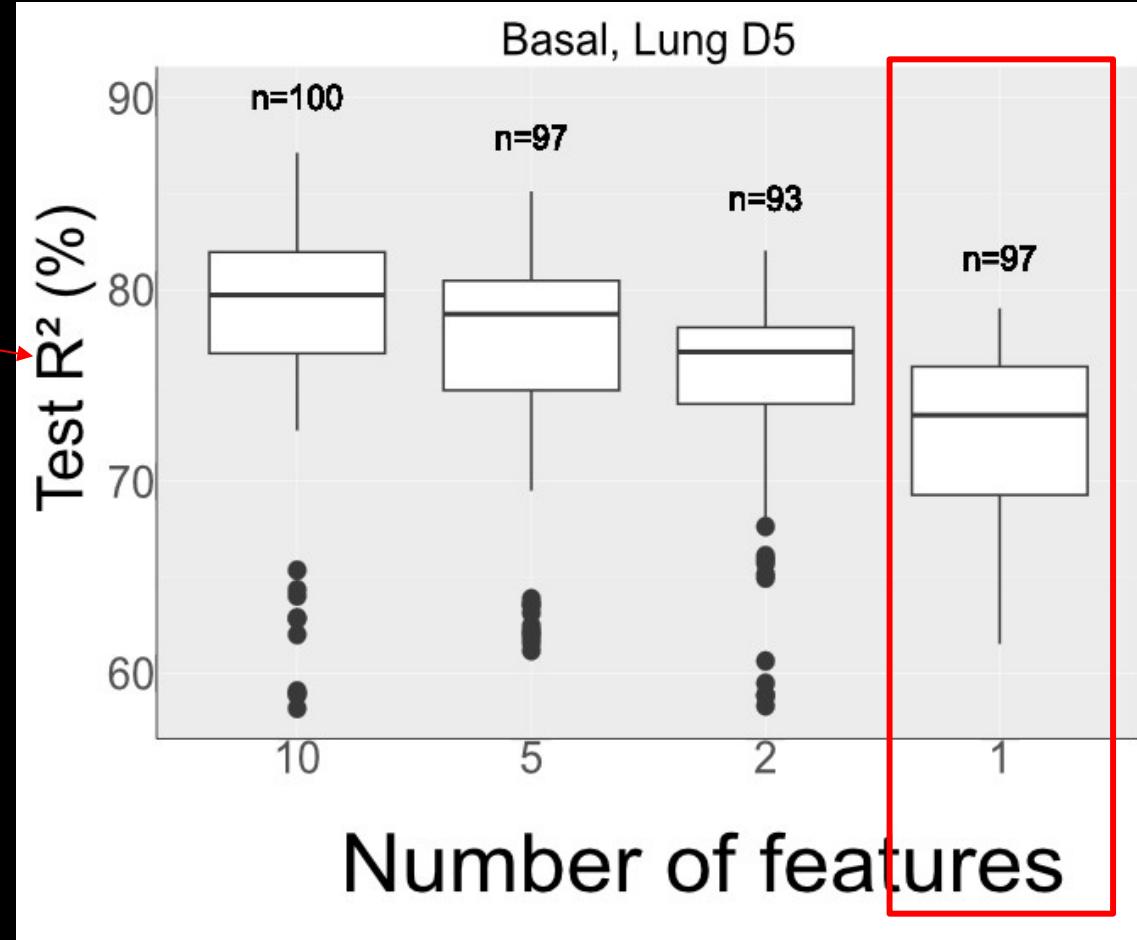
Basal, Lung D5



Basal, Lung D5

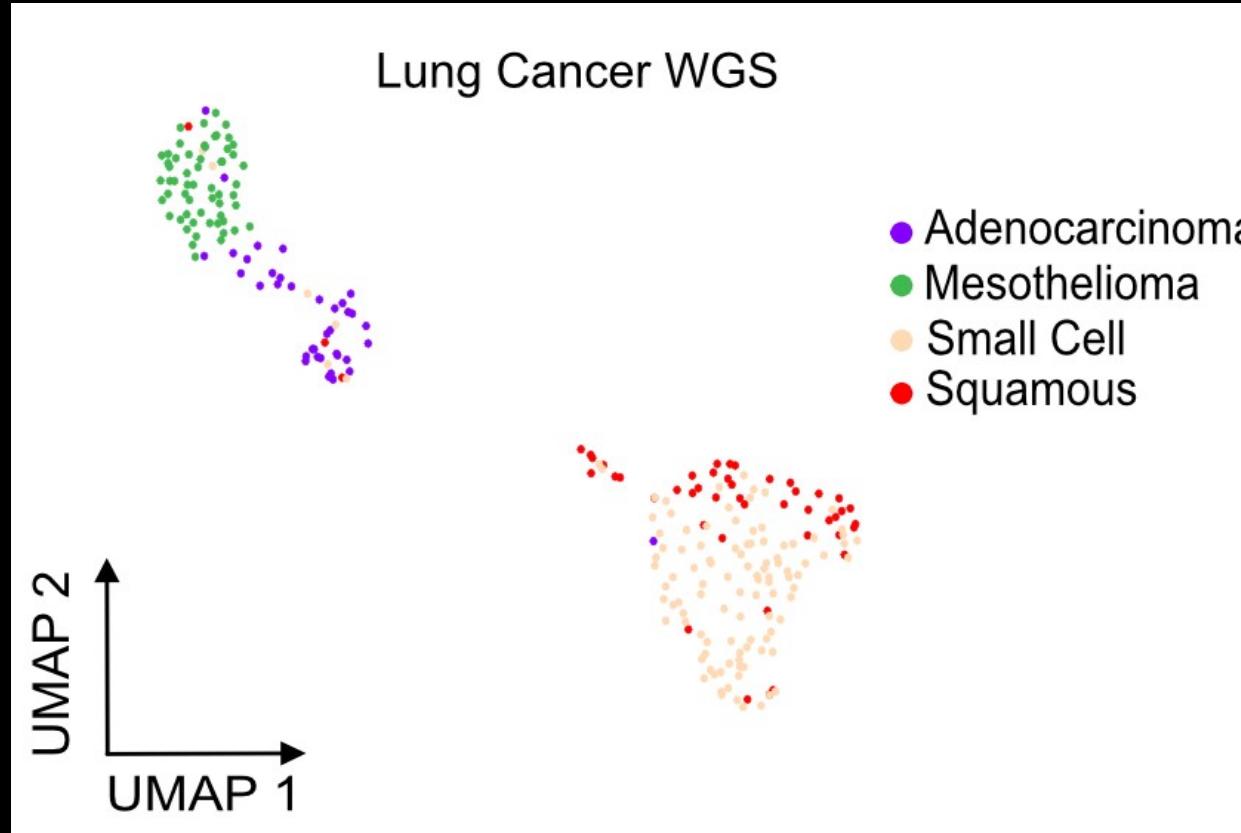


Aka variance explained

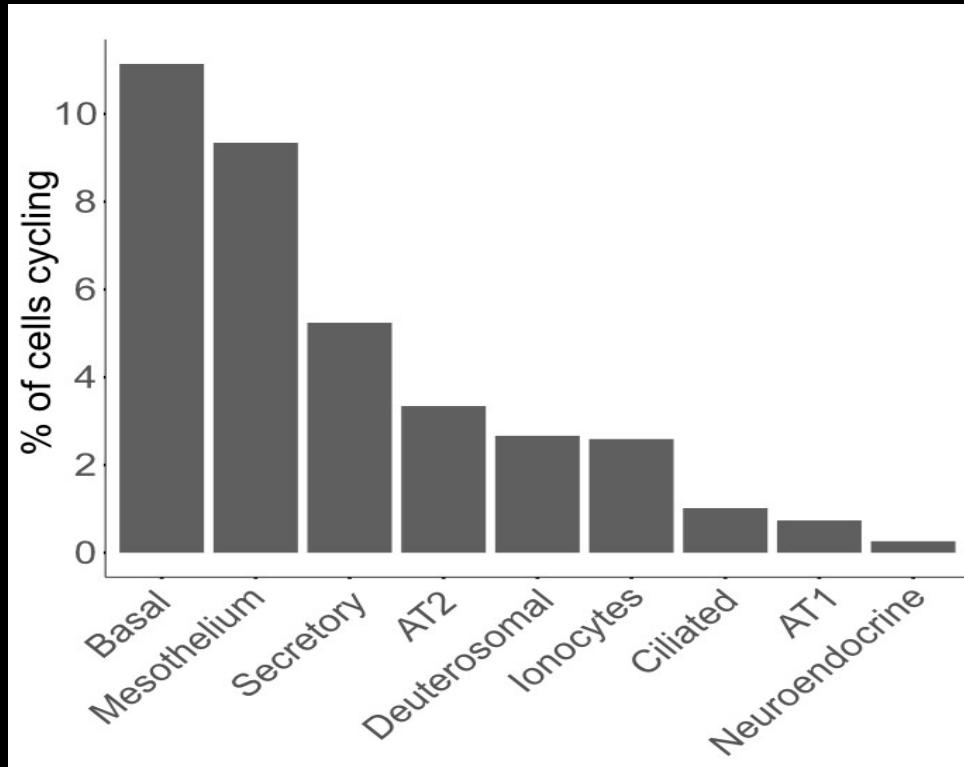


- Inherent similarity in mutational density between squamous cell carcinoma and small cell lung cancer

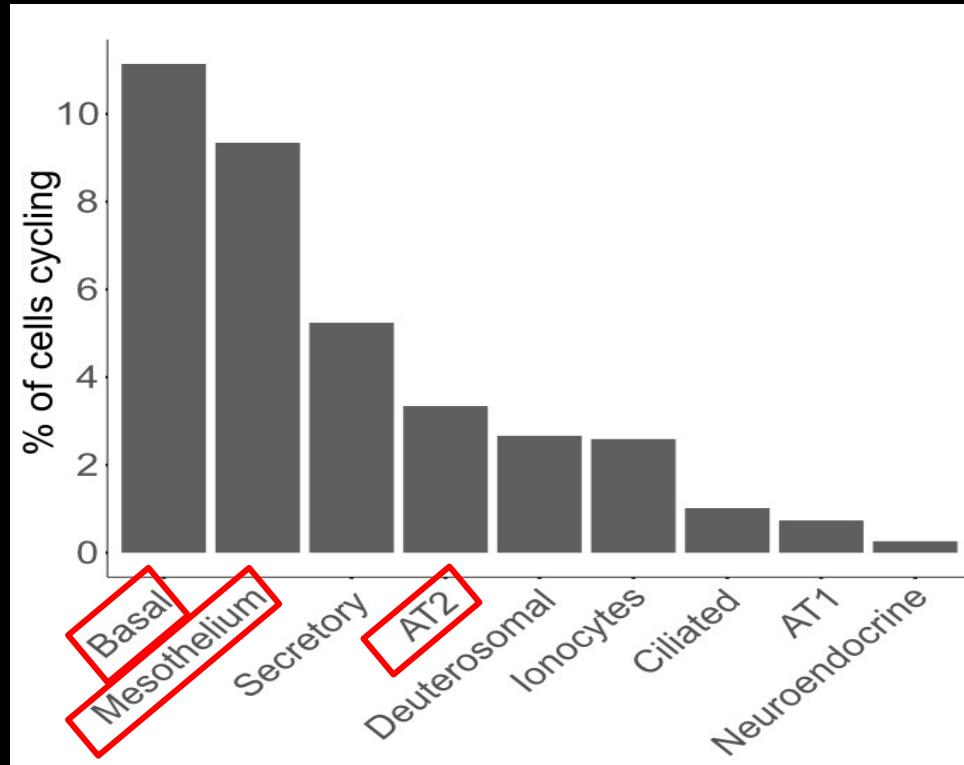
- Inherent similarity in mutational density between squamous cell carcinoma and small cell lung cancer



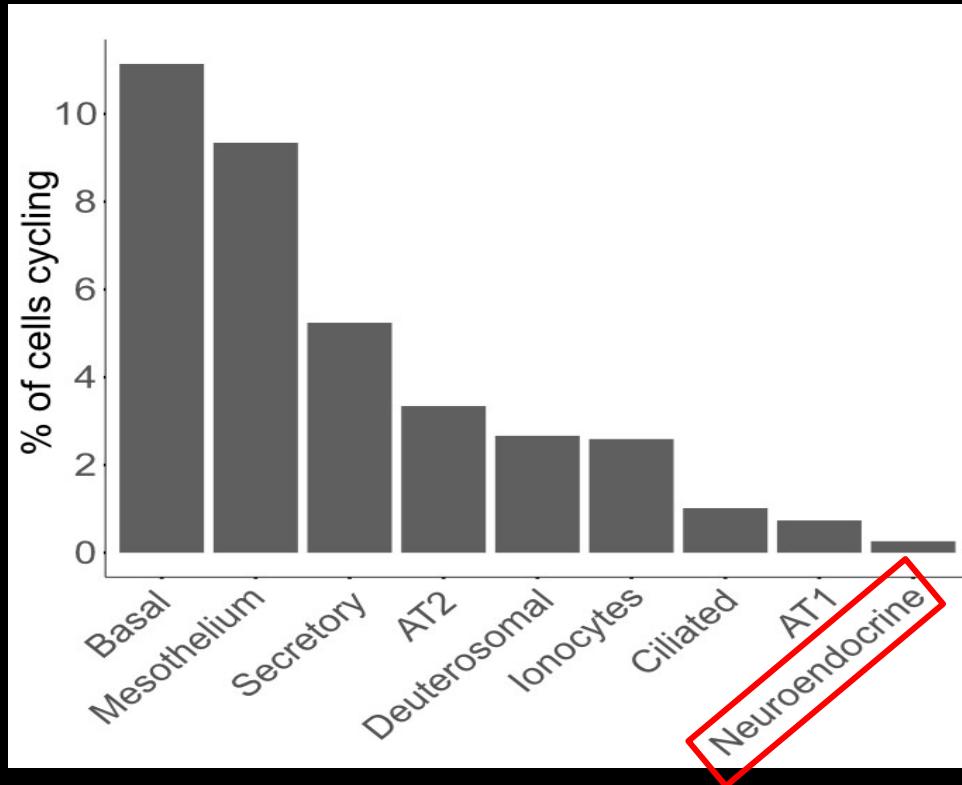
- Lung COOs are highly proliferative in homeostasis



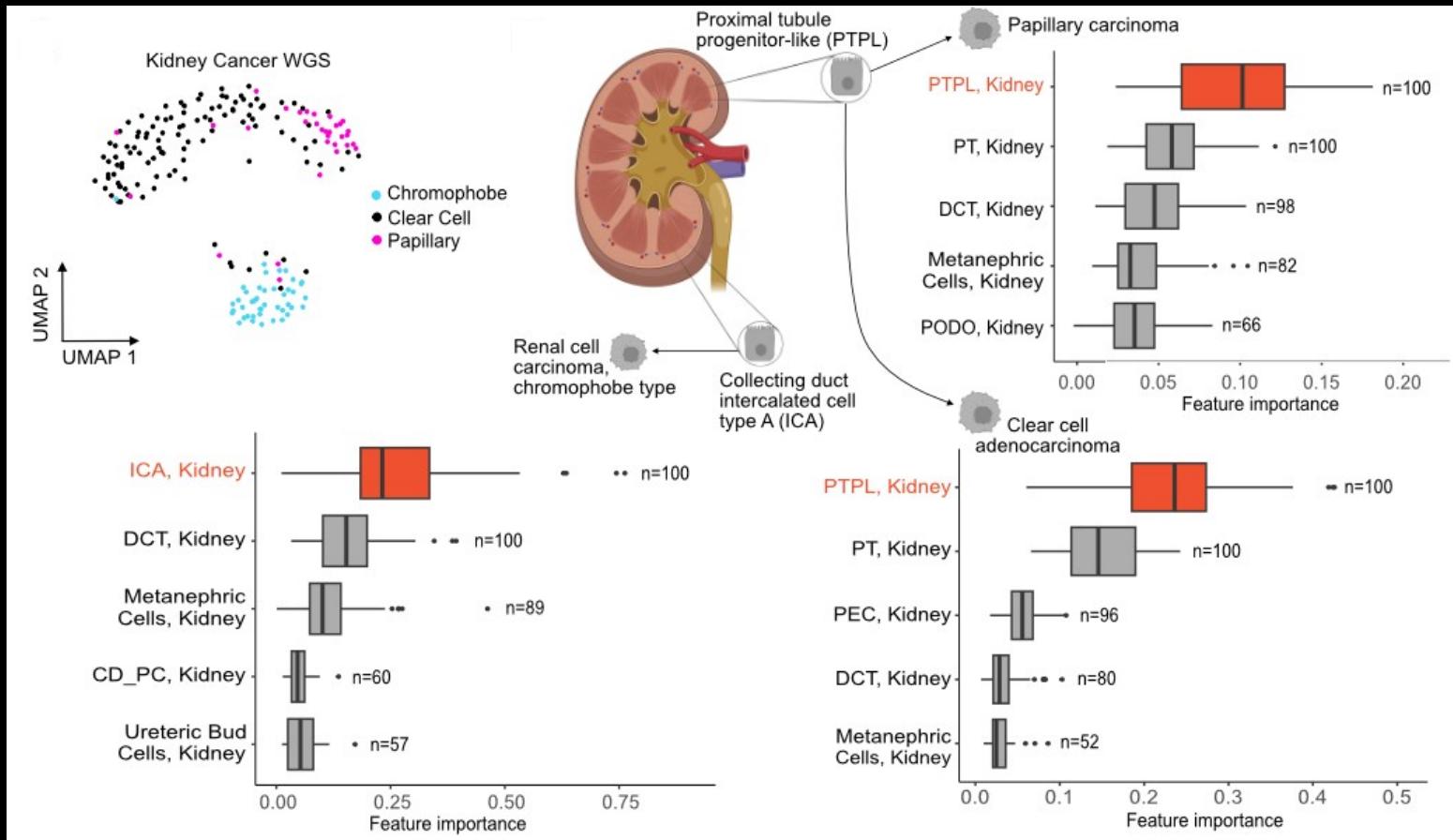
- Lung COOs are highly proliferative in homeostasis



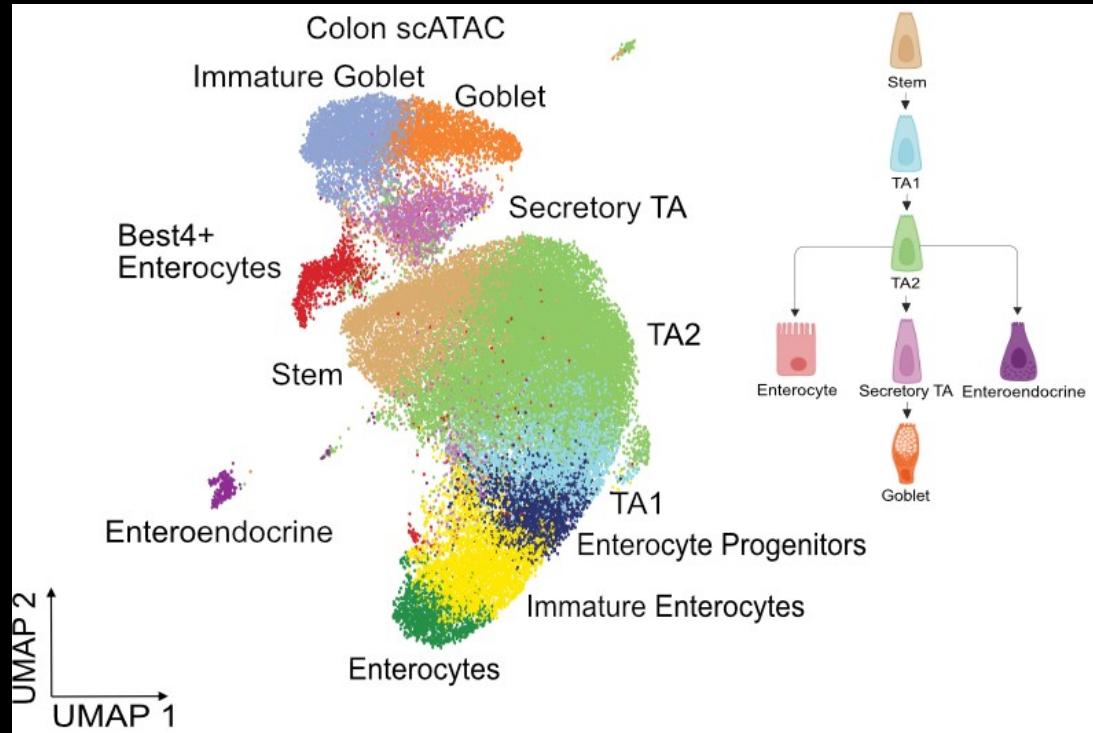
- Lung COOs are highly proliferative in homeostasis



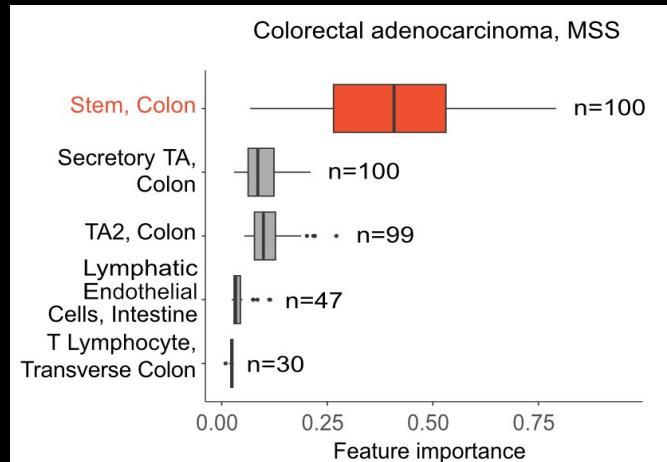
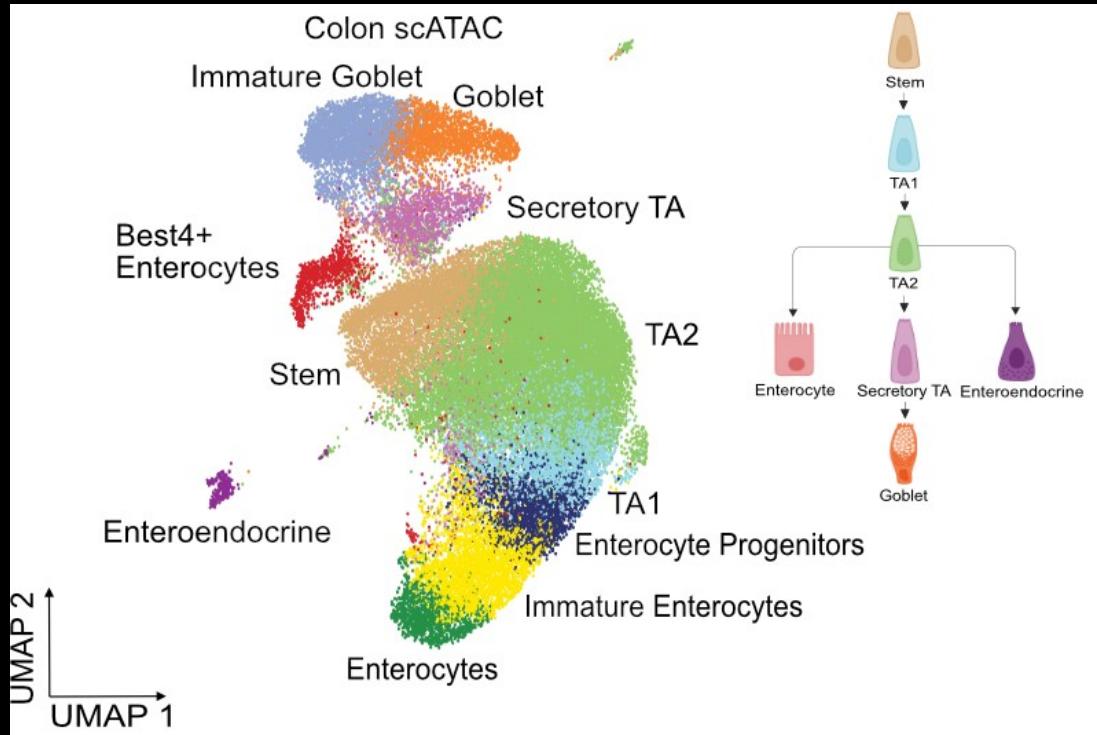
- Histological cancer subtypes are associated with different cells of origin



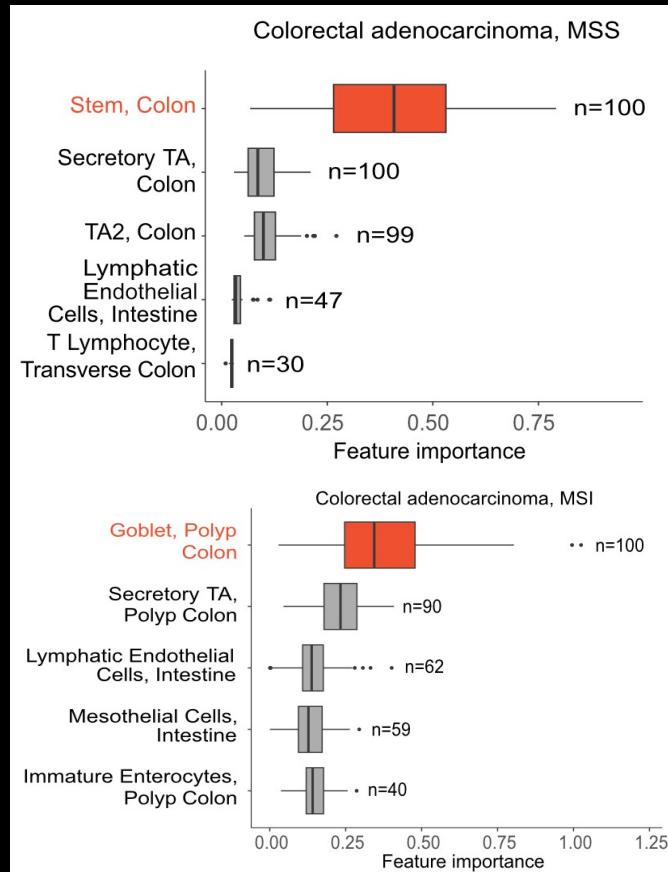
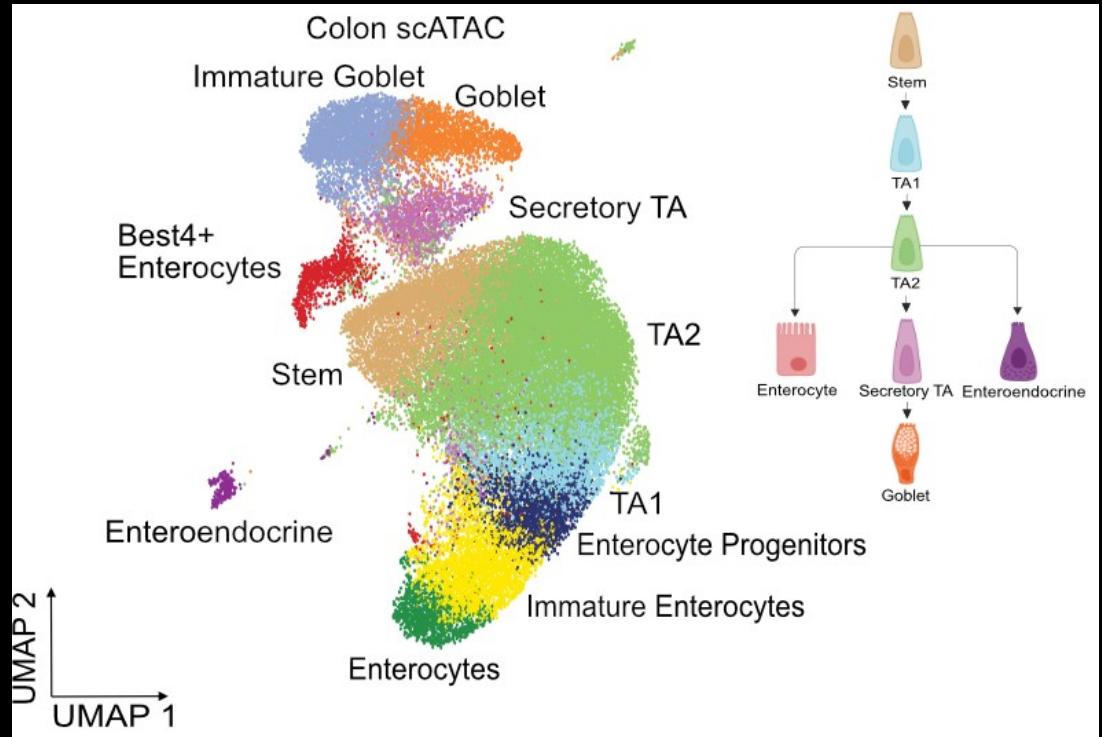
- Model exhibits cell subset specificity



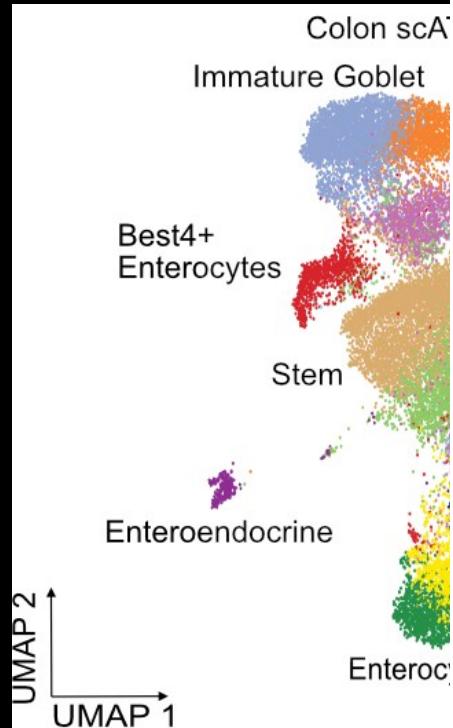
- Model exhibits cell subset specificity



- Model exhibits cell subset specificity



- Model exhibits

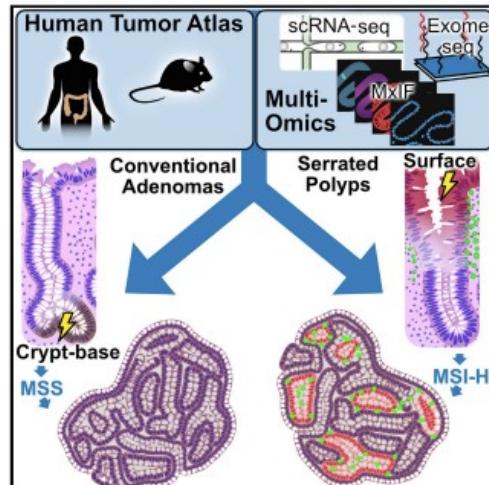


Cell

Article

Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps

Graphical abstract



Authors

Bob Chen, Cherie' R. Scurrah, Eliot T. McKinley, ..., Robert J. Coffey, Martha J. Shrubsole, Ken S. Lau

Correspondence

robert.coffey@vumc.org (R.J.C.), martha.shrubsole@vanderbilt.edu (M.J.S.), ken.s.lau@vanderbilt.edu (K.S.L.)

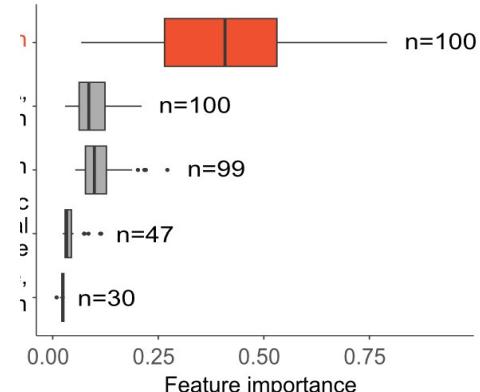
In brief

A single-cell resolution atlas of human colorectal polyps maps out distinct paths for pre-cancer to cancer transformation, accompanied by differential immune microenvironment features.

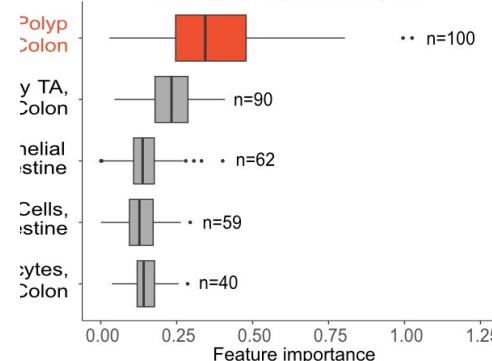
Highlights

- A single-cell resolution atlas of human adenomas and serrated polyps
- Serrated polyps arise from metaplasia as opposed to stem cell expansion
- Cytotoxic immunity in serrated polyps occurs independently of hypermutation
- Distinct immune microenvironments track tumor cell-differentiation states

Colorectal adenocarcinoma, MSS

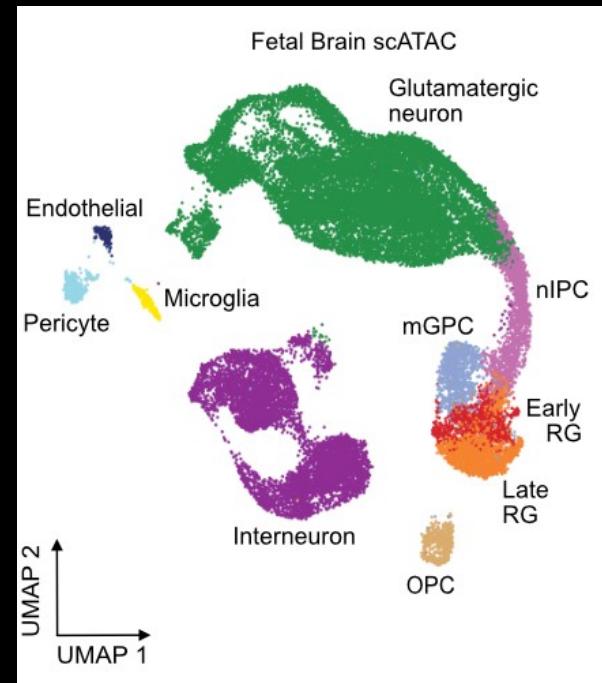


Colorectal adenocarcinoma, MSI

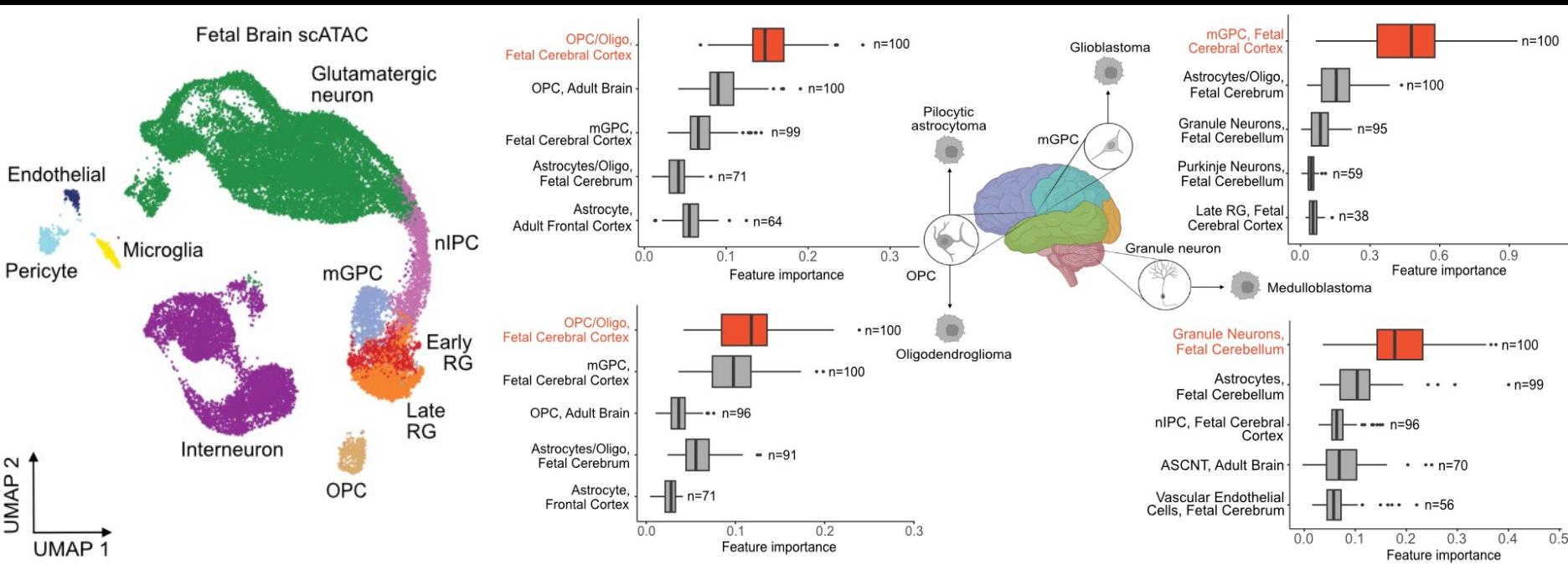


- Model predicts fetal-like progenitor populations as COOs for brain cancers

- Model predicts fetal-like progenitor populations as COOs for brain cancers



- Model predicts fetal-like progenitor populations as COOs for brain cancers



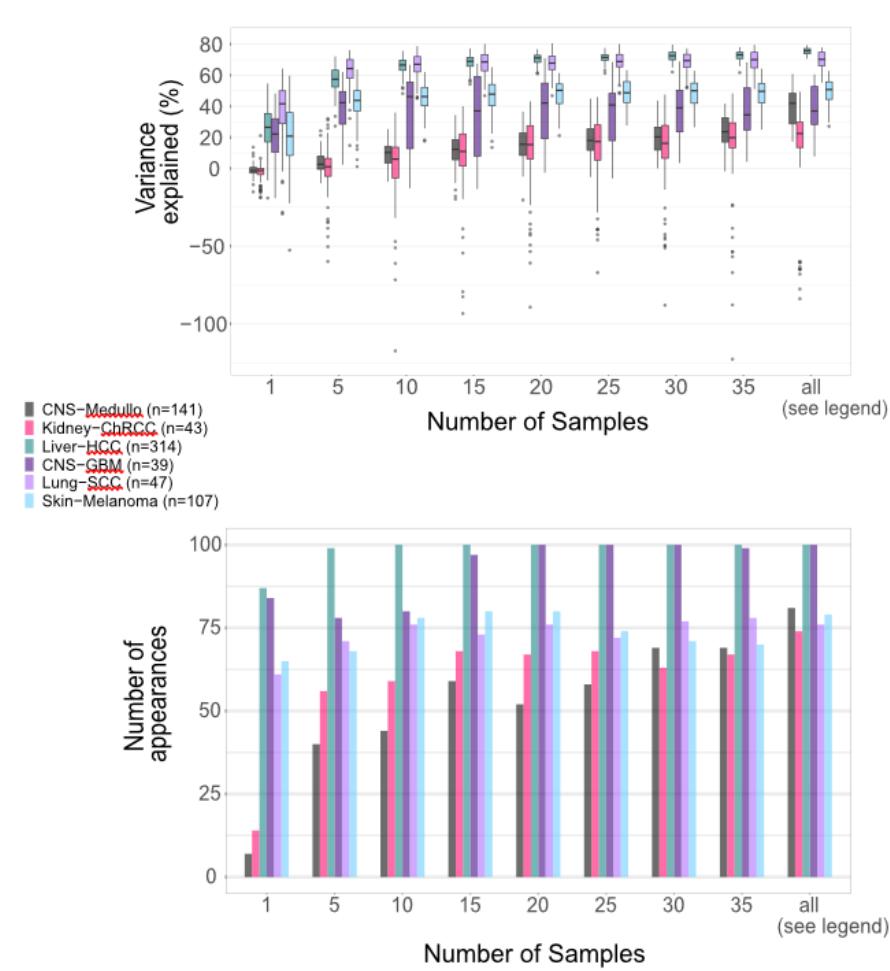
- **Limitations**
 - COO may vary even within the same type of cancer

- **Limitations**

- COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC

- **Limitations**
 - COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise

- Limitations
 - COO may
 - e.g.
 - per-



sparsity and noise

- **Limitations**
 - COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise
 - future direction: **data augmentation?**

- **Limitations**

- COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise
 - future direction: data augmentation?
- **Comprehensiveness of cell type coverage**

- **Limitations**

- COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise
 - future direction: data augmentation?
- Comprehensiveness of cell type coverage
- Diverse dataset sources

- **Limitations**
 - COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise
 - future direction: data augmentation?
 - Comprehensiveness of cell type coverage
 - Diverse dataset sources
 - **remediation: uniform data generation protocols**

- **Limitations**

- COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise
 - future direction: data augmentation?
- Comprehensiveness of cell type coverage
- Diverse dataset sources
 - remediation: uniform data generation protocols
- **Correlation does not imply causation!**

- **Limitations**

- COO may vary even within the same type of cancer
 - e.g Basal vs Neuroendocrine for SCLC
 - per-patient predictions are challenging due to data sparsity and noise
 - future direction: data augmentation?
- Comprehensiveness of cell type coverage
- Diverse dataset sources
 - remediation: uniform data generation protocols
- Correlation does not imply causation!
 - **experimental validation needed**

- Most important takeaway...
from the perspective of a cancer biologist

- Most important takeaway...
from the perspective of a cancer biologist

Can NOW generate novel hypotheses about the suspected COO
for a given cancer of interest at unprecedented cellular /
subcellular resolution using scATAC-seq.

- Most important takeaway...
from the perspective of a cancer biologist

Can NOW generate novel hypotheses about the suspected COO
for a given cancer of interest at unprecedented cellular /
subcellular resolution using scATAC-seq.

1. Sequence suspected normal TOOs comprehensively using
scATAC (or use our comprehensive scATAC dataset)

- Most important takeaway...
from the perspective of a cancer biologist

Can NOW generate novel hypotheses about the suspected COO
for a given cancer of interest at unprecedented cellular /
subcellular resolution using scATAC-seq.

1. Sequence suspected normal TOOs comprehensively using
scATAC (or use our comprehensive scATAC dataset)
2. Obtain WGS mutation profile for your cancer of interest

- Most important takeaway...
from the perspective of a cancer biologist

Can NOW generate novel hypotheses about the suspected COO
for a given cancer of interest at unprecedented cellular /
subcellular resolution using scATAC-seq.

1. Sequence suspected normal TOOs comprehensively using
scATAC (or use our comprehensive scATAC dataset)
2. Obtain WGS mutation profile for your cancer of interest
3. Use the framework highlighted in this presentation, which
will be made publicly available soon(ish) as

Acknowledgements

Chan
Zuckerberg
Initiative 

- Mount Sinai
Alex Tsankov
Bruno Giotti
Wooseung Lee
Tsankov lab members



- U. of Zagreb
Rosa Karlic
Paula Stanci



- Paz Polak



Thank you!

