

**A novel computational framework for  
analyzing single-cell transcriptomics data:  
applications to T cell exhaustion and  
cancer immunotherapy response**

*Christopher D. Klocke*

A dissertation presented in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy** in  
**Bioinformatics & Computational Biomedicine**  
to the **School of Medicine** at  
**Oregon Health & Science University.**

Department of Medical Informatics & Clinical Epidemiology

School of Medicine

Oregon Health & Science University

June, 2024



© COPYRIGHT 2024 BY CHRISTOPHER D. KLOCKE  
ALL RIGHTS RESERVED

Bioinformatics & Computational Biomedicine  
School of Medicine  
Oregon Health & Science University

---

CERTIFICATE OF APPROVAL

---

This is to certify that the PhD dissertation of  
Christopher D. Klocke  
has been approved.

---

Dr. Guanming Wu  
Advisor

---

Dr. Andrew Adey  
Chair

---

Dr. Amy Moran  
Committee Member

---

Dr. Shannon McWeeney  
Committee Member

---

Dr. Melissa Wong  
Committee Member

# TABLE OF CONTENTS

<b>List of Figures</b>	<b>6</b>
<b>List of Tables</b>	<b>12</b>
<b>1 Introduction</b>	<b>14</b>
1.1 Summary . . . . .	14
1.2 Cancer – The Clinical Problem . . . . .	14
1.3 Tumor Heterogeneity . . . . .	15
1.4 Single-Cell Assays . . . . .	16
1.5 A Systems-Level View . . . . .	17
1.5.1 Advantages of a Systems-Level View . . . . .	17
1.5.2 Biological Pathways . . . . .	17
1.5.3 Gene-Regulatory Networks . . . . .	18
1.6 The Immune System and Cancer . . . . .	19
1.6.1 Overview . . . . .	19
1.6.2 Immune Checkpoint Inhibitors . . . . .	20
1.6.3 T cell exhaustion . . . . .	23
1.6.4 Using scRNA-seq to Study Immune Checkpoint Inhibitor Response and T Cell Exhaustion . . . . .	24
1.7 Limitations and Open Questions . . . . .	26
<b>2 T cell exhaustion in the tumor immune microenvironment</b>	<b>27</b>
2.1 Introduction . . . . .	27
2.2 Results . . . . .	29
2.2.1 Tumor immune microenvironment and chronic viral infection datasets . . . . .	29
2.2.2 A novel computational framework to identify TFs that are significantly correlated with the T cell exhaustion process . . . . .	29

2.2. Inference of CD8 T cell exhaustion pseudotime trajectory	30
2.2. Ranking samples by sample-level CD8 T cell exhaustion score	30
2.2. Differential TF activity analysis for high- and low-exhaustion sample groups in immune cell types	31
2.2.3 Application of the computational framework to scRNA-seq datasets from human skin cancer and HIV infection	31
2.2.4 Identification of gene regulatory patterns shared across skin tumor datasets that are related to T cell exhaustion	32
2.2. Transcription factors up-regulated in terminally exhausted tumor-infiltrating CD8 T cells	35
2.2. Transcription factors down-regulated in terminally exhausted tumor-infiltrating CD8 T cells	35
2.2. Down-regulation of Kruppel-like factor (KLF) signaling in terminally exhausted tumor-infiltrating CD8 T cells	36
2.2.5 Transcription factors up-regulated in macrophages of highly exhausted tumor samples	36
2.2.6 Gene-regulatory network modules associated with CD8 T cell exhaustion	36
2.2. Exhaustion-associated CD8 T cell GRNs	36
2.2. Exhaustion-associated macrophage GRNs	37
2.2.7 Identification of pathways significantly related to T cell exhaustion in macrophages	38
2.3 Discussion	39
2.4 Data Availability	41
2.5 Code Availability	41
2.6 Results Availability	41
2.7 Acknowledgements	41
2.8 Methods	41
2.8.1 Datasets	41
2.8.2 Pre-processing, dimensionality reduction, batch correction, and clustering	42
2.8.3 Cell cluster annotation	42
2.8.4 Inference of CD8 T cell exhaustion trajectory with <i>Monocle3</i>	42
2.8.5 Calculation of sample-level exhaustion scores with the cell enrichment approach	42
2.8.6 Sample clustering and selection of extremes	43
2.8.7 Quantifying transcription factor regulon activity with <i>SCENIC</i>	43
2.8.8 Comparison of transcription factor activity between exhaustion-low and exhaustion-high samples in each cell type	43

2.8.9	Overlap analysis . . . . .	43
2.8.10	Analysis of exhaustion-associated gene-regulatory sub-networks . . . . .	44
2.8.11	Pathway enrichment analysis with <i>Reactome</i> . . . . .	44
2.8.12	Other package versions: . . . . .	44
<b>3</b>	<b>T cell exhaustion and immunotherapy response</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Results . . . . .	47
3.2.1	A novel computational framework to identify cell subpopulation activity related to cell state transition and clinical phenotypes . . . . .	47
3.2.1	Identification of suitable data for application of the framework	47
3.2.1	Pre-processing of single-cell data	47
3.2.1	Inference of cell state trajectory	49
3.2.1	Evaluating the cell state trajectory and providing additional support	49
3.2.1	Calculation of sample-level cell state scores and selection of extremes	49
3.2.1	Encoding of clinical phenotype	49
3.2.1	Identification of phenotype-associated cell subpopulations with <i>Scissor</i>	49
3.2.1	Additional support for phenotype association of selected cell subpopulations	50
3.2.1	Comparison of cell state-associated and phenotype-associated cell subpopulations	50
3.2.1	Pathway analysis	50
3.2.2	Application of framework to NSCLC tumor immune microenvironment data . . . . .	50
3.2.3	Data integration, pre-processing, annotation, and subsetting . . . . .	50
3.2.3	Cell counts checks	52
3.2.4	CD8 T cell exhaustion trajectory and sample-level exhaustion scores . . . . .	52
3.2.4	CD8 T cell exhaustion trajectory	52
3.2.4	Sample-level exhaustion scores	52
3.2.4	Support for the exhaustion trajectory – part 1	52
3.2.4	Support for the exhaustion trajectory – part 2	54
3.2.4	Selection of exhaustion-high and exhaustion-low tumor cell subpopulations	55
3.2.5	Identification of phenotype-associated lung tumor subpopulations with <i>Scissor</i> . . . . .	55
3.2.5	<i>Scissor</i> overview	55
3.2.5	Inputs	55
3.2.5	Selection of lung tumor cells related to immunotherapy response and resistance	56
3.2.5	Additional support for phenotype-associated tumor cells – part 1	57
3.2.5	Additional support for phenotype-associated tumor cells – part 2	57
3.2.5	Additional support for phenotype-associated tumor cells – part 3	57
3.2.6	Comparison of immunotherapy outcome-associated and T cell exhaustion-associated lung tumor subpopulations . . . . .	58
3.2.6	Sample-level comparison	58
3.2.6	DEG comparison	58
3.2.6	Pathway analysis	59
3.2.7	Application of the framework to melanoma data . . . . .	60

3.2. Application of ICI outcome-related gene signatures to bulk melanoma data	61
3.3 Discussion . . . . .	61
3.4 Methods . . . . .	62
3.4. Datasets	63
3.4.1 NSCLC data – single-cell	63
3.4.2 NSCLC data – bulk with ICI response annotation	63
3.4.3 Melanoma data – bulk with ICI response annotation	63
3.4.1 Pre-processing, merging, cell type annotation, and subsetting of scRNA-seq data . . . . .	63
3.4. Merging, batch correction, and pre-processing	63
3.4.1 Cell cluster annotation and subsetting	63
3.4.2 Inference of CD8 T cell exhaustion trajectory with <i>Monocle3</i> . . . . .	63
3.4.3 Calculation of sample-level exhaustion scores . . . . .	63
3.4.4 Sample clustering and selection of extremes . . . . .	64
3.4.5 Support for the exhaustion trajectory – part 1 . . . . .	64
3.4.6 Support for the exhaustion trajectory – part 2 . . . . .	64
3.4.7 Bulk RNA-seq data – pre-processing and DEG analysis . . . . .	64
3.4.8 Derivation of a binary response phenotype from RECIST scores . . . . .	64
3.4.9 Selection of ICI response- and resistance-associated tumor cells using <i>Scissor</i> . . . . .	65
3.4.10 Optimization of alpha parameter . . . . .	65
3.4.11 <i>Scissor</i> reliability test . . . . .	65
3.4.12 Additional support for phenotype-associated tumor cells – part 1 . . . . .	66
3.4.13 Additional support for phenotype-associated tumor cells – part 2 . . . . .	66
3.4.14 Additional support for phenotype-associated tumor cells – part 3 . . . . .	66
3.4.15 Differential gene expression analysis – exhaustion scores vs. ICI outcome . . . . .	67
3.4.16 Sample representation analysis – exhaustion scores vs. ICI outcome . . . . .	67
3.4.17 Pathway enrichment analysis with Reactome . . . . .	67
3.4.18 Generalization of NSCLC gene signature to other tumor types . . . . .	67
3.4.19 Other package versions: . . . . .	68
<b>4 Conclusion</b>	<b>69</b>
<b>5 Appendix</b>	<b>71</b>



## List of Figures

1	<b>Selection for resistance</b> . . . . .	15
2	<b>CD8 T cell destroying tumor cell</b> . . . . .	20
3	<b>Immunosurveillance</b> A) antigen recognition and T cell receptor binding B) immune checkpoint inhibits anti- tumor cytotoxic activity C) immune checkpoint inhibition with monoclonal antibodies . . . . .	21
4	<b>Relationship between immune cell activity and the degree of CD8 T cell exhaustion in the tumor immune microenvironment</b> . . . . .	28
5	<b>Analysis overview</b> A) input – scRNA-seq data from skin tumor microenvironment; B) select CD8 T cells using marker genes; C) infer pseudotime trajectory from progenitor (TCF7 high, TOX low) exhausted to terminally (TCF7 low, TOX high) exhausted CD8 T cells using Monocle3; D) calculate sample-level exhaustion score, quantifying the distribution of cells along the exhaustion trajectory by sample; E) select highest- and lowest-scoring sample groups; compare transcription factor activity in these samples for individual cell types; F) perform gene-regulatory network analysis of exhaustion-associated transcription factors. . . . .	30
6	CD8 T cell exhaustion in human skin tumor samples. A, B, D-F) imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints LAG3, TIGIT, and PDCD1; C) Monocle3 pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells; G) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score. . . . .	32
7	CD8 T cell exhaustion in human HIV samples A, B, D-F) imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints LAG3, TIGIT, and PDCD1 C) Monocle3 pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells G) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score. . . . .	33
8	Overlap in exhaustion-associated transcription factor activity between melanoma and basal cell carcinoma (BCC) datasets Overlap analysis was performed between melanoma and basal cell carcinoma datasets to identify the proportion of significantly exhaustion-associated TFs that were shared vs. non-shared and whether this overlap was significant. A-C, E-F) The two tumor datasets exhibit significant overlap in up- and down-regulation of CD8 T cell activity, up-regulation of macrophage activity, and up- and down-regulation of NK cell activity associated with CD8 T cell exhaustion. D) The overlap of TFs down-regulated in the most exhausted samples was not significant. . . . .	34



9	Exhaustion-related transcription factor activity in CD8 T cells Exhaustion-associated transcription factors (graph nodes) and their regulatory relationships, inferred with pyscenic (graph edges) A) transcription factors up-regulated in the CD8 T cells of the most CD8 T cell-exhausted immune microenvironments B) transcription factors down-regulated in the CD8 T cells of the most CD8 T cell-exhausted immune microenvironments. . . . .	37
10	Exhaustion-related transcription factor activity in macrophages Exhaustion-associated transcription factor activity in macrophages (graph nodes) and their regulatory relationships, inferred with pyscenic (graph edges) A) transcription factors up-regulated in the macrophages of the most CD8 T cell-exhausted immune microenvironments B) transcription factors down-regulated in the macrophages of the most CD8 T cell-exhausted immune microenvironments. . . . .	38
11	Pathway analysis results of exhaustion-associated DEGs in macrophages. Pathways enriched in genes over-expressed in macrophages of high-exhaustion samples relative to low-exhaustion samples; red line = adjusted p-value cutoff of 0.05 A) tumor-specific exhaustion-associated pathway activity B) viral-specific exhaustion-associated pathway activity C) exhaustion-associated pathway activity shared between tumor and viral contexts. . . . .	39
12	<b>Heterogeneity in tumor subpopulation modulation of and response to anti-tumor immune activity</b>	46
13	<b>Overview of the analysis workflow</b> . . . . .	48
14	<b>Lung tumor data – merged datasets</b> A) three NSCLC datasets integrated for analysis B) CD8 T cell and tumor cell populations were identified C) eight samples in the merged dataset contain tumor cells in addition to CD8 T cells and other immune cell types; cells labeled “background” are from 33 tumors from the <i>Liu et al., 2022</i> dataset, which does not contain sufficient tumor cells for analysis – these T cells are included as a reference to help delineate the CD8 T cell exhaustion trajectory D) Leiden clustering of merged dataset . . . . .	51
15	<b>CD8 T cell exhaustion in human skin tumor samples</b> A, B, D-F) imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints PDCD1, LAG3, and TIGIT; C) <i>Monocle3</i> pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells; G) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score. . . . .	53
16	<b>lung tumor cell subpopulations</b> A) Scissor-positive (ICI response- associated) and Scissor-negative (ICI resistance-associated) tumor cell subpopulations, superimposed on reference cells B) lung tumor cells with the lowest (microenvironment biased toward progenitor exhausted CD8 T cells) and highest (biased toward terminal exhaustion) exhaustion scores in their tumor immune microenvironments C) lung tumor cells from eight patients D) ICI response labels (RECIST) where available . . . . .	56

17	<b>proportion of cells in each sample that are scissor-positive (ICI response-associated) and scissor-negative (ICI resistance-associated), respectively</b> Note: samples are ranked from lowest to highest exhaustion score (left to right) . . . . .	59
18	<b>Pathway analysis of DEGs in lung tumor cells</b> Pathways enriched in genes up-regulated in exhaustion and ICI response groups; red line = adjusted p-value cutoff of 0.05 A) pathways up-regulated in exhaustion-high samples and ICI-responsive subpopulations B) pathways up-regulated in exhaustion-low samples and ICI- responsive subpopulations . . . . .	60
19	<b>CD8 T cell exhaustion in human skin tumor (basal cell and squamous cell carcinoma) samples a, b, d-f)</b> imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints LAG3, TIGIT, and PDCD1 c) Monocle3 pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells g) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score . . . . .	72
20	<b>gene expression vs. pseudotime in melanoma dataset</b> Expression of TCF7, the primary marker of the progenitor exhausted CD8 T cell state, starts high and is monotonically decreasing across pseudotime. TOX, the primary marker of the terminally exhausted CD8 T cell state, shows the opposite relationship. Immune checkpoint genes LAG3, PDCD1, and TIGIT show broadly similar expression patterns to TOX relative to pseudotime, although they decrease slightly for the highest pseudotime values. . . . .	73
21	<b>CD8 T cell UMAP plot – expression of key genes in melanoma dataset</b> . . . . .	74
22	<b>CD8 T cell UMAP plot – leiden clusters for melanoma dataset</b> . . . . .	75
22	<b>CD8 T cell UMAP plot – Monocle3 principal graph and clusters for melanoma dataset</b> top: clusters; bottom: principal graph and root cells <b>Parameters used:</b> Number of principal components: 18; Clustering: k=20, resolution=0.0001 . . . . .	76
23	<b>gene expression vs. pseudotime in basal cell carcinoma dataset</b> Expression of TCF7, the primary marker of the progenitor exhausted CD8 T cell state, starts high, increases very slightly and then decreases relative to pseudotime. TOX, the primary marker of the terminally exhausted CD8 T cell state, is monotonically increasing across pseudotime. Immune checkpoint genes LAG3, PDCD1, and TIGIT increase across pseudotime (with a slight decrease in LAG3 at the beginning). . . . .	77
24	<b>CD8 T cell UMAP plot – expression of key genes – BCC dataset</b> . . . . .	78
25	<b>CD8 T cell UMAP plot – Leiden clusters – BCC dataset</b> . . . . .	79
25	<b>CD8 T cell UMAP plot – Monocle3 principal graph and clusters – BCC dataset</b> top: clusters; bottom: principal graph and root cells <b>Parameters used:</b> Number of principal components: 15; Clustering: k=20, resolution=0.0001 . . . . .	80

26	<b>gene expression vs. pseudotime in chronic HIV dataset</b> Expression of TCF7, the primary marker of the progenitor exhausted CD8 T cell state, starts high before increasing slightly and then decreasing across pseudotime. TOX, the primary marker of the terminally exhausted CD8 T cell state, is monotonically increasing across pseudotime. Immune checkpoint genes LAG3, PDCD1, and TIGIT are also monotonically increasing across pseudotime. . . . .	81
27	<b>CD8 T cell UMAP plot – expression of key genes – HIV dataset</b> . . . . .	82
28	<b>CD8 T cell UMAP plot – Leiden clusters – HIV dataset</b> . . . . .	83
28	<b>CD8 T cell UMAP plot – Monocle3 graph and clusters – HIV dataset</b> top: clusters; bottom: principal graph and root cells <b>Parameters used:</b> Number of principal components: 10; Clustering: k=20, resolution=0.0001; ‘learn_graph()’: use_partition=FALSE . . . . .	84
29	<b>Pathway enrichment of genes up-regulated in the most exhausted macrophage samples – unabridged, tumor-specific</b> . . . . .	86
30	<b>Pathway enrichment of genes up-regulated in the most exhausted macrophage samples – unabridged, viral-specific</b> . . . . .	87
31	<b>Pathway enrichment of genes up-regulated in the most exhausted macrophage samples – unabridged, shared</b> . . . . .	88
32	<b>Pathway enrichment of genes up-regulated in the most exhausted CD8 T cell samples – unabridged, tumor-specific</b> . . . . .	89
33	<b>Pathway enrichment of genes up-regulated in the most exhausted CD8 T cell samples – unabridged, viral-specific</b> . . . . .	90
34	<b>Pathway enrichment of genes up-regulated in the most exhausted CD8 T cell samples – unabridged, shared</b> . . . . .	91
35	<b>Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, tumor-specific</b> . . . . .	92
36	<b>Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, viral-specific</b> . . . . .	93
36	<b>Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, viral-specific</b> . . . . .	94
37	<b>Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, shared</b> . . . . .	95
38	<b>Leiden clusters</b> <i>Hu et al., 2023</i> dataset (pre- and post-treatment tumor cells) . . . . .	96

39	<b>Patients</b> <i>Hu et al., 2023</i> dataset (pre- and post-treatment tumor cells) . . . . .	97
40	<b>Pre- vs. post-treatment</b> <i>Hu et al., 2023</i> dataset (pre- and post-treatment tumor cells) . . . . .	97
41	<b>Treatment outcome (RECIST)</b> <i>Hu et al., 2023</i> dataset (pre- and post-treatment tumor cells); PR: Partial Response; SD: Stable Disease . . . . .	98
42	<b>Group</b> <i>Hu et al., 2023</i> dataset (pre- and post-treatment tumor cells) . . . . .	99
43	<b>Scissor status</b> <i>Hu et al., 2023</i> dataset (pre- and post-treatment tumor cells); phenotype-associated cells selected by Scissor; Note: any Scissor-positive or Scissor-negative cells not from the Hu dataset not pictured here. Additionally, only pre-treatment cells were used for phenotype association analysis . . . . .	100
44	<b>Significant DEG overlap – Scissor-positive and exhaustion-high</b> x-axis: genes whose expression is up-regulated in immunotherapy response-related tumor cells relative to immunotherapy resistance-related tumor cells; y-axis: genes whose expression is up-regulated in high CD8 T cell exhaustion-associated tumor cells relative to low exhaustion-associated tumor cells; significance quantified by $-\log(\text{adjusted p-value})$ . . . . .	101
45	<b>Significant DEG overlap – Scissor-positive and exhaustion-low</b> x-axis: genes whose expression is up-regulated in immunotherapy response-related tumor cells relative to immunotherapy resistance-related tumor cells; y-axis: genes whose expression is up-regulated in low CD8 T cell exhaustion-associated tumor cells relative to high exhaustion-associated tumor cells; significance quantified by $-\log(\text{adjusted p-value})$ . . . . .	102
46	<b>Significant DEG overlap – Scissor-negative and exhaustion-low</b> x-axis: genes whose expression is up-regulated in immunotherapy resistance-related tumor cells relative to immunotherapy response-related tumor cells; y-axis: genes whose expression is up-regulated in low CD8 T cell exhaustion-associated tumor cells relative to high exhaustion-associated tumor cells; significance quantified by $-\log(\text{adjusted p-value})$ . . . . .	103
47	<b>Significant pathway overlap – Scissor-positive and exhaustion-high</b> x-axis: pathways whose expression is up-regulated in immunotherapy response-related tumor cells relative to immunotherapy resistance-related tumor cells; y-axis: pathways whose expression is up-regulated in high CD8 T cell exhaustion-associated tumor cells relative to low exhaustion-associated tumor cells . . . . .	104
48	<b>Percentage of variance explained by principal components – NSCLC CD8 T cells</b> five principal components were used for CD8 T cell trajectory . . . . .	105
49	<b>Marker genes for setup of CD8 T cell exhaustion trajectory – CD8 T cells from NSCLC tumors</b> . . . . .	106
50	<b>Leiden clusters and Monocle 3 principal graph – CD8 T cells from NSCLC tumors</b> principal point Y_83 used as root for initialization of Monocle 3 trajectory inference . . . . .	107

51	<b>Expression of key genes vs. pseudotime – CD8 T cells from NSCLC tumors</b> . . . . .	108
52	<b>Optimization of Scissor’s alpha parameter, part 1</b> An optimal value of Scissor’s alpha parameter was chosen for the lung tumor cells by considering the tradeoff between a higher value, which selects fewer phenotype-associated cells for downstream analysis, and a lower value, which may select more heterogeneous cells with a wider range of expression profiles. In order to choose a clearly defined subset of the tumor cells, in sufficient quantity, a value of 0.275 was determined to be optimal. . . . .	109
53	<b>Optimization of Scissor’s alpha parameter, part 2</b> See description of previous figure. . . . .	110
54	<b>biorender publication license for Figure 2</b> . . . . .	111
55	<b>biorender publication license for Figure 3</b> . . . . .	112
56	<b>biorender publication license for Figure 4</b> . . . . .	113
57	<b>biorender publication license for Figure 12</b> . . . . .	114

## List of Tables

1	scRNA-seq datasets. Three publicly available scRNA-seq datasets used – two from tumor immune microenvironment (melanoma, basal cell carcinoma) and one from chronic viral infection (HIV). . . .	29
2	<b>scRNA-seq datasets</b> Three publicly available scRNA-seq datasets from human NSCLC tumors were used	51
3	<b>Cell counts by sample</b> . . . . .	52
4	<b>Li et al. dataset, exhaustion high group, CD8 T cells</b> – test statistic = 1.77, p-value = $1.77 \times 10^{-34}$	54
5	<b>Li et al. dataset, exhaustion low group, CD8 T cells</b> – test statistic = 2.03, p-value = $4.24 \times 10^{-11}$	54
6	<b>Yost et al. dataset, exhaustion high group, CD8 T cells</b> – test statistic = 2.77, p-value = $1.91 \times 10^{-99}$	54
7	<b>Yost et al. dataset, exhaustion low group, CD8 T cells</b> – test statistic = 3.18, p-value = $2.92 \times 10^{-18}$	54
8	<b>Overlapping DEGs in lung tumor cells – T cell exhaustion vs. immunotherapy outcome</b> Genes are categorized based on their expression levels in subsets of tumor cells. Genes are categorized as “exhaustion-high” if their expression in the subset of tumor cells from samples with high CD8 T cell exhaustion scores is significantly up-regulated compared to the subset from low-exhaustion samples. Similarly, genes are categorized as “ICI-responsive” if their expression in the subset of tumor cells associated with ICI response is significantly up-regulated compared to the subset associated with ICI resistance. The inverses are true for “exhaustion-low” and “ICI-resistant.” . . . . .	60
9	<b>Generalization of ICI outcome DEGs from lung tumor cells to melanoma data</b> – test statistic = 2.96, p-value = $2.83 \times 10^{-187}$ . . . . .	61
10	<b>Cell counts by cell type for each of three datasets</b> . . . . .	96



# **1 Introduction**

## **1.1 Summary**

Cancer immunotherapies show great potential for treating patients with otherwise incurable disease. Over the past decade, a type of immunotherapy known as immune checkpoint inhibition (ICI) has been approved to treat an increasing number of cancer types, including melanoma, non-small cell lung cancer, renal cell carcinoma, and some head and neck cancers [1]. However, these treatments are only effective in a subset of patients, which cannot be reliably identified prior to treatment; the mechanisms by which some patients exhibit resistance to ICI are not fully understood [2]. A better understanding of the gene expression and regulation patterns of heterogeneous malignant cell subpopulations and various immune cell types in the tumor microenvironment contributing to ICI treatment failure is critically needed. This will enable stratification of potential recipients into responders and non-responders and the development of novel therapies or combinations that succeed where current options fail.

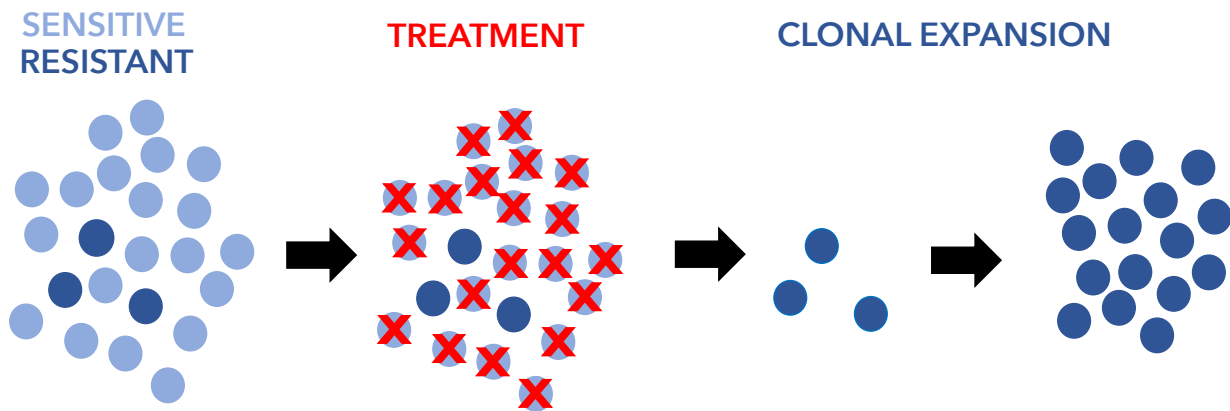
To this end, single-cell sequencing technologies have been used to interrogate the expression patterns of heterogeneous cell subpopulations in tumors, and have shown success in their application to questions related to topics such as T cell exhaustion and immunotherapy response [3, 4, 5, 6, 7]. Despite rapid development of methods and tools to analyze data produced by these powerful new assays, new approaches are still needed. A systems-level approach that considers cell type-specific changes in expression and regulation over time that contribute to T cell exhaustion and treatment failure will facilitate new, clinically relevant insights. The consideration of tumor cell subpopulations and the differential impact of key cell states is also critically needed. This type of analysis framework is necessary to enable more effective therapeutic targeting of the tumor immune microenvironment with ICI treatment. Such additions to the computational biology toolkit will enable progress on this important clinical question and, once developed, can be applied to other important biological questions as well.

## **1.2 Cancer – The Clinical Problem**

Cancer is a category of diseases characterized by an uncontrolled proliferation of cells within the patient's body. Cancer has a range of causes and occurs when normal biological processes become dysregulated, allowing a rogue population of cells to multiply and spread throughout the body [8, 9]. Cancer causes significant morbidity and mortality worldwide, leading to around six hundred thousand deaths annually in the United States alone [10, 11]. Various cancers account for almost one in six deaths globally, killing 9.6 million people annually [12].

This pressing clinical need has driven massive investment of skill, time, and money into researching this group of diseases. Billions of dollars are spent on cancer research every year [13]. Some progress has been made, and survival rates for most cancers have improved over the past several decades [14]. However, the remaining disease burden demands further research and new solutions. While there are many approved cancer treatments, including the classical approaches of surgery, radiation therapy, and chemotherapy and more recent targeted therapies, these treatments often





**Figure 1: Selection for resistance**

carry significant, burdensome side effects and often fail to elicit a durable response. Additionally, radiation and chemotherapy can induce iatrogenic secondary cancers [15]. These shortcomings have spurred an urgent search for new therapeutic approaches that address the limitations of existing treatment modalities.

### 1.3 Tumor Heterogeneity

Tumors are not homogeneous. Two types of heterogeneity must be considered when studying cancer. The first, inter-tumor heterogeneity, refers to the significant variability between patients. Since tumors arise out of healthy cells in each patient, and there are many routes to broadly similar states of molecular and cellular dysregulation that characterize tumors, cancer cell populations from different patients tend to differ significantly in their mutation profiles and histopathology characteristics even within the same cancer type [16]. The second, intra-tumor heterogeneity, describes the way in which individual tumors comprise genetically and epigenetically diverse subpopulations of cells [17]. This second type of heterogeneity is of primary interest when considering mechanisms of treatment resistance, as drugs can act as Darwinian selective pressures that cull certain groups of malignant cells and leave resistant subpopulations unharmed [18].

Intra-tumor heterogeneity has been understood for over four decades, since the “clonal” nature of cancer was first characterized. In 1976, Peter Nowell described cases in which all cells in a primary tumor exhibited the same abnormal karyotype, plasma cell clone (in the case of plasma cell cancers), or X chromosome (in women), suggesting they were descended from a common progenitor cell. Additionally, he suggested sequential selection of mutant subpopulations as a model for cancer evolution [19]. When a cell within a tumor sustains an additional genetic alteration, which can range from the modification of a single nucleotide to a loss-of-heterozygosity event in which a large section of a chromosome is lost, this change can be inherited by daughter cells that result when the cell divides. Importantly, genes and proteins that are clinically relevant, sometimes to the mechanism of targeted therapies, are often only expressed by a subset of cancer cells within a tumor [20]. This can lead to selection for and clonal expansion of resistant subpopulations in some patients, leading to treatment failure (Figure 1) [18]. In addition to genetic alterations, epigenetic changes, such as the

methylation of particular regions of the genome, can lead to functionally important changes in expression activity. These epigenetic changes have been studied more in recent years as new assays have become available [17].

Another form of heterogeneity that appears in tumor sample data comes from the mixture of other cell types present within and around a tumor. Cancer-associated fibroblasts (CAFs), immune cells (including tumor-infiltrating lymphocytes), and other adjacent healthy cells are present in the tumor microenvironment. This is not tumor heterogeneity in the same sense but will show up as heterogeneity in experimental data [21]. In addition to considering many cell types present within a tumor, researchers must also analyze various cell states that may arise due to differentiation trajectories or changing biological conditions. The relative abundances of various cell states within a cell type can impact treatment efficacy. In the case of targeted therapies, the primary heterogeneity of interest has to do with drug-resistant subclonal populations of cancer cells that may need to be targeted with novel therapies or combinations. When it comes to cancer immunotherapies, this becomes more complicated, because the mechanism of treatment response relies on an interaction between multiple cell types (immune cells and cancer cells).

This heterogeneity poses a major challenge to cancer researchers. When patients with a given disease mechanism or treatment response profile vary in their expression of individual genes and proteins, it becomes difficult to achieve the statistical power necessary to detect these mechanisms even if there is shared underlying biology. To address this, researchers must take a systems-level view, discussed in greater detail below.

## **1.4 Single-Cell Assays**

While the advent of high-throughput sequencing technologies has allowed for the generation of vast amounts of genotype and expression data from a range of tissue types, bulk sequencing generates data in which the signal from all cells in the sample is mixed together. While computational methods can be used to deconvolve this bulk signal to a degree, they have some limitations, particularly in their ability to track cell type-specific expression changes in genes expressed in multiple cell types within a tissue [22]. Bulk RNA-seq data, while useful for some research questions, does not provide the resolution needed to perform detailed analysis of different cell subpopulations within a tissue sample. This complicates the study of tumor samples that contain heterogeneity within tumor cell populations, mixtures of cell types, and mixtures of cell states within each cell type, all of which may be relevant to tumorigenesis and treatment response.

Over the past decade, new assays have been developed that can deliver information on cellular features at a single-cell level of resolution. These single-cell assays, such as single-cell DNA-seq (genotype), single-cell RNA-seq (expression), and others, allow researchers to interrogate tissues on a cell-by-cell basis [23, 24, 25, 26, 27]. Performing single-cell RNA-seq (scRNA-seq) on a tissue sample yields information about which genes are being expressed in each cell in the tissue, which can be used to determine the cell type composition of the tissue and examine the expression in each of these cell types. This is particularly useful when studying clinical samples from heterogeneous tumors. Various cell types and phenotypically distinct cancer cell subpopulations can be identified and analyzed.

These assays have been used extensively for cancer research in recent years. Many papers have used single-cell assays, including scRNA-seq, to study patient tumor samples with greater granularity than is possible with bulk data [6, 7, 3]. In addition to this, several, more comprehensive efforts are underway to develop large, public resources with many samples and huge number of cells from a variety of cancer types [28, 29]. This proliferation of high-resolution cancer datasets will facilitate new insights in tumor biology for years to come.

## **1.5 A Systems-Level View**

### **1.5.1 Advantages of a Systems-Level View**

Systems biology has been enabled by the collection of systematic biological data. For example, high-throughput transcriptomics measures the expression of every protein-coding gene in a sample. With this type of data, researchers can start to look at the interacting units within a system, their relationships, and how information passes through the system under various circumstances [30]. An analysis that may not be sufficiently statistically powered to detect a gene or protein that may vary between two conditions may be able to detect a difference at a pathway or gene-regulatory network level by aggregating across a group of genes based on some pre-defined relationship. A subset of patients may all exhibit alterations in the same pathway, even if the same genes are not affected. In this way, the heterogeneity problem can be addressed [31].

This type of approach works because, when studying molecular processes within a cell, the information passed through a network of interacting components can be more important than any of the individual components themselves. These networks can involve protein-protein interactions (such as signal transduction cascades and ligand-receptor interactions), gene-regulatory networks, metabolic reactions, and post-translational modifications of proteins and protein complexes [31].

A comprehensive systems biology approach must take these relationships into account, moving beyond the individual gene or protein level to identify gene sets related to a biological or clinical phenomenon of interest. We can overlay these pathways and gene-regulatory network modules onto single-cell transcriptomics data and determine their activity levels in cell populations of interest. The mechanistic insights gleaned from systems-level analysis of biological systems can then inform treatment decisions that are more suited to an individual patient and the specifics of their tumor biology [31].

### **1.5.2 Biological Pathways**

Studying members of a pathway as a group instead of individually helps to address the heterogeneity problem. Different patients with a shared disease mechanism or treatment response may exhibit modification of different elements within the same pathway. These biological pathways hierarchically organize sets of genes based on shared functionality, physical interactions, and regulatory relationships. These elements may be subunits of a protein complex. Alternately, one might be upstream of another in a cascade of protein-protein interactions within a cell [30]. These

pathways and their members are delineated in publicly available resources such as Reactome [32, 33, 34, 35]. To augment their analyses, researchers can access information describing protein-protein interactions, protein complexes, and metabolic reactions that have been curated from the biological literature.

### 1.5.3 Gene-Regulatory Networks

Genes do not act in isolation. Throughout development and under normal physiological conditions, the expression of each gene is carefully regulated. Some genes code for transcription factors, proteins which bind to cis-acting DNA elements such as promoters and enhancers to activate or repress the expression of target genes [36, 37]. These transcription factor-target interactions can be characterized using assays such as chromatin immunoprecipitation followed by sequencing (ChIP-seq), and experimental evidence indicates that specific DNA sequence patterns that interact with such proteins are found throughout much of the human genome [38, 36]. Transcription factors are coded for by genes that themselves must be regulated; these regulatory relationships are organized into networks of regulatory relationships known as gene-regulatory networks (GRNs) [39].

Researchers must take a system-level view of gene expression regulation, instead of studying individual genes or pairs of interacting genes, to better understand the function of biological systems. The information encoded in GRNs allows organisms to undergo development, perform necessary biological functions, and respond to changing environmental conditions. Additionally, the function that results from the interactions of a group of genes is determined by the topology of the gene-regulatory sub-network to which they belong. By mapping out the logic of these interactions and the recurring motifs that appear, their behavior can be better understood [39].

GRNs are not static. Mutations in cis-regulatory element sequences, which allow or block interactions with transcription factors, can change the behavior of a regulatory sub-network. In fact, this mechanism of GRN rewiring appears throughout evolution as prevalent traits change in response to various selective pressures [40]. In addition, GRNs can be rewired during cancer, driving changes in cellular behavior associated with tumorigenesis [41, 42]. Characterizing the rewiring of GRNs in cancer has the potential to shed light on mechanisms of treatment resistance. There is also a convenient and beneficial side-effect of taking a GRN-level approach when analyzing scRNA-seq data from human tumor samples. By studying co-regulated gene sets rather than individual genes, one can mitigate the sparsity limitations inherent to single-cell assays while still leveraging their power to interrogate heterogeneous cancer cell populations and other inter-mixed cell types [43].

A number of software tools for inferring GRN relationships from scRNA-seq data have emerged in recent years. These methods may be used to identify potential transcription factor-target relationships and prioritize them for further study, allowing bench researchers to focus on the most promising candidates for validation [44]. Given unlimited time and resources, existing experimental methods such as ChIP-seq could be used to reconstruct networks of transcription factor-target interactions under various biological conditions. However, practical constraints become prohibitive when trying to do this in a high-throughput manner, making this slow, costly, and inefficient [38]. Instead, computational methods can often be used to infer these relationships with some degree of confidence. Computationally predicted

TF-target relationships can then be experimentally validated (as resources allow), using computational inference to narrow the search space and more efficiently deploy resources to validate selected interactions. Several of these GRN inference algorithms, which are based primarily on gene co-expression, have shown reasonably good performance in independent benchmarking analyses [45, 46, 47, 48].

Some of these methods infer a so-called “regulon,” or set of genes, that is regulated by a given transcription factor in a given biological context. Quantifying the expression levels of these regulons enables another advantage of the GRN approach: the activity of this gene set can act as a proxy for the expression and activity of the TF protein, rather than just its mRNA expression. Correlation between gene expression and protein expression is typically somewhat high, but it varies significantly by gene and is not always a sufficient proxy [49].

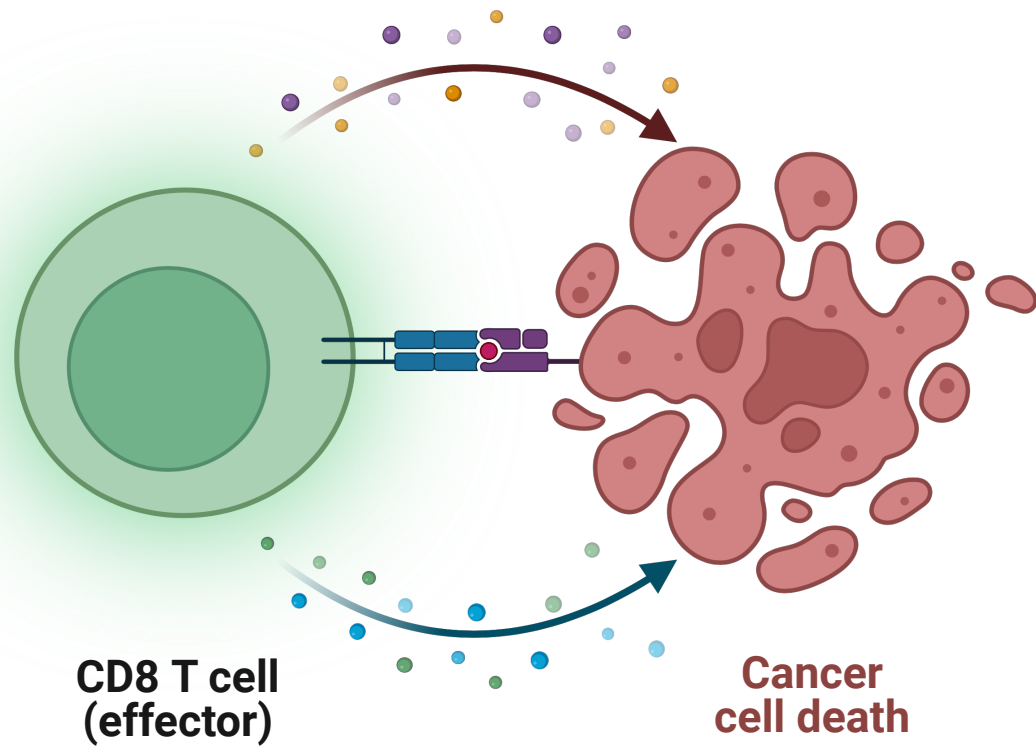
## **1.6 The Immune System and Cancer**

### **1.6.1 Overview**

The primary function of the immune system is to perform self / non-self determinations as accurately as possible. One’s own healthy cells must be recognized as such and left alone. Invading pathogens, as well as aberrant cells that may lead to uncontrolled proliferation, must be identified and destroyed. Designations of non-self inevitably involve some degree of Type I and Type II Error. False positives take the form of autoimmunity, where the immune system attacks its own healthy cells, while false negatives occur as a failure to eliminate nascent cancer or infection.

The immune system is one of the first lines of defense against the dysregulated proliferation of a group of cells which may become cancerous, and any tumor that progresses must find a way to escape this constant surveillance. The avoidance of immune destruction has been added as one of the “hallmarks of cancer” [9, 50]. A well-functioning immune system helps to prevent the development of tumors in several ways. Normal immune function protects the individual against viral infections, some of which are implicated in tumorigenesis. It also resolves inflammation throughout the body, a physiological condition that, if sustained, can promote the development of tumors [51]. Finally, normal immune function, some of which is carried out by CD8 T cells, can find and destroy malignant cells before they can progress into a more dangerous tumor (Figure 2). This process is known as immunosurveillance [52]. Malignant cells often display non-self antigens resulting from the myriad mutations in cancer genomes, which leads to T cell receptor (TCR) binding and cell destruction when these antigens are presented on MHC Class I [53] (Figure 3A).

Immunosurveillance plays a key role in the related process of immunoediting, in which some cancer cells are destroyed and others evolve to evade this response, whereby the adaptive immune system selects for malignant cells with lower immunogenicity. Immunoediting typically proceeds in three phases: elimination, equilibrium, and escape. In the elimination phase, cancer cells are targeted and destroyed through routine immunosurveillance. In the equilibrium phase, cancer cells are not eliminated completely but growth is limited as their proliferation is kept in check via immune control. Finally, in the escape phase, the cancer cell population can no longer be controlled by the immune system and will expand; to evade immunosurveillance, cancer cells may stop expressing MHC Class I, which presents



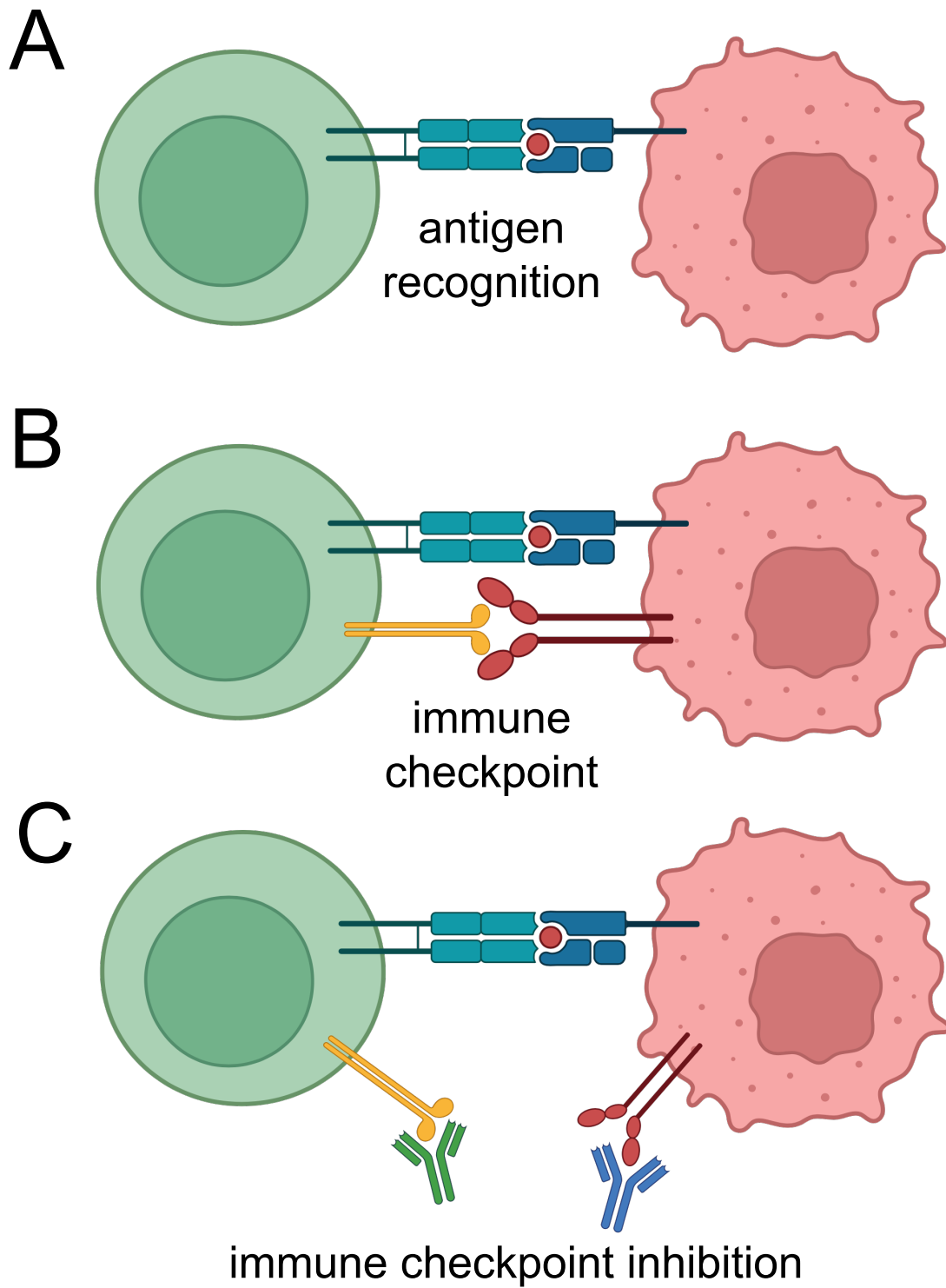
**Figure 2: CD8 T cell destroying tumor cell**

antigens to the adaptive immune system, or they may start expressing various immune checkpoint proteins to deactivate the immune response [51, 53].

In recent years, researchers have sought novel treatment modalities to influence the immune response to cancer, known as cancer immunotherapies. These treatments attempt to modulate the immune system so that it targets cancer cells more effectively. They include adoptive cell therapies such as chimeric antigen receptor T-cell (CAR-T) therapy, cancer vaccines, and immune checkpoint inhibitor (ICI) therapy [54, 55, 56, 53]. ICI therapies are of particular interest, as a range of these drugs have been approved for use in a variety of cancers over the past decade and have shown great success in a subset of patients [56].

### **1.6.2 Immune Checkpoint Inhibitors**

Immune checkpoints are inhibitory pathways that have evolved to maintain self-tolerance and prevent autoimmunity. These inhibitory pathways act to limit the duration and amplitude of immune responses against pathogens to prevent excess damage to surrounding healthy tissue [57, 58]. However, tumors can co-opt these immune checkpoints in order to evade immunosurveillance and escape destruction by CD8+ T cells capable of recognizing tumor-specific antigens [57, 59](Figure 3B). In the past two decades, researchers have developed techniques to interfere with these inhibitory



**Figure 3: Immunosurveillance** A) antigen recognition and T cell receptor binding B) immune checkpoint inhibits anti-tumor cytotoxic activity C) immune checkpoint inhibition with monoclonal antibodies

pathways and re-initiate anti-tumor immune responses; these pathways can be targeted with antibodies that disrupt their inhibitory signaling, known as immune checkpoint inhibitors (ICI) [57](Figure 3C).

While there exist a variety of proteins playing critical roles in immune checkpoint signaling that could be targeted to modulate immune response, antibodies against programmed cell death protein 1 (PD-1) and cytotoxic T-lymphocyte antigen 4 (CTLA-4) have been the first ICI therapies to show promise for cancer treatment and have led to durable treatment response in a subset of patients [57, 58, 60]. These two proteins are surface receptors, expressed primarily on the cell membranes of some T cells [51]. The function of these two immune checkpoint proteins, and thus the mechanism of action of treatments that inhibit them, are slightly different. CTLA-4 affects T cells during their initial activation upon exposure to their cognate antigen (often in lymph nodes), whereas PD-1 inhibits effector function of T cells in the tumor or infected tissue itself [60]. CTLA-4 is mainly expressed on CD4+ T cells, so the effect of anti-CTLA-4 therapy on CD8+ T cell function is likely to be indirect, via its effect on the activity of CD4+ helper and regulatory T cells [58]. Ipilimumab, an antibody against CTLA-4, was approved by the United States Food and Drug Administration (FDA) in 2011 for the treatment of advanced melanoma when Phase III clinical trials demonstrated that this treatment extended overall survival of patients when compared to standard of care [58]. Approvals for antibodies targeting PD-1 were soon to follow, with the FDA approvals of pembrolizumab and nivolumab in 2014 [61].

In spite of the profound success of these treatments in a subset of patients, most patients do not exhibit durable response following ICI therapy [51]. For this reason, attempts to develop new ICI treatments are underway. While anti-CTLA-4 and anti-PD-1 drugs have shown the most early success and regulatory approval, there are ongoing clinical trials testing the efficacy of similar monoclonal antibody blocking of the immune checkpoint signaling of TIM-3, LAG-3, and TIGIT, often in combination with anti-PD-1 drugs [62, 63, 64, 65, 66, 67, 68]. Importantly, some of these immune checkpoint proteins are co-regulated. For example, it was recently discovered that the immune checkpoints PD-1 and TIGIT are both regulated by the gene Blimp1 [69]. For this reason, a systems-level understanding of the gene regulatory relationships during ICI therapy is critical to achieve progress in this research area.

Researchers seek a better understanding of why ICI therapy succeeds in some patients and fails in others for two key reasons: patient stratification and therapy development. Mechanisms of resistance to ICI therapy may arise from the tumor immune microenvironment or may be intrinsic to the malignant cells themselves. While these mechanisms of resistance are not yet well-understood, some evidence suggests that the mutational landscape, interferon signaling, or other immune-related signaling pathways active in malignant cells may be involved [2]. Elevated neoantigen load, particularly when common across subclones, is also associated with successful ICI treatment outcomes [51]. The presence of certain immune cell types within the tumor immune microenvironment seems to contribute to ICI treatment outcome as well; for example, patients with tumors exhibiting larger numbers of CD8+ T cells, CD4+ helper T cells, and natural killer cells tend to fare better while those with tumors infiltrated by CD4+ T regulatory cells and macrophages often have worse outcomes [70]. Additionally, recent research has highlighted the importance of various cell states exhibited by T cells and revealed the importance of a phenomenon known as T cell exhaustion in ICI therapy [71]. Notwithstanding these initial prognostic markers, an urgent need remains for more and concrete biomarkers to



effectively stratify patients into those likely to respond to ICI therapies and those who are not [59]. Additionally, understanding mechanisms of treatment response and failure is a critical step to the development of new treatments or combinations that overcome the limitations of those that are currently available. ICI therapy is also extremely expensive and can come with serious side effects, such as cytokine release syndrome, so the failure to control or eliminate a patient's disease is not the only potential downside to this therapeutic approach; these treatments must be administered only when a patient has a reasonable likelihood of benefit [72, 73, 74].

### **1.6.3 T cell exhaustion**

Since immune checkpoint inhibitors act on the inhibitory receptors that are often expressed by exhausted T cells, many researchers argue that understanding the process of T cell exhaustion is central to understanding the mechanisms of ICI response and resistance [71]. T cell exhaustion is a stable cell state characterized by reduced effector function, the expression of various inhibitory receptors, including immune checkpoints like PD-1, LAG3, TIM3, TIGIT, and CD160, and various metabolic changes and epigenetic alterations [75, 76, 77, 78, 79]. The extent of the epigenetic changes may determine whether a given T cell's state of exhaustion is reversible or not [78]. These cells also exhibit gene expression profiles that differ from healthy T cells in their various stages of differentiation [75]. T cell exhaustion was first studied in mice chronically infected with lymphocytic choriomeningitis virus (LCMV), but researchers have since discovered that a similar process can be observed in tumor-infiltrating lymphocytes (TILs) found in human cancers [80, 75]. T cell "exhaustion" is somewhat of a misnomer, since exhausted T cells still exert some control over an infection or tumor [71]. Since exhausted T cells exhibit some effector function, and an evolved mechanism against autoimmunity could have developed to simply delete these cells, some authors suggest that this phenotype with reduced, but not absent, effector activity and expression of these inhibitory receptors may have evolved to keep a chronic infection somewhat in check without the cost of significant tissue damage [81].

The exhaustion process is a disruption of the normal trajectory undergone by CD8+ T cells in response to pathogens. Under the conditions of a typical acute infection, naïve CD8+ T cells interact with their cognate non-self antigens in an inflammatory environment. These cell populations then expand and differentiate into effector cells, which target cells infected with the pathogen. When the pathogen is cleared, most of these cells apoptose and a fraction of them differentiate into long-lived central memory and effector memory T cells, which can then be re-activated, proliferate, and regain effector function upon second exposure to the same antigen [82, 75, 79]. If T cell receptors (TCRs) specific to a particular antigen are chronically stimulated, however, they can undergo this exhaustion process instead. This may be due to an infection or a developing tumor that cannot be fully eliminated and thus provides ongoing exposure of an antigen to the immune system. This, plus chronic interferon signaling, drives CD8 T cells into an exhausted state [71, 79]. T cell exhaustion is of significant interest to those grappling with the clinical problem of ICI therapy failure. T cell exhaustion is associated with worse patient outcomes in multiple cancer types, and exhausted CD8 T cells are the main intended target of ICI therapy, as well as a number of other immunotherapies [79].

Research in mice has identified two major cell states that fall into the category of T cell exhaustion: progenitor exhausted T cells and terminally exhausted T cells [71]. Progenitor exhausted T cells are somewhat similar to memory T cells in that they are long-lived and exhibit the capacity for self-renewal. Terminally differentiated exhausted T cells exhibit effector genes, while progenitor exhausted T cells do not. [71]. While terminally exhausted T cell subsets exhibit cytotoxic effects, they become non-functional over time and therefore have a limited capacity to drive sustained tumor control and elimination [79].

Whether the degree of T cell exhaustion found in human cancers is a progressive series of cell states that runs parallel to the degree of differentiation of these cells is not yet clear. Some suggest that exhaustion is initially reversible but becomes fixed with further epigenetic changes [71]. These two cell states are not always evenly distributed; progenitor exhausted CD8 T cells are primarily found in stroma, lymph nodes and peripheral blood, while terminally exhausted CD8 T cells are more likely to be found within the tumor [79]. Further research is needed to differentiate between the ICI therapy-induced migration of new CD8 T cell populations into the tumor microenvironment vs. the induction of cell state change in cells already residing within the tumor.

Most of the work done to characterize the process of T cell exhaustion has been done in mice with chronic LCMV infection [75, 82]. While this early work has been useful, there is a pressing need to characterize the cell states, epigenetics, and expression patterns associated with T cell exhaustion with data from human tumors rather than assuming that findings from the mouse LCMV research will generalize well to this new biological context [71].

#### **1.6.4 Using scRNA-seq to Study Immune Checkpoint Inhibitor Response and T Cell Exhaustion**

In the past several years, numerous experiments have used scRNA-seq to characterize the cell types present within a tumor and their relationships to ICI response and T cell exhaustion. Much of this work has been performed on samples from human melanoma tumors [83, 4, 5, 84]. There appears to be significant transcriptional heterogeneity in both malignant and immune cell populations, with some but not all cells exhibiting any particular inhibitory receptor or exhaustion signature of interest [83].

One analysis identified a resistance program related to CDK4/6 signaling in malignant cells within a melanoma tumor that was associated with less infiltration of T cells in the tumor immune microenvironment, which is known to be associated with worse response to ICI treatment. This resistance program was found to be up-regulated following treatment in patients with resistance to ICI therapy. By inhibiting this CDK4/6 signaling, researchers were able to improve anti-PD-1 treatment response in mice [4].

Another analysis using scRNA-seq identified TOX, a transcription factor, as a regulator of the expression of several immune checkpoints (PD-1, TIM-3, TIGIT, CTLA-4) in melanoma and non-small cell lung cancer (NSCLC). More recent research has confirmed that TOX drives the transition of progenitor exhausted CD8 T cells into a terminally exhausted state [79]. Expression of this transcription factor was also negatively associated with ICI treatment response; as a result, experimenters concluded that TOX inhibition may improve ICI efficacy and that considering its expression

may be useful for patient stratification efforts [85]. Yet another treatment approach attempts to use interleukin-10 signaling to promote self-renewal in terminally exhausted CD8 T cell populations [79].

Another analysis of scRNA-seq data from pre- and post-ICI melanoma tumors, focusing solely on immune cells, showed that some exhausted CD8+ T cells express TCF7 and share the self-renewal and stem-like cell state associated with memory T cells. This result is consistent with what has been discovered in mice, and the abundance of these TCF7+ CD8+ T cells was shown to be predictive of ICI response. The authors identified exhausted-like and memory-like T cell states but were not able to clearly identify the order and transitions between these cell states [5]. Pre-clinical research is ongoing to determine the efficacy of ectopic TCF-1 (the protein encoded by the TCF7 gene) expression and TOX knockout [79].

Yet another analysis used flow cytometry, RNA-seq, and assay for transposase-accessible chromatin by sequencing (ATAC-seq) to identify four transcriptionally and epigenetically distinct states of exhaustion in T cells. However, these data came from mice with chronic LCMV infections [86]. As has been mentioned, these findings may not generalize well to T cell exhaustion in human tumors, and further work is needed using human data to ascertain the relevance of these findings to ICI response [71]. That being said, while the delineation between progenitor exhausted and terminally exhausted CD8+ T cells has mostly been done in mice, an analysis of immune cells from human renal cell carcinoma tumors was able to identify these two immune cell populations [87, 88].

One analysis of particular interest obtained paired pre- and post-ICI skin tumor (basal cell and squamous cell carcinoma) samples from treatment responders and non-responders. The authors then performed scRNA-seq and also used T cell receptor sequencing (TCR-seq) to track distinct TCR clonotypes longitudinally throughout treatment. With this approach, they discovered large populations of CD8+ T cells present after treatment from clonotypes not found in the tumor beforehand, suggesting that ICI may recruit new T cells from outside the tumor immune microenvironment [3].

Finally, an investigation into T cell exhaustion was published in November 2022 that challenges and updates the current model of a linear cell state progression from progenitor exhausted to terminally exhausted CD8 T cells, at least in the murine viral infection context. In this work, authors developed a single-cell multi-omic atlas of T cell exhaustion in LCMV-infected mice, which included single-cell chromatin accessibility, transcriptome and T cell receptor (TCR) sequencing of LCMV antigen-specific CD8+ T cells. A bifurcating cell fate trajectory was observed, in which progenitor and intermediate exhausted CD8 T cells could go on to enter one of two downstream cell states. The classic terminally exhausted ( $\text{Tex}^{\text{Term}}$ ) CD8 T cell population was observed. They also found a killer cell lectin-like receptor (KLR)-expressing cytotoxic CD8 T cell ( $\text{Tex}^{\text{KLR}}$ ) population. The  $\text{Tex}^{\text{KLR}}$  cells had greater cytotoxic function and proliferative capacity than their  $\text{Tex}^{\text{Term}}$  counterparts. They also observed that TCR signaling strength may determine CD8 T cell differentiation trajectory, with high-avidity TCR clones (relative to viral epitopes) more likely to follow the terminal exhaustion path and low-avidity TCR clones more likely to end up in the KLR-expressing exhaustion state.

They suggest that cytokine and inhibitory receptor signaling may be involved in determining this cell fate decision as well [89].

To check for generalization beyond the mouse context, the Tex<sup>KLR</sup> gene signature was applied to scRNA-seq data of CD8 T cells from human tumors. They found a subpopulation of cells in the human data that showed up-regulation of the Tex<sup>KLR</sup> signature and down-regulation of the Tex<sup>Term</sup> signature from the mouse analysis, and another cell population that exhibited the opposite [89]. These findings suggest a need to refine the current understanding of the CD8 T cell exhaustion process and its relevance to immunotherapy.

These single-cell analyses have expanded our understanding of ICI response, resistance, and T cell exhaustion. Nevertheless, there are several key issues that have yet to be addressed. It is still unclear how cell-extrinsic factors, such as the gene expression of other immune cell types, contribute to T cell exhaustion. It is also unclear how gene expression in malignant cell subpopulations associated with ICI resistance changes in response to treatment. Past approaches have been overly gene-centric; a systems-level view that characterizes the pathway and gene regulatory network patterns related to ICI treatment outcome is therefore critically needed to advance understanding of this biological phenomenon.

Importantly, clinical interference in a highly complex system for determining self vs. non-self always comes with risks. The modification or removal of critical safeguards in the immune system can lead to autoimmunity which can pose a threat to patients already dealing with serious medical problems. Keeping this in mind, researchers must seek to maintain this delicate balance and proceed with caution.

## **1.7 Limitations and Open Questions**

There is a need to examine the impact of other cell types within the tumor immune microenvironment on the CD8 T cell exhaustion process. This is not straightforward with existing scRNA-seq analysis methods. Pseudotime analyses allow for the analysis of changes in molecular activity along a trajectory of changing cell states, but the consideration of how other, proximal cell types may be impacting this cell state change requires new methods and frameworks. Additionally, the ways in which various tumor subpopulations relate to T cell exhaustion and immunotherapy response must be considered. These heterogeneous cell populations may differentially relate to these processes in multiple ways; some tumor populations may show different antigen-presenting activity, which may affect the ability of tumor-antigen-specific CD8 T cells to identify and destroy these cells. Alternately, tumor subpopulations may release signals that affect the CD8 T cells directly and modify their ability to target all tumor cells, beyond that specific subpopulation. New approaches are needed to fully elucidate the ways in which CD8 T cells of different states, other immune cell types, and tumor cell subpopulations interact and determine ICI therapy response and resistance.

## 2 T cell exhaustion in the tumor immune microenvironment

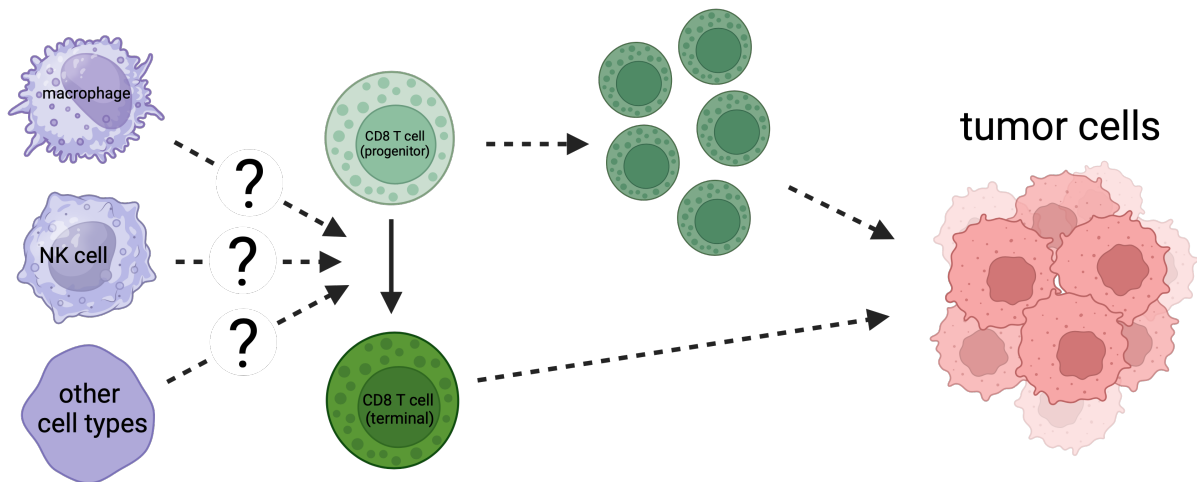
### ABSTRACT

While immune checkpoint inhibitors show success in treating a subset of patients with certain late-stage cancers, these treatments fail in many other patients as a result of mechanisms that have yet to be fully characterized. The process of CD8 T cell exhaustion, by which T cells become dysfunctional in response to prolonged antigen exposure, has been implicated in immunotherapy resistance. Single-cell RNA sequencing (scRNA-seq) produces an abundance of data to analyze this process; however, due to the complexity of the process, contributions of other cell types to a process within a single cell type cannot be simply inferred. We constructed an analysis framework to first rank human skin tumor samples by degree of exhaustion in tumor-infiltrating CD8 T cells and then identify immune cell type-specific gene-regulatory network patterns significantly associated with T cell exhaustion. Using this framework, we further analyzed scRNA-seq data from human tumor and chronic viral infection samples to compare the T cell exhaustion process between these two contexts. In doing so, we identified transcription factor activity in the macrophages of both tissue types associated with this process. Our framework can be applied beyond the tumor immune microenvironment to any system involving cell-cell communication, facilitating insights into key biological processes that underpin the effective treatment of cancer and other complicated diseases.

### 2.1 Introduction

Cancer immunotherapies show great potential for treating patients with otherwise incurable disease. Over the past decade, a type of immunotherapy known as immune checkpoint inhibition (ICI) has been approved to treat an increasing number of cancer types. However, in spite of the profound success of these treatments in a subset of patients, most patients do not exhibit durable response following ICI therapy [51, 90]. Furthermore, the subset of patients for which these therapies are effective cannot be reliably identified prior to treatment, and the mechanisms by which some patients exhibit resistance to ICI have yet to be fully characterized [2]. An urgent need remains for biomarkers to effectively stratify patients into those likely to respond to ICI therapies and those who are not [59]. Moreover, obtaining a better understanding of mechanisms of treatment response and failure is a critical step in the development of new treatments or combinations that overcome the limitations of those currently available.

Recent research has highlighted the importance of various cell states exhibited by CD8 T cells and revealed the importance of a phenomenon known as T cell exhaustion in ICI therapy [71]. The process of CD8 T cell exhaustion, by which CD8 T cells become dysfunctional in response to prolonged antigen exposure, can be observed in the tumor immune microenvironment and in chronic viral infection and has been implicated in immunotherapy resistance [71, 91]. Since immune checkpoint inhibitors act on the inhibitory receptors that are often expressed by exhausted T cells, and



**Figure 4: Relationship between immune cell activity and the degree of CD8 T cell exhaustion in the tumor immune microenvironment**

successful ICI response relies on the effector function of these cells, understanding the process of T cell exhaustion is central to understanding the mechanisms of ICI response and resistance [71].

Research in mice has identified two major cell states that fall into the category of T cell exhaustion: progenitor exhausted T cells and terminally exhausted T cells [71]. Progenitor exhausted T cells are somewhat similar to memory T cells in that they are long-lived and exhibit the capacity for self-renewal. They express high levels of the protein TCF7 and low levels of the protein TOX. Terminally exhausted T cells, on the other hand, express low levels of TCF7 and high levels of TOX [71, 91]. Most of the early work to characterize the process of T cell exhaustion has been performed in mouse models, using chronic lymphocytic choriomeningitis (LCMV) infection [75, 82]. While this early work has been useful, there is a pressing need to characterize the cell states, epigenetics, and expression patterns associated with T cell exhaustion with data from human tumors rather than assuming that findings from the mouse LCMV research will generalize well to this new biological context [71]. Newer research has explored this process further in human tissue, but deeper knowledge of this process is integral to overcoming ICI therapy resistance.

Cell state changes like T cell exhaustion occur within a complex molecular context involving changes within and external to the cell. Gene-regulatory networks (GRNs) consisting of transcription factors that regulate the expression of other genes control the states and activities of cells. These, in turn, are influenced by signal transduction involving molecular cascades that may originate in other cell types. Looking at these GRNs and cell-cell signaling patterns can allow us to take a system-level view of the cell state change, rather than considering each gene in isolation.

While a standard single-cell RNA sequencing (scRNA-seq) analysis pipelines allows for the analysis of cell-intrinsic gene expression and regulation changes associated with a given cell state change, contributions of other cell types to a process within a single cell type, such as the CD8 T cell exhaustion, are challenging to infer. To address this, we developed a novel computational framework. In brief, we first inferred a pseudotemporal ordering of CD8 T cells that

Dataset	GEO ID	Disease	Total cells	CD8 T cells
<i>Li et al., 2019</i>	GSE123814	melanoma	37561	19741
<i>Yost et al., 2019</i>	GSE123139	basal cell carcinoma	20954	8136
<i>Wang et al., 2020</i>	GSE157829	HIV	21941	15202

**Table 1:** scRNA-seq datasets. Three publicly available scRNA-seq datasets used – two from tumor immune microenvironment (melanoma, basal cell carcinoma) and one from chronic viral infection (HIV).

started with the progenitor exhausted CD8 T cells (TCF7 high, TOX low) and progressed gradually to the terminally exhausted CD8 T cells (TCF7 low, TOX high). We then calculated a sample-level CD8 T cell “exhaustion score,” based on the distribution of a sample’s CD8 T cells along this pseudotime trajectory. Subsequently, we used this score to rank samples from least to most exhausted. Cells from samples placed at the two extremes of this ranking were compared, one cell type at a time, to identify TFs that are significantly associated with the T cell exhaustion process (Figure 4). We applied our computational framework to scRNA-seq data generated from human melanoma samples [84]. To validate the results and identify common patterns, we also applied this framework to a basal cell carcinoma (BCC) dataset and a chronic HIV infection dataset [3, 92]. Overlap analysis between these datasets revealed common transcription factor (TF) patterns in multiple cell types in the tumor immune microenvironment associated with CD8 T cell exhaustion.

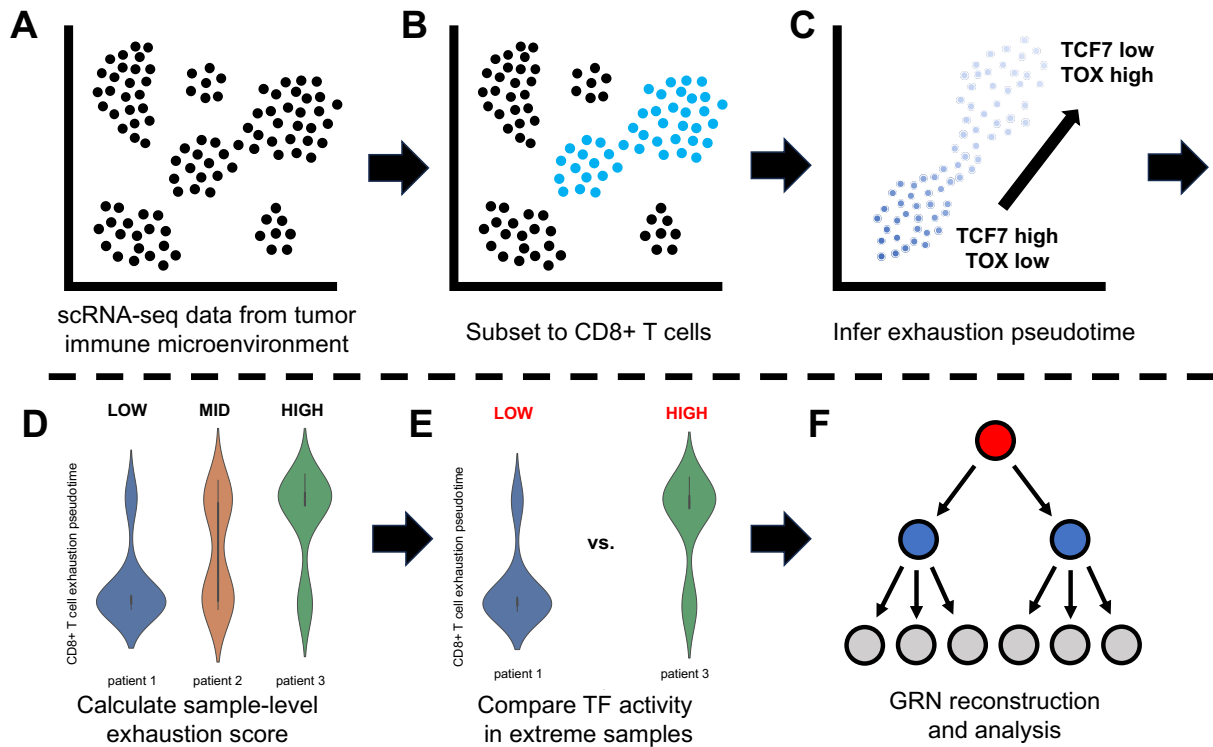
## 2.2 Results

### 2.2.1 Tumor immune microenvironment and chronic viral infection datasets

In order to investigate the relationship between CD8 T cell exhaustion and molecular activity in neighboring cell types in the tumor immune microenvironment and chronic viral infection, three publicly available single-cell RNA-seq (scRNA-seq) datasets were obtained (Tables 1, 10). The first dataset contains human melanoma tumor samples from sixteen patients [84]. The second dataset contains human BCC tumors from fourteen patients [3]. The third dataset contains peripheral blood mononuclear cell (PBMC) samples from six patients with chronic HIV infections [92]. These data were pre-processed using a standard scRNA-seq analysis pipeline (see Methods) and then analyzed with our novel pipeline, described below.

### 2.2.2 A novel computational framework to identify TFs that are significantly correlated with the T cell exhaustion process

To identify molecular mechanisms underlying the CD8 T cell exhaustion process in the tumor immune microenvironment, we have developed a novel analysis framework (Figure 5) to analyze scRNA-seq datasets. This framework consists of the following steps, applied after standard scRNA-seq pre-processing: subset to CD8+ cells, infer exhaustion pseudotime, calculate sample level exhaustion scores, infer TF activities for individual cells, conduct differential activity analysis for samples at the two extremes of the inferred T cell exhaustion trajectory, construct GRN, and then compare the patterns between different datasets. Some of the major steps in the pipeline are detailed below.



**Figure 5: Analysis overview** A) input – scRNA-seq data from skin tumor microenvironment; B) select CD8 T cells using marker genes; C) infer pseudotime trajectory from progenitor (TCF7 high, TOX low) exhausted to terminally (TCF7 low, TOX high) exhausted CD8 T cells using *Monocle3*; D) calculate sample-level exhaustion score, quantifying the distribution of cells along the exhaustion trajectory by sample; E) select highest- and lowest-scoring sample groups; compare transcription factor activity in these samples for individual cell types; F) perform gene-regulatory network analysis of exhaustion-associated transcription factors.

**2.2.2.1 Inference of CD8 T cell exhaustion pseudotime trajectory** After pre-processing and preliminary analysis, the datasets were subsetted to just CD8 T cells (Figure 5B). *Monocle3* trajectory inference was then used to find a pseudotemporal ordering of cells starting from the progenitor exhausted population, which exhibited high expression of TCF7 and low expression of TOX, and proceeding to a terminally differentiated exhausted population, which exhibited low expression of TCF7 and high expression of TOX, as well as increased expression of immune checkpoint genes PDCD1, LAG3, and TIGIT (Figure 5C) [93].

**2.2.2.2 Ranking samples by sample-level CD8 T cell exhaustion score** With a cell-level CD8 T cell exhaustion trajectory, differentially expressed genes and differentially active TFs in CD8 T cells across this trajectory can be determined. However, we sought to identify molecular activity in other cell types in the tumor immune microenvironment that is significantly related to CD8 T cell exhaustion as well. To achieve this, we developed a new approach analogous to Gene Set Enrichment Analysis [94], calculating the degree of CD8 T cell exhaustion for individual patient samples according to the distribution of each sample’s CD8 T cells along this exhaustion trajectory (Figure 5D). Afterward, samples were clustered by their exhaustion score and the highest- and lowest-scoring sample



groups, with CD8 T cell distributions skewed most heavily toward either the progenitor exhausted or terminally exhausted populations, were selected for downstream comparison (Figure 5E).

### **2.2.2.3 Differential TF activity analysis for high- and low-exhaustion sample groups in immune cell types**

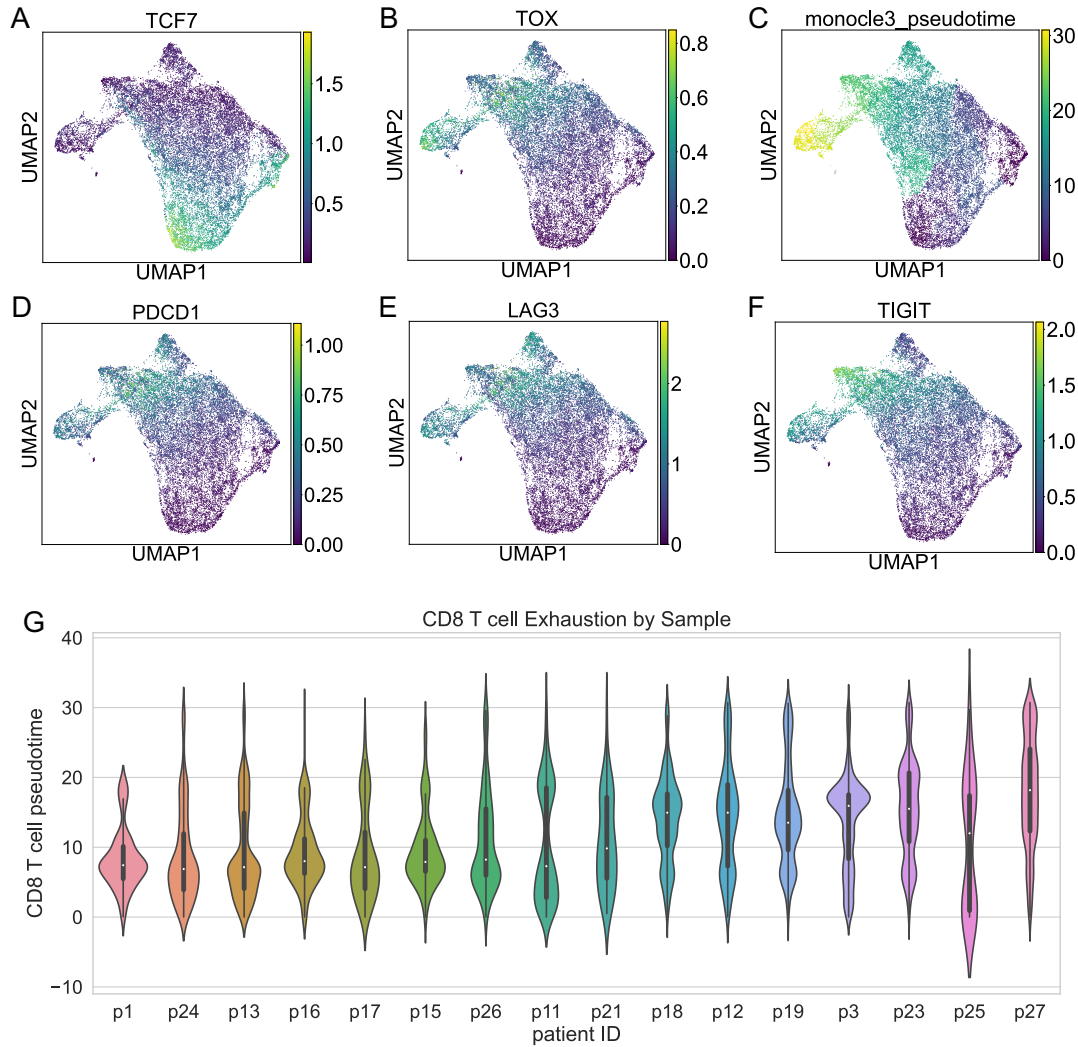
Once the samples on the far ends of the CD8 T cell exhaustion spectrum were identified, TF activity in other immune cell populations was then compared between these two groups. For the melanoma dataset, TF activity in CD8 T cells, macrophages, natural killer (NK) cells, B cells, and plasma cells was considered. For the BCC dataset, TF activity in CD8 T cells, macrophages, and NK cells was considered. For the HIV dataset, we looked at CD8 T cells, macrophages, B cells, and plasma cells.

### **2.2.3 Application of the computational framework to scRNA-seq datasets from human skin cancer and HIV infection**

We first applied this novel computational framework to a human melanoma scRNA-seq dataset (Figure 6) [84]. Once the CD8 T cells were identified, a pseudotime trajectory was inferred (Figure 6C, 20-22). The CD8 T cells in this dataset have a TCF7-high state (Figure 6A), where pseudotime starts, and transition gradually to a TOX-high state (Figure 6B). This transition also shows a progressive increase in expression of immune checkpoint genes PDCD1, LAG3, and TIGIT (Figures 6D-F, 20). The sample-level T cell exhaustion score analysis showed that samples p23, p25, and p27 have the highest T cell exhaustion scores and samples p1 and p24 have the lowest exhaustion scores (Figure 6G). The differential TF activity between these two sample groups (as quantified by *pyscenic* regulons and *AUC<sub>Cell</sub>*) was then compared.

We also applied this computational framework to a BCC dataset to investigate the generality of this pattern in the skin tumor microenvironment (Figure 19). Furthermore, we applied the framework to a chronic HIV infection dataset (Figure 7). Chronic viral infection data from human patients was included to enable the comparison of the T cell exhaustion process between the human tumor and viral infection contexts. The similarities and differences in the etiology and regulation of this cell state change between these two biological contexts is not yet fully clear. To address this, we compare and contrast the molecular profiles between the two.

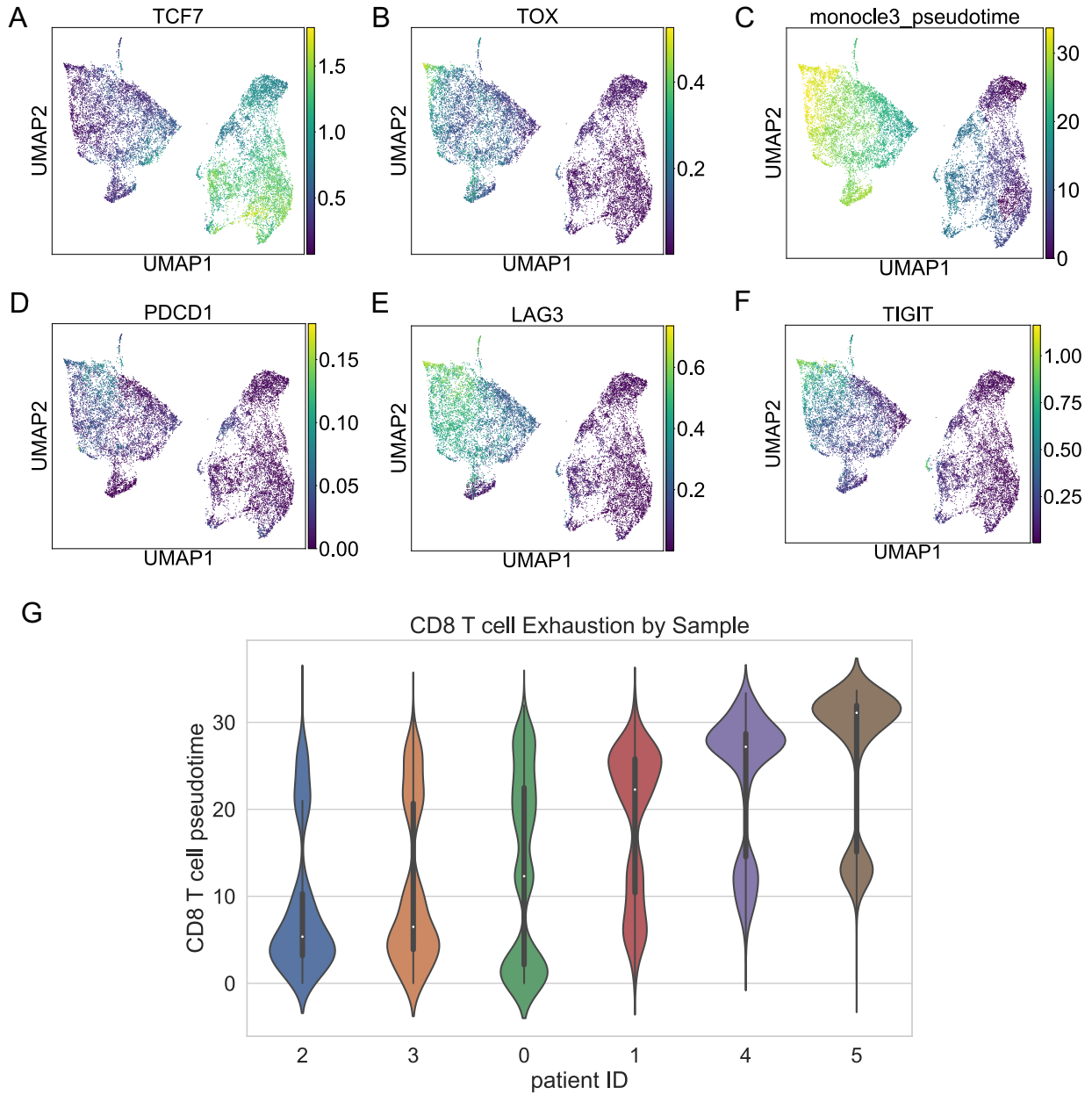
A similar trajectory of CD8 T cell state change was observed in these two datasets (Figures 19C, 7C, 23-28). We observed a similar trajectory of key markers of stemness, differentiation, and immune checkpoints (Figures 24, 27). For the BCC data, fourteen patient samples were scored and ranked, the high exhaustion group (samples su008, su013, su014) was compared against the low exhaustion group (002, 007) and differential TF activity between these groups was considered in each cell type (Figure 19G). For the chronic HIV data, six patient samples were scored and ranked and the high (samples 4 and 5) and low (samples 2 and 3) exhaustion groups were similarly compared (Figure 7G).



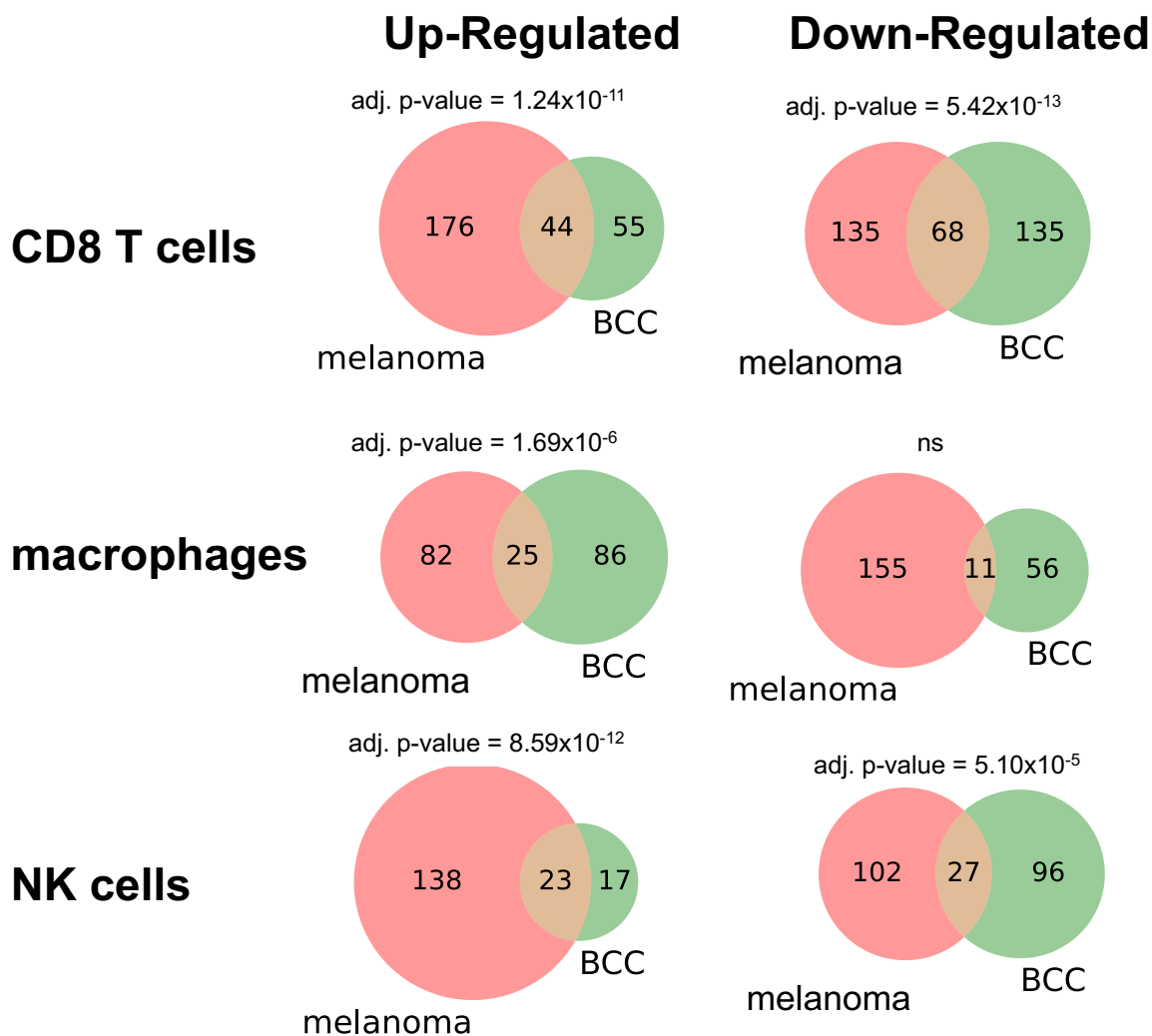
**Figure 6:** CD8 T cell exhaustion in human skin tumor samples. A, B, D-F) imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints LAG3, TIGIT, and PDCD1; C) Monocle3 pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells; G) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score.

## 2.2.4 Identification of gene regulatory patterns shared across skin tumor datasets that are related to T cell exhaustion

To uncover the cell type-specific transcription factor (TF) activity associated with the CD8 T cell exhaustion process that is consistently found in the skin tumor microenvironment, overlap analysis was performed via Fisher's exact test for results obtained from two skin cancer datasets (Methods). Cell types shared across both skin cancer datasets, including CD8 T cells, macrophages, and NK cells, were considered in the overlap analysis. We found significant overlap in TFs up-regulated (44 shared, adjusted p-value =  $1.24 \times 10^{-11}$ ) and down-regulated (68 shared, adjusted p-value =  $5.42 \times 10^{-13}$ ) in the CD8 T cells of the most exhausted samples, suggesting a shared mechanism in the T cell exhaustion process across skin tumor microenvironments (Figure 8A, B). Additionally, there was significant overlap in



**Figure 7:** CD8 T cell exhaustion in human HIV samples A, B, D-F) imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints LAG3, TIGIT, and PDCD1 C) Monocle3 pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells G) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score.



**Figure 8:** Overlap in exhaustion-associated transcription factor activity between melanoma and basal cell carcinoma (BCC) datasets. Overlap analysis was performed between melanoma and basal cell carcinoma datasets to identify the proportion of significantly exhaustion-associated TFs that were shared vs. non-shared and whether this overlap was significant. A-C, E-F) The two tumor datasets exhibit significant overlap in up- and down-regulation of CD8 T cell activity, up-regulation of macrophage activity, and up- and down-regulation of NK cell activity associated with CD8 T cell exhaustion. D) The overlap of TFs down-regulated in the most exhausted samples was not significant.

TFs up-regulated (25 shared, adjusted p-value =  $1.69 \times 10^{-6}$ ) in the macrophages of the most exhausted samples, as well as up- (23 shared, adjusted p-value =  $8.59 \times 10^{-12}$ ) and down-regulated (27 shared, adjusted p-value =  $5.10 \times 10^{-5}$ ) TFs in the NK cells (Figure 8C-F). However, we failed to observe a significant overlap of TFs down-regulated in macrophages from the most exhausted skin tumor microenvironments, which may suggest that the exhaustion-associated activity in macrophages mostly involves signaling pushing CD8 T cells toward the terminally exhausted state rather than holding them in the progenitor state (Figure 8D).

**2.2.4.1 Transcription factors up-regulated in terminally exhausted tumor-infiltrating CD8 T cells** We observed the up-regulation of certain CD8 T cell-specific transcription factor activity in the most exhausted samples in both tumor datasets, including IRF2, CTCF, E2F1, E2F2, E2F8, ETV7, and EOMES. Some of these have been found to be related to the CD8 T cell exhaustion process in previous experiments. IRF2 expression in tumor-infiltrating CD8 T cells has been shown to drive T cell exhaustion [95]. It pushes CD8 T cells toward the terminal exhaustion state by acting on TOX [79]. CTCF works with TCF1 as a cofactor to promote proliferative homeostasis in CD8 T cells [96]. It has been found to control the relative abundance of terminally exhausted CD8 T cells [97].

Members of the E2F transcription factor family are known cell cycle regulators. E2F1 is a regulator of cell cycle and proliferation; however, it may not be necessary for progression to the exhausted CD8 T cell state, based on viral experiments in mice [98]. E2F2 and E2F8 expression has been found in exhausted T cells in COVID-19 infection [99]. Additionally, E2F2 is most active in PD1+/CD39+ CD8 T cells [100].

ETV7 drives cell proliferation in some cell types [101, 102]. Some previous findings indicate that chronically stimulated CD8 T cells show ETV7 activity that down-regulates inflammation, potentially via ETS1 and TBX21, which we also found to be significantly up-regulated in the CD8 T cells of the most exhausted tumor samples in both datasets [103]. EOMES, which we found to be up-regulated in the most exhausted tumor samples, is a key regulator of CD8 T cell exhaustion as well, and binds competitively to the same sites as T-Bet (TBX21); however, the exact roles and functions of these factors is not yet clear [104, 105, 106, 107]. ETS1 is known to be important in regulating T cell state [108, 109]. Additionally, ETS1 knockdown leads to increased anti-tumor cytotoxic activity of CD8 T cells [110].

**2.2.4.2 Transcription factors down-regulated in terminally exhausted tumor-infiltrating CD8 T cells** We also observed the down-regulation of certain CD8 T cell-specific transcription factor activity in the most exhausted samples in both tumor datasets, including LYL1, BCL11A, PLAG1, and BCL-6. Several of these are known to be responsible for maintaining the stem-like state of the progenitor exhausted CD8 T cells. LYL1 is responsible for stem / progenitor state and self-renewal in thymocytes and T cells [111, 112, 113]. BCL11A slows differentiation of T cells to maintain self-renewal [114]. It is required for normal lineage development of lymphocytes [115, 116]. PLAG1 helps to maintain self-renewal in hematopoietic stem cells [117]. BCL-6, acting in opposition to PRDM1 (encodes BLIMP-1), maintains the relative proportion of progenitor (as opposed to terminally differentiated) exhausted CD8 T cell populations; TCF1, which drives self-renewal in progenitor exhausted CD8 T cells, activates BCL-6 and suppresses PRDM1 [118, 119].

**2.2.4.3 Down-regulation of Kruppel-like factor (KLF) signaling in terminally exhausted tumor-infiltrating CD8 T cells** We observed down-regulation of KLF2, KLF3, and KLF4 signaling in the CD8 T cells of the most exhausted melanoma samples, with KLF3 and KLF4 down-regulated in the BCC samples as well. This activity concords with the findings of previous mouse LCMV research [120]. The most exhausted HIV samples also exhibited down-regulation of KLF3 and KLF4; however, they showed up-regulation of KLF2, as did the BCC data.

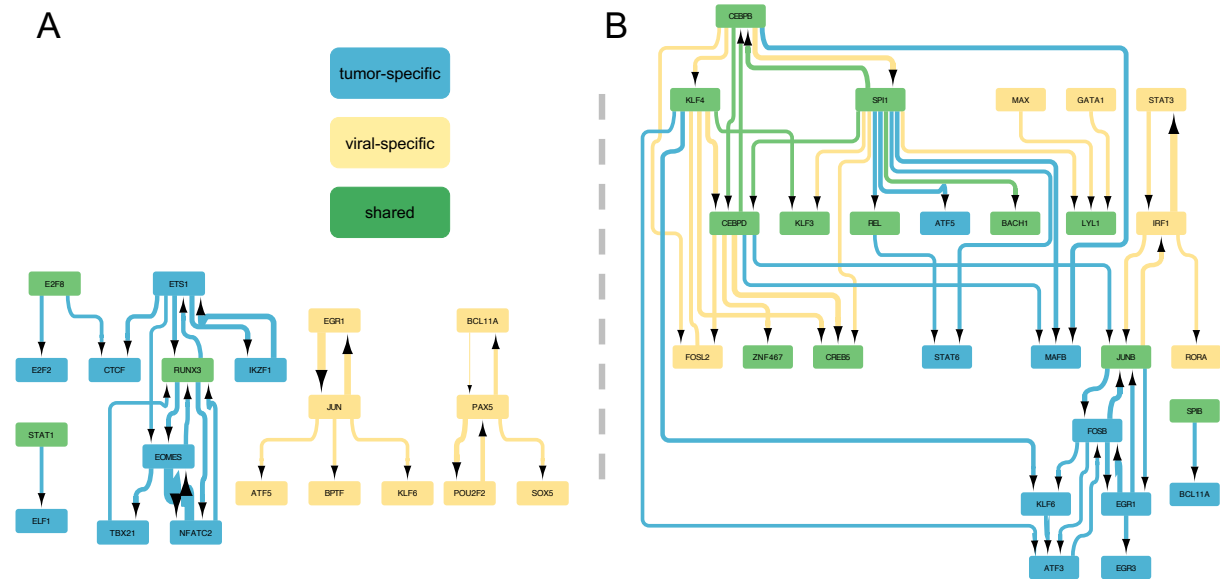
It is worth noting that KLF4 has been linked to increased effector function in these terminally exhausted CD8 T cell subsets [121]. This concords with the finding that the most exhausted samples exhibited down-regulation of KLF4 in their CD8 T cells and also had poorer response to immunotherapy. We also observed down-regulation of AP-1 signaling factors such as FOS, FOSB, JUN, JUNB in CD8 T cells across multiple datasets, the activity of which has been linked to KLF4 signaling in other work [121].

## **2.2.5 Transcription factors up-regulated in macrophages of highly exhausted tumor samples**

In addition to identifying TF activity in CD8 T cells associated with T cell exhaustion, some of which was in concordance with previous findings, we also identified TF activity in macrophages associated with the degree of CD8 T cell exhaustion in a sample. In both tumor datasets, significant up-regulation of TBX21 (T-Bet), PRDM1, RUNX3, CREM, IRF1, FOXP3, STAT2, NFKB1 and NFKB2 was observed. Some of these transcription factors, including TBX21 and PRDM1, also exhibit up-regulation in the most exhausted CD8 T cells themselves, suggesting that these two cell types may be responding to some of the same signals in the exhausted tumor immune microenvironment. Significant up-regulation of NFKB subunits NFKB1 and NFKB2 was observed in the macrophages of the most exhausted samples across all three datasets. Additionally, NFKB subunit RELB was up-regulated in the most exhausted BCC and HIV samples, with RELA up-regulated in the melanoma macrophages. These results suggest the up-regulation of NF Kappa-B signaling in macrophages of both highly exhausted tumor and HIV patients.

## **2.2.6 Gene-regulatory network modules associated with CD8 T cell exhaustion**

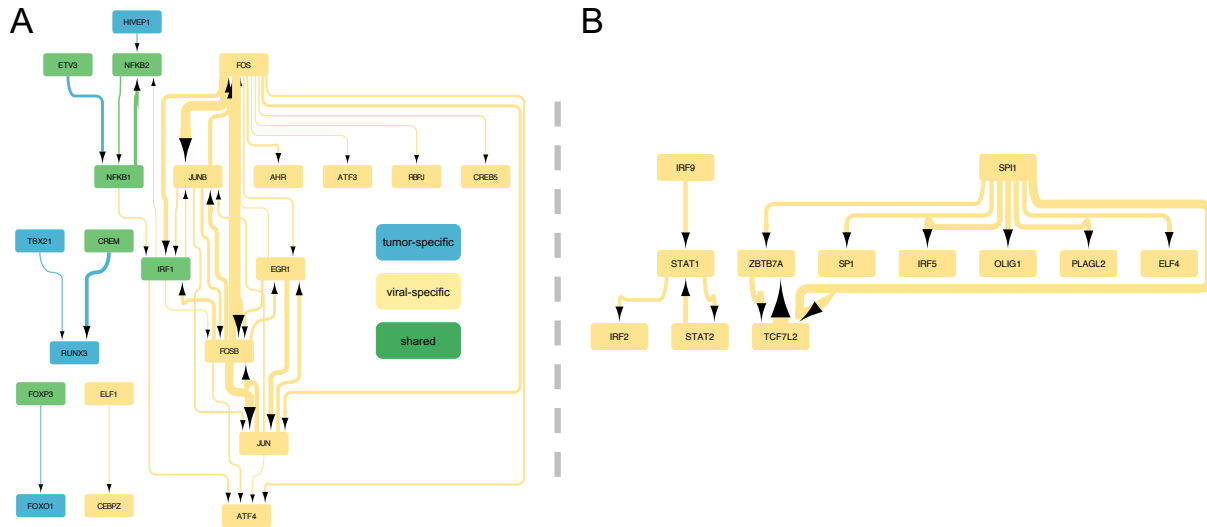
**2.2.6.1 Exhaustion-associated CD8 T cell GRNs** To better understand the regulatory relationships among TFs associated with CD8 T cell exhaustion, we reconstructed gene-regulatory network modules using a subset of the highly exhaustion-associated TFs whose activity in CD8 T cells was up-regulated in the most exhausted samples (Figure 9A). Regulatory relationships between these significant TFs were inferred using *pyscenic* gene regulatory network inference, which is based on co-expression between a transcription factor and its potential target followed by search for a TF's binding motif in the promoter region of a potential target (Methods). Edge widths represent the strength of the inferred relationship. We identified a larger module consisting primarily of tumor-specific exhaustion-associated TF regulation, which include TBX21 and EOMES. We also identified two up-regulated GRN modules comprising viral-specific exhaustion-associated TFs. In one, EGR1 activates JUN (forms AP-1 complex with FOS), which then activates ATF5, BPTF, and KLF6. The other includes BCL11A, PAX5, POU2F2, and SOX5.



**Figure 9:** Exhaustion-related transcription factor activity in CD8 T cells Exhaustion-associated transcription factors (graph nodes) and their regulatory relationships, inferred with pyscenic (graph edges) A) transcription factors up-regulated in the CD8 T cells of the most CD8 T cell-exhausted immune microenvironments B) transcription factors down-regulated in the CD8 T cells of the most CD8 T cell-exhausted immune microenvironments.

The same GRN module reconstruction was performed for TFs whose activity was found to be down-regulated in the CD8 T cells of the most exhausted samples (Figure 9B). This module included some TF down-regulation shared across the tumor and viral contexts, including KLF3 and KLF4, as well as tumor- and viral-specific activity. Interestingly, KLF6 was up-regulated in the most exhausted viral samples (adjusted p-value: 0.0401) but down-regulated in the most exhausted tumor samples (adjusted p-values for melanoma, BCC:  $7.34 \times 10^{-7}$ ,  $6.51 \times 10^{-54}$ ), suggesting a potential difference in mechanism behind these cell state trajectories.

**2.2.6.2 Exhaustion-associated macrophage GRNs** GRN module reconstruction was also performed for exhaustion-associated TF activity in macrophages (Figure 10). We identified a large GRN module of up-regulated TF activity in viral patients, including TFs related to FOS/JUN/AP-1 signaling, which was connected to an NFKB module via IRF1. This NFKB / IRF1 activity was found to be up-regulated in the most exhausted samples across both tumor datasets and the viral dataset, indicating that NFKB activity in macrophages may be associated with CD8 T cell exhaustion in both disease contexts. We also observed tumor-specific up-regulation of TBX21 and RUNX3, whose activity is associated with exhaustion in the CD8 T cells themselves as well. A number of transcription factors and complexes were found to be significantly up-regulated in the macrophages of the most exhausted samples across all 3 datasets, including NFKB, IRF1, BHLHE40, FOSL2, NFIL3, and CREM. Previous research indicates that NFKB regulates IRF1 [122, 123]. Additionally, there is cross-regulation among NFKB, AP-1, and IRF1 signaling [123]. We also identified two viral-specific GRN modules of TFs down-regulated in the macrophages of the most exhausted



**Figure 10:** Exhaustion-related transcription factor activity in macrophages Exhaustion-associated transcription factor activity in macrophages (graph nodes) and their regulatory relationships, inferred with pyscenic (graph edges) A) transcription factors up-regulated in the macrophages of the most CD8 T cell-exhausted immune microenvironments B) transcription factors down-regulated in the macrophages of the most CD8 T cell-exhausted immune microenvironments.

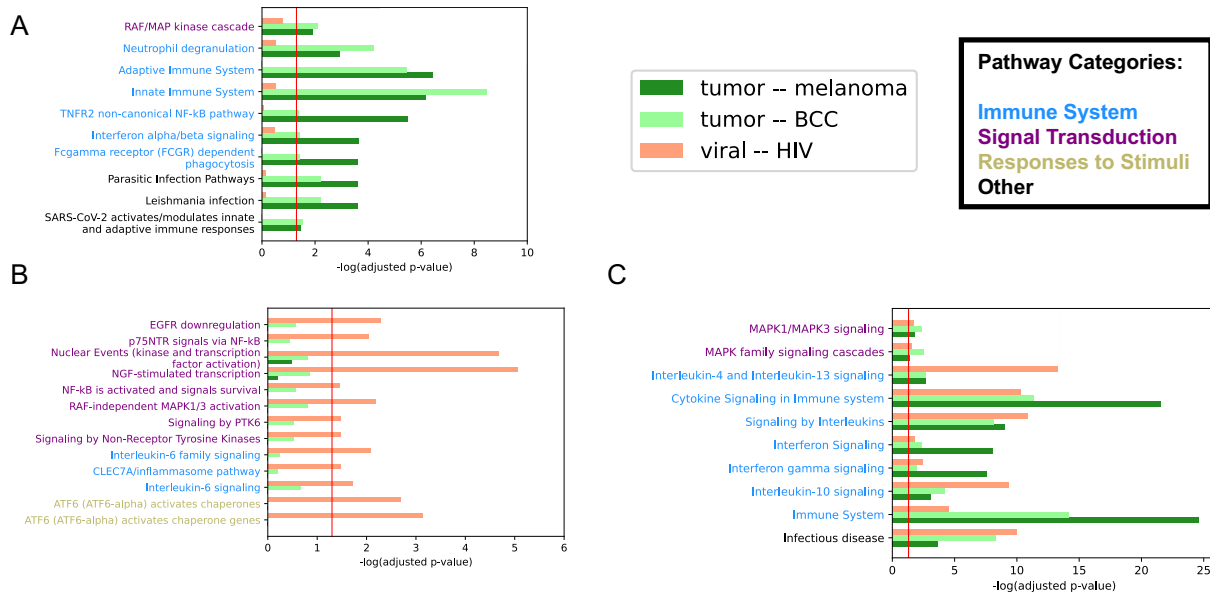
samples, which included members of the interferon regulatory factor (IRF2, IRF5, IRF9) and STAT (STAT1, STAT2) families.

### 2.2.7 Identification of pathways significantly related to T cell exhaustion in macrophages

In order to identify biological pathway activity in macrophages that may play a significant role in the exhaustion process, we performed pathway enrichment analysis on differentially expressed genes in the macrophages of high- and low-exhaustion samples using *Reactome* (Figure 11, Methods) [33, 35, 34, 32]. We observed some tumor-specific exhaustion-associated pathways in macrophages, including a number of immune pathways. We saw an increase in TNFR2 non-canonical NF- $\kappa$ B signaling in the tumor context and an increase in p75NTR signals via NF- $\kappa$ B in the viral context, which aligns with our observation of an up-regulation of NFKB1 and NFKB2 in the most exhausted samples across the tumor and viral contexts, but indicates that the mechanism here may be slightly different.

In both tumor and viral infection, we observed up-regulation of the interleukin 10 signaling pathway (Figure 11C). Importantly, we see up-regulation of NFIL3 in the macrophages of high-exhaustion samples (see above). Published results show that NFIL3 drives IL-10 activity, further supporting the possible causal relationship between NFIL3 and the IL-10 pathway [124, 125]. While the role of IL-10 on CD8 T cell state and activity appears to be multifaceted, this anti-inflammatory cytokine may drive exhaustion of CD8 T cells via signaling by PRDM1, which was identified as an up-regulated TF in the high-exhaustion macrophages [126, 127, 128]. We also observed some MAPK-related signaling pathways up-regulated in exhaustion-associated macrophages, such as MAPK family signaling cascades, RAF/MAP kinase cascades, MAPK1/MAPK3 signaling, and RAF-independent MAPK1/3 activation. In addition to this, we also





**Figure 11:** Pathway analysis results of exhaustion-associated DEGs in macrophages. Pathways enriched in genes over-expressed in macrophages of high-exhaustion samples relative to low-exhaustion samples; red line = adjusted p-value cutoff of 0.05 A) tumor-specific exhaustion-associated pathway activity B) viral-specific exhaustion-associated pathway activity C) exhaustion-associated pathway activity shared between tumor and viral contexts.

observed up- and down-regulation of a number of immune, cell cycle-related, and other pathways in the CD8 T cells themselves (Figures 32-37).

## 2.3 Discussion

The promising yet incomplete success of immune checkpoint inhibitors has led to significant interest in and scrutiny of the CD8 T cell exhaustion process due to its association with treatment outcome. Progress in this research area can change the way cancer and chronic infection are treated. This process has garnered much attention in recent years as many attempts to find ways to modulate this process, with the hope of improving immunotherapy outcomes and expanding the population of potential responders to this class of treatments through the development of drugs that modulate the tumor immune microenvironment to improve the likelihood of response to ICI therapy.

To better understand the molecular mechanisms of the CD8 T cell exhaustion process, we have developed a new computational approach to analyze scRNA-seq datasets collected from tumor immune microenvironments. The advantage of this framework lies in its ability to consider the potential contribution of gene or TF activity in multiple cell types to a change in cell state in a cell type of interest, rather than only considering activity within that single cell type. Existing tools allow us to determine which genes and transcription factors are differentially active across the exhaustion pseudotime trajectory; our approach produces a sample-level exhaustion score, therefore allowing the differential activity analysis of transcription factors in other cell types along with CD8 T cells. By applying our novel computational framework to single-cell transcriptomics data from human tumor and viral infection samples, we identified the potential contribution of various immune cell types to this process. This work contributes to ongoing

efforts that seek to understand how, when, and why ICI therapy succeeds and fails and, ultimately, to improve clinical outcomes in patients with late-stage cancers.

Our novel computational approach successfully identified TFs in CD8 T cells associated with CD8 T cell exhaustion, implying their potential involvement in immune checkpoint inhibitor response. Many of these TFs recapitulate published results from previous investigations of this process. By examining the literature, we verified the importance of transcription factors such as TBX21 (T-bet), BLIMP1, the E2F family, and the KLF family. More importantly, this approach also identified TFs in other cell types, including macrophages, which sheds light on the state of the surrounding tumor immune microenvironment and how it may contribute to CD8 T cell exhaustion. In the macrophages of samples with the greatest degree of CD8 T cell exhaustion, across both tumor datasets and the HIV data, we found up-regulation of NF $\kappa$ B signaling, as well as IRF1, BHLHE40, FOSL2, NFIL3, and CREM. Some of these TFs form a gene-regulatory network module (Figure 10A). Additionally, NFIL3 has the potential to regulate the activity of a key cytokine, IL-10, which has been linked to the CD8 T cell exhaustion process by previous research [126, 127, 128]. The activity of cytokines like IL-10 may partially explain the association between the abundance of macrophages in the tumor immune microenvironment and ICI response [129]. In addition to the up-regulation of IL-10 regulator NFIL3 itself in exhaustion-associated macrophages, we also observed up-regulation of genes in the IL-10 signaling pathway in exhaustion-associated macrophages. This provides further support for the association between IL-10 signaling in the tumor immune microenvironment and CD8 T cell exhaustion. The activity of this anti-inflammatory cytokine as it relates to CD8 T cell exhaustion is worthy of further investigation. By identifying cell type-specific TF activity in the tumor immune microenvironment associated with the degree of CD8 T cell exhaustion in a patient sample, and by organizing these relevant TFs into a gene-regulatory network that can be studied at the system level, we contribute to a clearer understanding of this critical biological process.

This work fulfills the need to compare and contrast the CD8 T cell exhaustion process between the tumor immune microenvironment and the chronic viral infection context in human tissue as well. We found some exhaustion-associated GRN modules in CD8 T cells and macrophages that were viral-specific, while others were a combination of tumor-specific and shared. As these results come from human data, they are more relevant to the human disease context than earlier findings from LCMV in mice, which they may confirm or contradict. In CD8 T cells, we observed up-regulation of two viral-specific GRN modules, with one related to AP-1 signaling. We also observed viral-specific up-regulation of AP-1-related TF activity in exhaustion-associated macrophages, as well as down-regulation of an IRF9-regulated GRN module. One caveat to bear in mind when comparing these tumor and viral samples is that the difference in tissue context (peripheral blood vs. skin tumor microenvironment) likely drives some of the observed differences in expression patterns.

In addition to providing insights into the CD8 T cell exhaustion process, this computational pipeline can also be applied to other biological questions. This approach allows researchers to consider the potential contribution of other cell types to a cell state change of interest, whether in cancer, development, or another biological context. While we used this functionality to identify potential drivers of CD8 T cell exhaustion throughout the tumor immune microenvironment, it

can be applied to any process wherein a continuous trajectory of cell states can be identified in scRNA-seq data and mapped in pseudotime. By moving to the sample level and ranking samples based on the distribution of their cells across a trajectory, we build on currently available methods to add this important perspective.

To summarize, this work illuminates some key differences between T cell exhaustion in human tumors and chronic infections and reveals gene-regulatory networks in immune cell populations within a tumor that may mediate response to immune checkpoint inhibition therapy. The novel framework we developed also allows researchers to consider the potential contribution of other cell types to a cell state change of interest, giving it wider applicability in the field of single-cell transcriptomics research.

## 2.4 Data Availability

The three scRNA-seq datasets used in this paper have been previously published, and are available via Gene Expression Omnibus: *GSE123814* [3], *GSE123139* [84], *GSE157829* [92] – <https://www.ncbi.nlm.nih.gov/geo/>

## 2.5 Code Availability

Code used in this paper can be located in the following GitHub repository:

<https://github.com/christopher-klocke/exhaustion-microenvironment>

DOI: <https://doi.org/10.5281/zenodo.13786867>

## 2.6 Results Availability

Hosted at *zenodo* with DOI: <https://doi.org/10.5281/zenodo.10088918>

## 2.7 Acknowledgements

This research was supported by the US NLM Training Grant (*T15-LM007088*, CK) and US NIH (*U41 HG003751*, *U24HG012198* and *U01CA239069*, GW) grants. The authors would also like to thank Brett Davis and Nathaniel Evans for discussion of original ideas.

## 2.8 Methods

### 2.8.1 Datasets

Publicly available datasets were downloaded from Gene Expression Omnibus using the following identifiers: *GSE123814* [84], *GSE123139* [3], *GSE157829* [92]. The *Li et al.* dataset contains human melanoma samples from 16 patients. The *Yost et al.* dataset contains human basal cell and squamous cell carcinoma samples from 15 patients. The *Wang et al.* dataset contains human peripheral blood mononuclear cell (PBMC) samples from six HIV-infected patients. For the *Yost et al.* dataset, only pre-treatment (with immune checkpoint inhibition) samples were used.

## 2.8.2 Pre-processing, dimensionality reduction, batch correction, and clustering

Data files were unzipped and converted to the *AnnData* format [130] using in-house Python and Bash scripts and stored as .h5ad files. Datasets were subsetted to just protein-coding genes per *UniProt* [131]. Pre-processing was performed using the standard *scanpy* (1.9.1) [130] pipeline for scRNA-seq data, including quality filtering, normalization, and log-transformation steps. Doublets were identified and removed using *DoubletDetection* (4.2) [132]. Dimensionality reduction was performed with PCA. In order to account for potential batch effects between samples, batch correction was performed with *Batchelor* (1.14.1) [133]. Clustering was performed using the Leiden algorithm (*leidenalg* 0.8.10) [134]. The dimensionality was further reduced to two dimensions for visualization using UMAP (*umap-learn* 0.5.3) [135, 136].

## 2.8.3 Cell cluster annotation

*MAGIC* (*magic-impute* 3.0.0) [137] was used to impute expression for marker genes collected from databases and the literature. Cell types were manually assigned to clusters based on the following marker genes for the *Li et al.* and *Wang et al.* data sets: CD8 T cells: CD3, CD8; CD4 T cells: CD3, CD4; regulatory T cells: CD3, CD4, FOXP3; macrophages: CD68, CD163, CD14; natural killer cells: CD56, CD16, CD3, KLRB1, NKG7, NKG2A, GZMB, NCAM1; B cells: CD19, CD27, CD38; plasma cells: SDC1, CD20. We used the previous cell type annotation for *Yost et al.* dataset.

## 2.8.4 Inference of CD8 T cell exhaustion trajectory with *Monocle3*

Datasets were subsetted to just CD8 T cell clusters for trajectory inference. A pseudotemporal ordering of cells from progenitor exhausted to terminally exhausted was inferred for CD8 T cells using *Monocle3* (1.3.1) [93]. Root cells were chosen from the principal graph based on expression of key exhaustion-associated genes, including TCF7 (high), TOX (low), AND LAG3 (low).

## 2.8.5 Calculation of sample-level exhaustion scores with the cell enrichment approach

To assess the distribution of a sample's CD8 T cells along the exhaustion pseudotime path relative to the overall distribution (across all samples), a cell enrichment approach was used. To do this, the Gene Set Enrichment Analysis (GSEA) method was repurposed (*gseapy* 0.9.5) [94]. In GSEA, an enrichment score is calculated by quantifying the over-representation of a pre-defined set of genes in the most highly differentially expressed genes between two conditions. This is often a comparison in expression levels between two treatment conditions in bulk gene expression data.

To calculate the cell enrichment score, the same mathematical framework is used as in GSEA, but cell sets and a cell ranking are substituted for gene sets and a gene ranking. Whereas GSEA focuses on gene sets, the cell enrichment approach considers cell sets, where a set of cells comprises all cells of a given type derived from one patient sample. That sample's cells are compared to the trajectory as a whole, constructed from many samples. In this way, the degree to which a subset of cells (derived from a given sample) is enriched at one end of a ranking of cells (e.g. exhaustion

pseudotime across all samples) is quantified. The resulting score captures the degree to which a sample's CD8 T cells are biased toward the progenitor exhausted or terminally exhausted state.

### **2.8.6 Sample clustering and selection of extremes**

We used the Jenks natural breaks algorithm (*jenksy 0.3.2*) to divide the samples into clusters based on similar exhaustion scores [138]. Sensitivity analyses were performed to ensure reasonable robustness to the choice of cluster number. The sample clusters with the highest and lowest scores were chosen for further analysis. By comparing the extreme samples from the ends of the distribution, we increased power to detect changes in other cell types associated with these different immune microenvironments. Additionally, by using multiple samples from each end, we minimized the effect of individual patient background and increased the number of cells used to calculate the results. For every dataset, at least two samples were used at each end of the sample ranking.

### **2.8.7 Quantifying transcription factor regulon activity with SCENIC**

We used the *SCENIC* gene-regulatory network inference algorithm, implemented in the *pyscenic (0.11.2)* Python library. Briefly, *GRNBoost* was used to infer correlation modules, which were then pruned by *CisTarget* to obtain regulons, each consisting of a transcription factor and its inferred, directly regulated targets (*arboreto 0.1.6*). The activity of each regulon was then quantified for each cell using *AUCell* [47, 139, 140].

### **2.8.8 Comparison of transcription factor activity between exhaustion-low and exhaustion-high samples in each cell type**

*AUCell* scores for a given regulon were used as a proxy for the activity of the corresponding transcription factor. To determine whether a given transcription factor was significantly up- or down-regulated in a given cell type in one sample group relative to the other, the following comparison was made: using a Mann-Whitney U test (*scipy 1.9.0*), the distributions of TF activity scores were compared between the two sample groups. Moving to the sample level for this comparison allows for the assessment of potentially exhaustion-associated activity in other cell types within the tumor immune microenvironment. P-values were then corrected using the Benjamini-Hochberg method.

### **2.8.9 Overlap analysis**

Once exhaustion-associated cell-type specific TF activity was identified for each dataset, the results were compared across datasets. We used a Fisher's exact test to perform an overlap analysis, determining whether a given cell type had shared significant up- or down-regulated TFs between two datasets. The resulting p-values were corrected using the Bonferroni method.

### 2.8.10 Analysis of exhaustion-associated gene-regulatory sub-networks

TF-target relationships inferred using *pyscenic* were used to connect significant TFs into gene-regulatory sub-networks. Two thresholds were then used to generate the GRN figures. Importance scores were exported from the *GRNBoost* results for each TF-target relationship. If both tumor datasets had an inferred GRN graph edge for a TF-target relationship at a score above 1 but the viral dataset did not, the edge was labeled as tumor-specific. If the viral score was above one but neither of the tumor scores were above 1, it was listed as viral-specific. If all 3 had a score above 1, the edge was considered to be shared across these biological contexts. The GRN was then pruned using the higher threshold; if the maximum score for a potential TF-target edge exceeded the higher threshold, the edge and its associated nodes were included in the final network visualization. This threshold varied from 25 to 30 for different GRN figures. All GRN visualizations were created using *Cytoscape* (v3.9.1) [141]. A subset of the top-scoring TFs were displayed in the GRN figures; a full list of exhaustion-associated TFs and their p-values can be found in the Supplemental Results files.

### 2.8.11 Pathway enrichment analysis with *Reactome*

Using *Reactome* (v85) pathway gene sets and pathway hierarchy, we performed a pathway enrichment analysis on genes differentially expressed (per two-sided Mann-Whitney U test, *scipy*) in CD8 T cells and macrophages between the high- and low-exhaustion samples [33, 35, 34, 32]. A Fisher's exact test was used to quantify overrepresentation of a pathway's genes in these differentially expressed genes. Out of 20376 total protein-coding genes (per *UniProt* v2022\_01), any gene with an adjusted p-value less than 0.05 was considered to be differentially expressed. These were then tested for overrepresentation in each pathway's gene set.

### 2.8.12 Other package versions:

*matplotlib*=3.5.1; *matplotlib-base*=3.5.1; *networkx*=2.8.5; *numpy*=1.22.3; *numpy-base*=1.22.3; *pandas*=1.4.2; *python*=3.10.4; *scikit-learn*=1.1.2; *seaborn*=0.11.2

### 3 T cell exhaustion and immunotherapy response

#### ABSTRACT

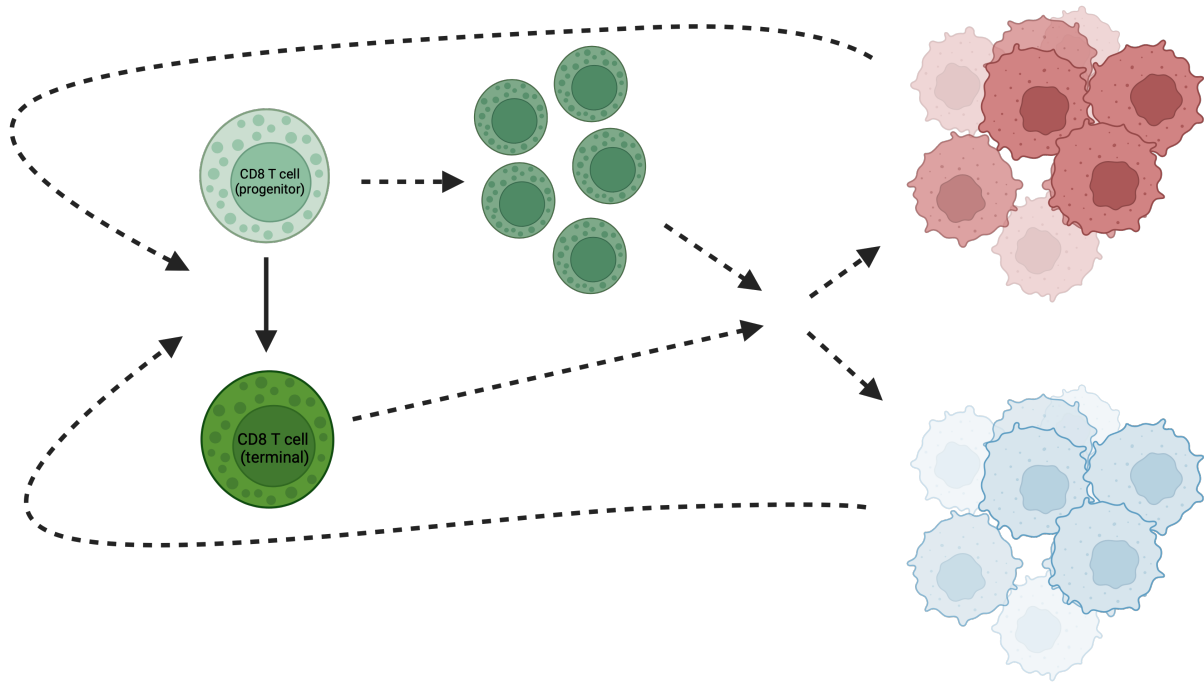
The process of CD8 T cell exhaustion has been studied in the context of immune checkpoint inhibitor therapy. However, its role in immunotherapy response is not fully clear. To better address these questions, we developed an analysis framework that combines our previously developed sample score framework with another tool called Scissor, which enables the use of clinically labeled bulk RNA-seq data to select subpopulations of tumor cells associated with clinical outcome. By combining these two tools, cellular processes like T cell exhaustion can be related to subpopulations of tumor cells and their relationships with a clinical phenotype such as immunotherapy response. This framework was successfully applied to human lung cancer data and was able to identify some unexpected relationships between the tumor immune microenvironment and immunotherapy response. This analysis framework may be applied to any biological question where single-cell expression data and clinically labeled bulk expression data are available, and in which subpopulations of a heterogeneous cell type need to be compared to a cell state change in another cell type as well as a clinical phenotype.

#### 3.1 Introduction

Immune checkpoint inhibitor (ICI) therapy is a class of cancer immunotherapy that has shown success in treating patients with some late-stage cancers, including non-small cell lung cancer (NSCLC) and melanoma. Despite some success, these therapies fail in a significant proportion of patients; only 20-30% of patients responded in the first applications of ICI treatment to cancer [142, 143]. One biological process that has been studied in connection with responsiveness to ICI therapy is CD8 T cell exhaustion. This process occurs when CD8 T cells in the tumor immune microenvironment are chronically exposed to tumor antigen over a prolonged period and develop a decrease in anti-tumor cytotoxic activity.

Exhausted CD8 T cells, which typically express immune checkpoints proteins, exist in several states: progenitor exhausted CD8 T cells exhibit high TCF7, low TOX, and exhibit self-renewal and a stem-like cell state [71]. The abundance of these TCF7+ CD8+ T cells has been shown to be predictive of ICI response [5]. Terminally exhausted T cells are more differentiated than the upstream progenitor exhausted population, exhibit low TCF7, high TOX, and have reduced (but still present) anti-tumor cytotoxic activity[71].

There are two main ways in which tumor subpopulations may have a differential effect on ICI response (Figure 12). A group of tumor cells may exhibit behavior by which just that subpopulation is made more or less susceptible to immune recognition and destruction. A down-regulation MHC expression leading to decreased tumor antigen presentation that facilitates greater immune evasion is one such example. Alternately, a group of tumor cells may signal to immune cells



**Figure 12: Heterogeneity in tumor subpopulation modulation of and response to anti-tumor immune activity**

such as CD8 T cells, either directly or via a modulation of other activity within the tumor immune microenvironment, and thus have an impact on the degree to which all cells within the tumor are targeted.

Studying the relationship between CD8 T cell exhaustion and immunotherapy response is challenging for several reasons. The tumor microenvironment is heterogeneous, both within and across cell types. Single-cell RNA sequencing (scRNA-seq) can help to break down this heterogeneity by allowing for the identification of different cell types and even multiple cell states within a given cell type. However, the number of patients represented in a given scRNA-seq dataset is typically small. Additionally, these datasets often do not contain the clinical annotations needed to link cell states and processes to the phenotype of interest. Bulk RNA-seq datasets often have larger sample sizes and clinical labels of interest. By making use of clinically labeled bulk RNA-seq data to identify tumor sub-populations in single-cell data related to a phenotype of interest, the complementary strengths of these two sources of data may be harnessed.

This leaves a final challenge. The consideration of how one target cell type is related to cell state change in another reference cell type is not straightforward using standard scRNA-seq analysis frameworks. We previously developed an approach to address this challenge, which quantifies cell state change at the sample level and then makes comparisons between groups of samples with very different sample-level scores (see Chapter 2), within each cell type of interest. This enables comparisons within these other cell types (e.g. tumor cells) that relate to the distribution of cell states within the reference cell type (e.g. CD8 T cells).



To better investigate the link between CD8 T cell exhaustion and ICI treatment response, a novel analysis framework was developed. It makes use of this previously developed approach for quantifying a balance of cell states at the sample level (for consideration of associated molecular activity in other cell types). It also utilizes a tool called *Scissor*, leveraging clinically labeled bulk RNA-seq data to select subsets of phenotypically-associated single cells [144]. By pairing these two methods, we can answer questions that cannot be addressed by either approach individually, or by any other existing approach. This framework was then applied to data from lung and skin tumor patients in order to better understand how CD8 T cell exhaustion in the human tumor context relates to ICI response. This framework is capable of indicating which molecular activity within the target cell type (e.g. tumor cells) is related to the clinical phenotype (e.g. immunotherapy response) via its relationship with the cell state change of interest in the reference cell type (e.g. exhaustion trajectory in CD8 T cells), versus through some other biological process or relationship.

This framework can also be applied to biological and clinical questions beyond cancer immunotherapy. It can be used for any analysis in which the association of a target cell type to a cell state change of interest in a different cell type is considered in the context of a particular clinical outcome. This is done by considering the relationships between the cell state change-associated subpopulations and the phenotype-associated subpopulations.

## **3.2 Results**

### **3.2.1 A novel computational framework to identify cell subpopulation activity related to cell state transition and clinical phenotypes**

In order to investigate the relationship between lung tumor subpopulations, T cell exhaustion, and ICI response, a novel analysis framework was developed (Figure 13). While the use case of interest here involved T cell exhaustion and immunotherapy response, the framework can be applied to any biological question in which there is a cell state trajectory of interest and another cell type that may relate to this cell state change and a clinical phenotype of interest. The following is a brief overview of the steps used to accomplish this.

**3.2.1.1 Identification of suitable data for application of the framework** To make use of this framework, several types of input data must be identified. scRNA-seq data from enough patients to set up a sample-level ordering is needed. The necessary sample size depends in part on how well the relevant cell types are distributed across samples, as well as on how significantly the distribution of cells across the relevant cell state trajectory varies by sample. Additionally, bulk RNA-seq data from the correct tissue context, labeled with the clinical outcome of interest, are required. Ideally, at least two bulk RNA-seq datasets will be used; the first to select phenotype-associated single-cells and the second (or more) to provide additional support for the validity of this selection.

**3.2.1.2 Pre-processing of single-cell data** Standard scRNA-seq and bulk RNA-seq pre-processing steps are used. If multiple single-cell datasets are used, they are merged, and batch correction is performed to integrate across samples. Following clustering and cell type annotation, the data are subsetted down to the cell types of interest: one in which the

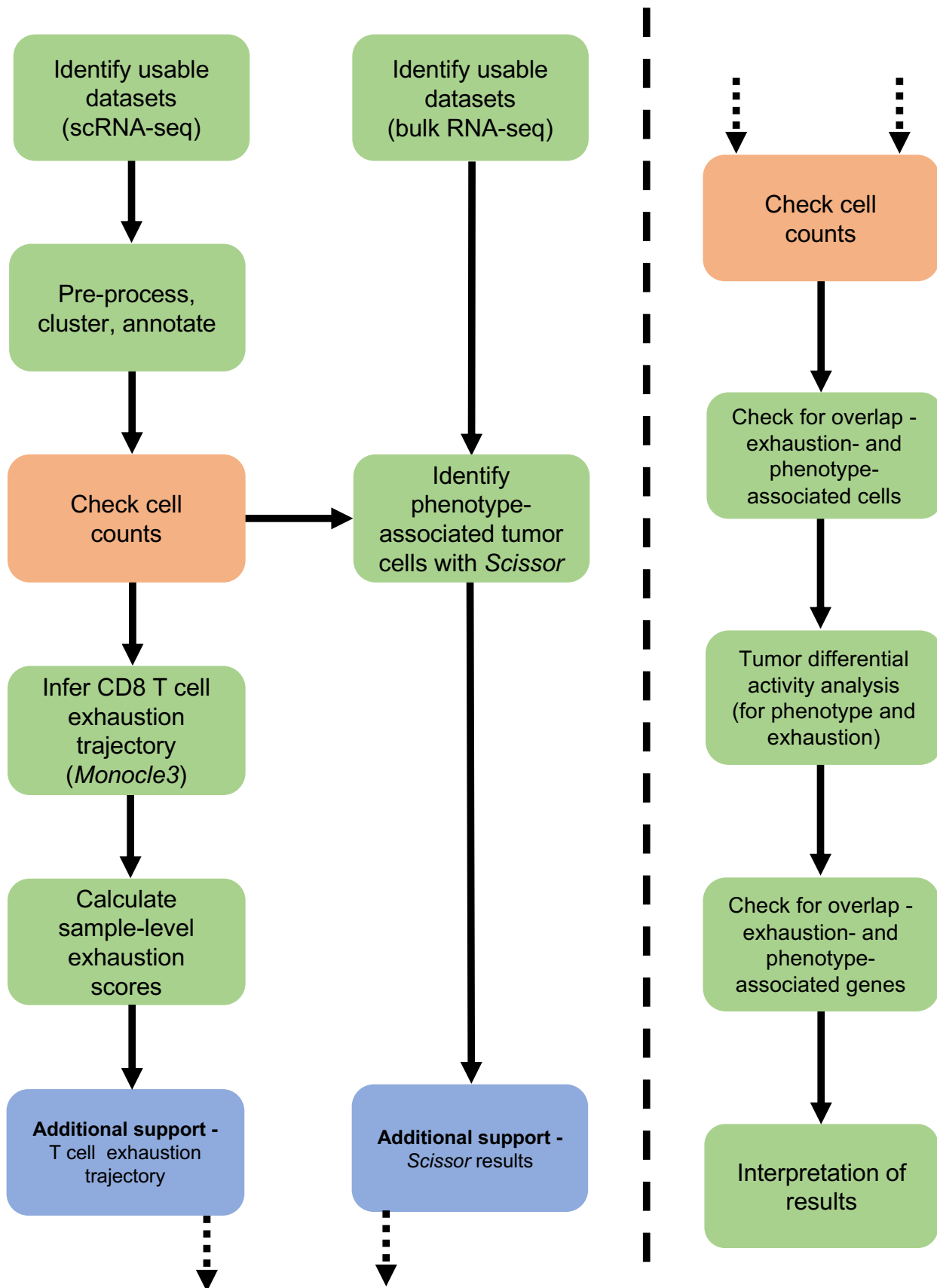


Figure 13: Overview of the analysis workflow

cell state change is occurring, and the other in which cell state-associated cell subsets (associated with state change in the other cell type, e.g. CD8 T cells) and phenotype-associated cells will be selected.

**3.2.1.3 Inference of cell state trajectory** In the data subset containing the cell type that is changing state, the beginning cell state is identified using marker gene visualization and domain knowledge. This beginning state is used as a root to initialize a trajectory inference algorithm, such as Monocle3, which finds a path from the beginning cell state to the final cell state.

**3.2.1.4 Evaluating the cell state trajectory and providing additional support** Once a cell state change trajectory has been inferred, it should first be evaluated by considering the change in expression of key marker genes associated with each cell state relative to pseudotime. In general, a monotonic decrease of marker genes associated with the beginning state and a monotonic increase of marker genes associated with the end state should be observed. There may be cases where more complex patterns are observed if certain genes are associated with an intermediate state between the two states of interest, or are transiently expressed to drive the change. Existing domain knowledge should be leveraged here. The simple case is a linear cell state change trajectory, but modifications can be made to accommodate branching and other trajectory patterns.

Ideally, if the data are available, differential gene expression activity between the two cell states of interest should be compared with other datasets where these cell states have been compared previously. Significant overlap provides further support for the claim that the same biological process is driving the trajectory of expression states.

**3.2.1.5 Calculation of sample-level cell state scores and selection of extremes** The cell state trajectory comprises cells from a number of samples. Each sample will not have cells distributed along the trajectory in the same way. Some samples will have a distribution of cells that is biased toward the early part of the trajectory, while others will be biased toward the later end. In order to quantify each sample's distribution of cells along the trajectory, a sample-level cell state score is calculated using a previously established method (see Methods). These sample-level scores are then split into groups and the two groups containing the lowest and highest scores are chosen for further comparison.

**3.2.1.6 Encoding of clinical phenotype** The bulk RNA-seq data used to select phenotype-associated subpopulations in the single-cell data must be labeled with a clinical phenotype of interest, which can be encoded as a binary or continuous variable.

**3.2.1.7 Identification of phenotype-associated cell subpopulations with *Scissor*** The *Scissor* algorithm is then used to select phenotype-associated single cells based on expression similarity with the labeled bulk RNA-seq data.

**3.2.1.8 Additional support for phenotype association of selected cell subpopulations** In order to ensure that the selected cells are truly associated with the clinical phenotype of interest, additional data should be leveraged where possible. If additional clinically labeled bulk RNA-seq data from the same tissue type are available, these data can be used as a primary validation. The differentially expressed genes (DEGs) between the two phenotype-associated single-cell populations (in the case of a binary clinical phenotype) should overlap significantly with genes differentially expressed between the clinical categories in the bulk data. Additionally, if any of the single-cell samples are annotated with the same clinical phenotype of interest, these samples should be checked to ensure they are enriched for cells associated with the matching clinical category. Finally, any other data available to support this selection should be considered. For example, if the clinical phenotype of interest is response to a particular cancer treatment, then response-associated single-cells should be over-represented in the pre-treatment relative to post-treatment samples.

**3.2.1.9 Comparison of cell state-associated and phenotype-associated cell subpopulations** Once cell state-associated and clinical phenotype-associated cell subpopulations were identified, the resulting target cell subsets were then compared. First, two sets of differentially expressed genes were calculated. Expression differences between the early trajectory-biased and late trajectory-biased target cell subpopulations, based on the distribution of reference cell states from the corresponding samples along the trajectory, were quantified. Differential expression between the two phenotype-associated cell subsets was also calculated. An overlap analysis was then performed to determine how many genes are associated with one or several of these four categories (early trajectory, late trajectory, phenotype 1, phenotype 2). Next, the relationship between sample-level trajectory scores and the relative proportions of each class of phenotype-associated cells within each sample was considered.

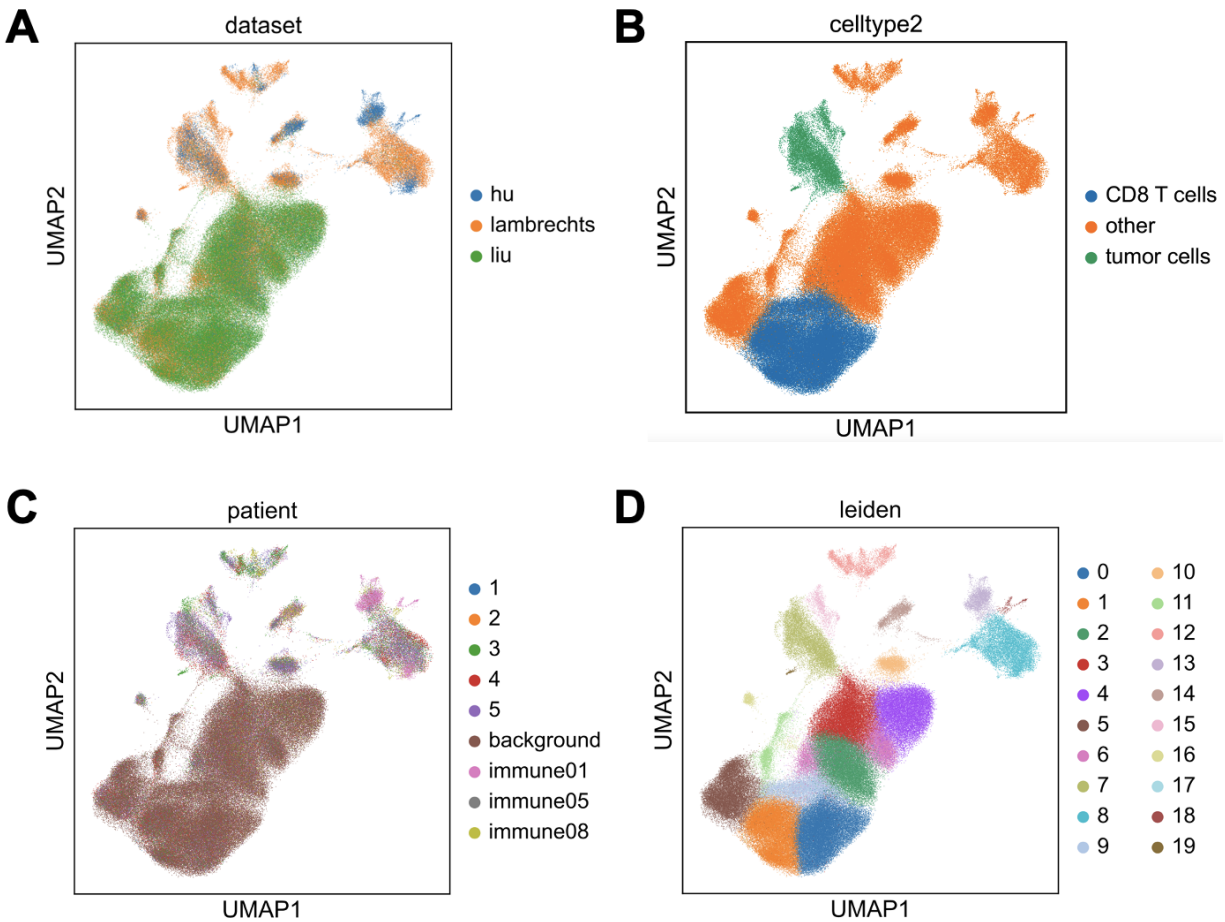
**3.2.1.10 Pathway analysis** After differential gene activity between the cell state scores and the phenotype association groups was identified, the Reactome pathway knowledgebase was used to identify pathways enriched for these differentially expressed genes. Pathways associated with cell state distributions and clinical outcomes were identified.

### **3.2.2 Application of framework to NSCLC tumor immune microenvironment data**

In order to evaluate the efficacy of this framework and obtain a clearer understanding of the relationship between CD8 T cell exhaustion and ICI response, the framework was applied to publicly available lung cancer data. To do this, multiple bulk and single-cell RNA-seq datasets from patients with NSCLC tumors were identified (see Methods). The bulk samples were collected prior to treatment with immunotherapy and are labeled with ICI response phenotypes.

### **3.2.3 Data integration, pre-processing, annotation, and subsetting**

First, the three single-cell NSCLC datasets were integrated, batch-corrected, and pre-processed (Figure 14a, Table 2, Methods). CD8 T cell and tumor cell clusters were identified in the resulting combined lung single-cell dataset (Figure



**Figure 14: Lung tumor data – merged datasets** A) three NSCLC datasets integrated for analysis B) CD8 T cell and tumor cell populations were identified C) eight samples in the merged dataset contain tumor cells in addition to CD8 T cells and other immune cell types; cells labeled “background” are from 33 tumors from the *Liu et al., 2022* dataset, which does not contain sufficient tumor cells for analysis – these T cells are included as a reference to help delineate the CD8 T cell exhaustion trajectory D) Leiden clustering of merged dataset

Dataset	GEO ID	Tumor Type	Total cells	CD8 T cells	Tumor cells
<i>Lambrechts et al., 2018</i>	<i>see Methods</i>	NSCLC	51232	10445	7218
<i>Hu et al., 2023</i>	GSE207422	NSCLC	13457	1145	2157
<i>Liu et al., 2022</i>	GSE179994	NSCLC	137660	56781	–
All datasets	–	NSCLC	<b>202349</b>	<b>68371</b>	<b>9375</b>

**Table 2: scRNA-seq datasets** Three publicly available scRNA-seq datasets from human NSCLC tumors were used

14b-d, Methods). The data were then subsetted to a CD8 T cell-specific dataset and a tumor cell-specific dataset. The resulting CD8 T cell and tumor datasets were then re-clustered (Figures 15 and 16).

**3.2.3.1 Cell counts checks** In the combined single-cell NSCLC dataset, the two cell types of interest, CD8 T cells and tumor cells, are well-distributed across samples, with a number of samples containing hundreds or thousands of each cell type (Table 3). This indicates that these data are suitable for downstream analysis.

### 3.2.4 CD8 T cell exhaustion trajectory and sample-level exhaustion scores

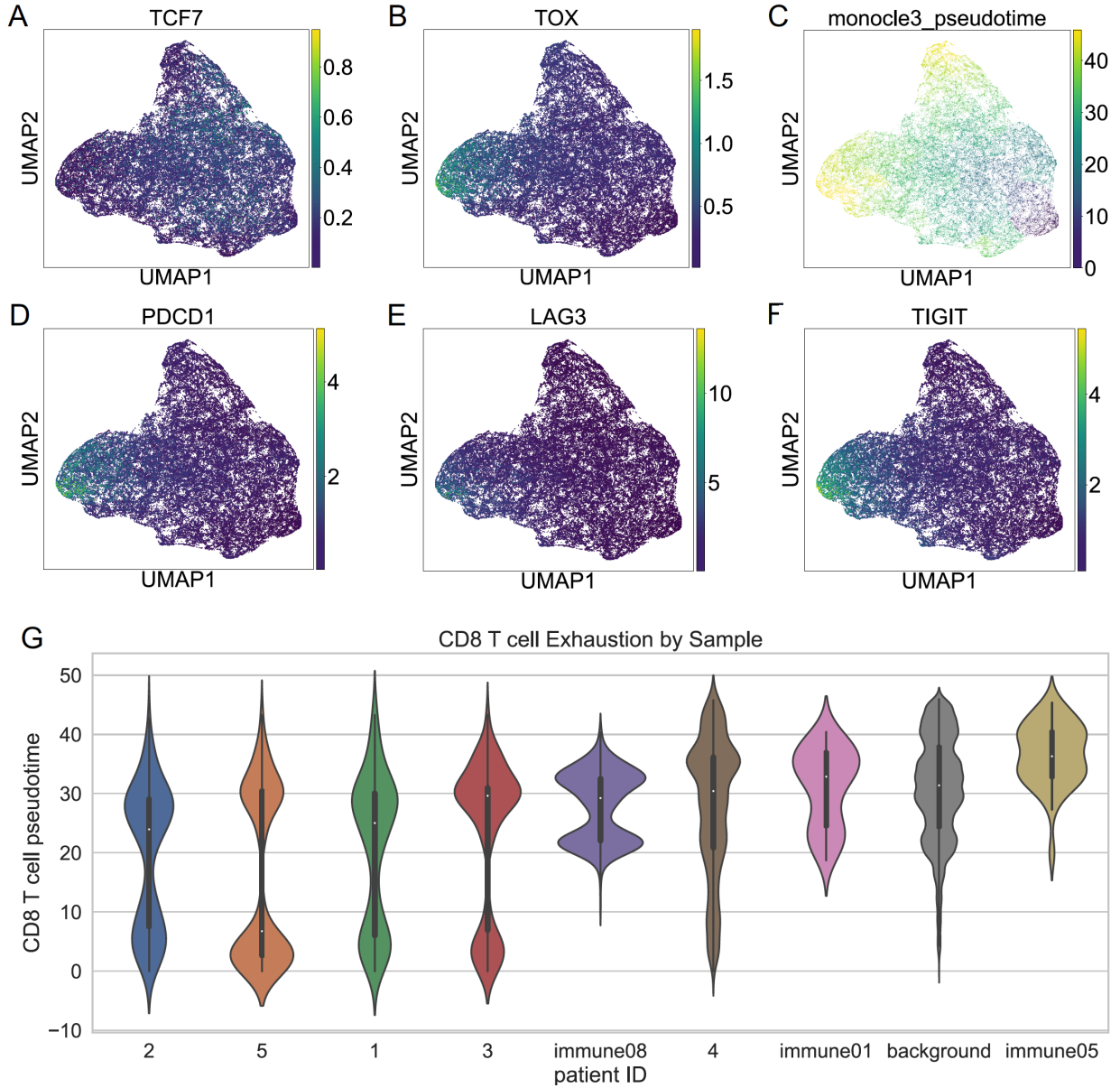
**3.2.4.1 CD8 T cell exhaustion trajectory** Next, a trajectory of CD8 T cells was inferred, progressing from a TCF7-high, progenitor exhausted state to a TOX-high, terminally exhausted state (Figure 15a-f, Methods).

**3.2.4.2 Sample-level exhaustion scores** In order to determine the relative degree of CD8 T cell exhaustion at a sample level, the distribution of cells along this exhaustion pseudotime trajectory was quantified with sample-level exhaustion scores. Samples were then ranked by exhaustion score and the most extreme samples (least exhausted and most exhausted by CD8 T cell distribution) were identified (Methods). Samples 2 and 5 were designated as “exhaustion-low,” with CD8 T cell distributions relatively biased toward the progenitor exhausted cell state. Sample immune05 was designated as “exhaustion-high,” with a CD8 T cell distribution biased toward the terminally exhausted cell state (Figure 15g, Methods).

**3.2.4.3 Support for the exhaustion trajectory – part 1** To support the validity of the exhaustion trajectory of CD8 T cells, the change in expression level of several exhaustion-related genes relative to pseudotime was considered. Average expression of TCF7 in CD8 T cells decreased monotonically along the trajectory, while average expression of

sample	dataset	cell type	
		CD8 T cells	tumor cells
1	<i>Lambrechts et al., 2018</i>	408	286
2	<i>Lambrechts et al., 2018</i>	368	162
3	<i>Lambrechts et al., 2018</i>	2247	1446
4	<i>Lambrechts et al., 2018</i>	4757	2459
5	<i>Lambrechts et al., 2018</i>	2665	2865
immune01	<i>Hu et al., 2023</i>	206	669
immune05	<i>Hu et al., 2023</i>	215	1478
immune08	<i>Hu et al., 2023</i>	724	10
background (36 samples)	<i>Liu et al., 2022</i>	56781	–
total	–	68371	9375

**Table 3: Cell counts by sample**



**Figure 15: CD8 T cell exhaustion in human skin tumor samples** A, B, D-F) imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints PDCD1, LAG3, and TIGIT; C) *Monocle3* pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells; G) exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score.

TOX, PDCD1, LAG3, and TIGIT generally increased along the trajectory (Figure 51). These expression trends are in concordance with previous study of these cell states [71].

**3.2.4.4 Support for the exhaustion trajectory – part 2** To provide further support to the validity of the exhaustion trajectory and sample-level exhaustion scores, we compared the overlap of differentially expressed genes in the CD8 T cells of exhaustion-high and exhaustion-low samples across several datasets (Methods, Tables 4-7) (see Chapter 2 for skin tumor dataset details).

		<b>DEGs – melanoma data</b>	
		up-regulated	not significant
<b>DEGs – NSCLC data</b>	up-regulated	922	2034
	not significant	2446	9538

**Table 4: *Li et al.* dataset, exhaustion high group, CD8 T cells** – test statistic = 1.77, p-value =  $1.77 \times 10^{-34}$

		<b>DEGs – melanoma data</b>	
		up-regulated	not significant
<b>DEGs – NSCLC data</b>	up-regulated	128	575
	not significant	1404	12833

**Table 5: *Li et al.* dataset, exhaustion low group, CD8 T cells** – test statistic = 2.03, p-value =  $4.24 \times 10^{-11}$

		<b>DEGs – BCC data</b>	
		up-regulated	not significant
<b>DEGs – NSCLC data</b>	up-regulated	979	1489
	not significant	2389	10083

**Table 6: *Yost et al.* dataset, exhaustion high group, CD8 T cells** – test statistic = 2.77, p-value =  $1.91 \times 10^{-99}$

		<b>DEGs – BCC data</b>	
		up-regulated	not significant
<b>DEGs – NSCLC data</b>	up-regulated	97	279
	not significant	1435	13129

**Table 7: *Yost et al.* dataset, exhaustion low group, CD8 T cells** – test statistic = 3.18, p-value =  $2.92 \times 10^{-18}$



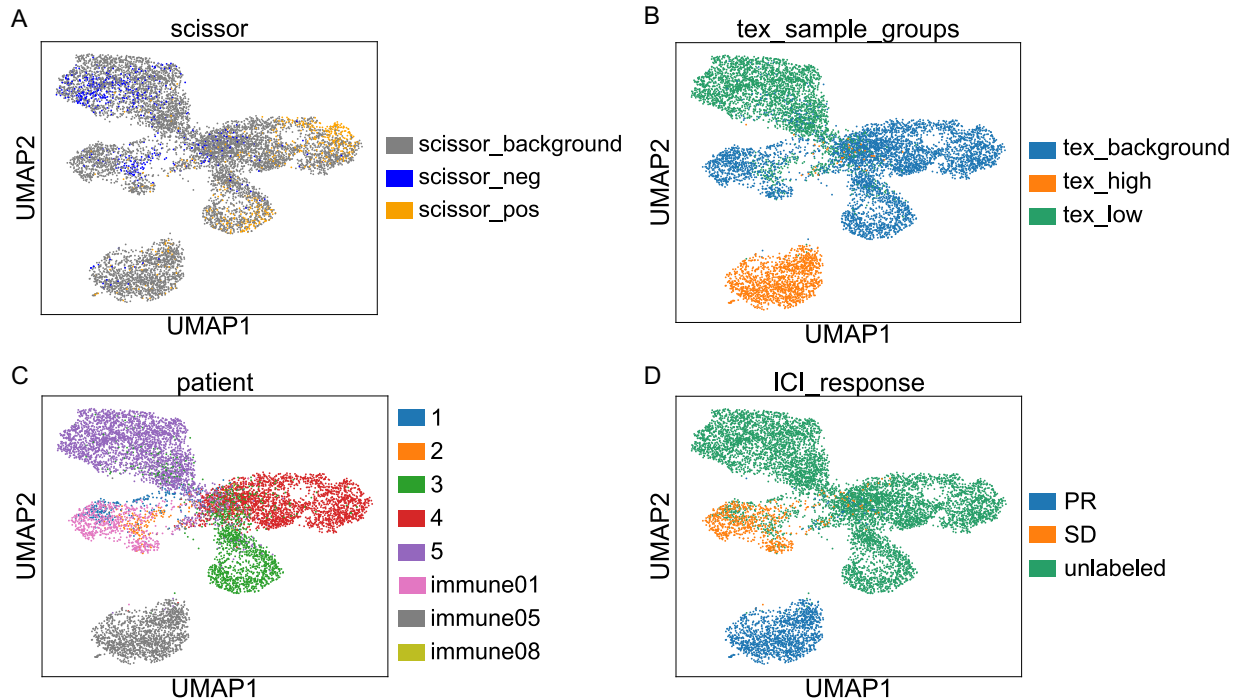
Significant overlap in genes up-regulated in the CD8 T cells of exhaustion-high and exhaustion-low sample groups was observed between the lung data and two independent lung tumor datasets, as indicated by p-values that were all equal to or smaller than  $4.24 \times 10^{-11}$ . This indicates a successful identification of samples with strongly contrasting distributions of these two cell states, which can be observed in a range of tumor tissue contexts.

**3.2.4.5 Selection of exhaustion-high and exhaustion-low tumor cell subpopulations** Once the sample-level exhaustion scores have been calculated, they can be used to identify subpopulations of other cell types associated with various degrees of CD8 T cell exhaustion. In this way, the tumor cells from samples 2 and 5 were determined to come from “exhaustion-low” tumor immune microenvironments, while the tumor cells from sample immune05 came from an “exhaustion-high” microenvironment (Figure 16b, c). Comparisons can then be performed between the tumor cell populations from these exhaustion-low and exhaustion-high immune microenvironments.

### **3.2.5 Identification of phenotype-associated lung tumor subpopulations with *Scissor***

**3.2.5.1 *Scissor* overview** The *Scissor* algorithm is a powerful computational tool that can be used to leverage the clinical labels and higher sample size of bulk RNA-seq data to identify subpopulations of scRNA-seq data associated with a phenotype of interest [144]. It takes as input single-cell RNA seq data and bulk RNA-seq data from the same tissue type, where the bulk data samples are labeled with a clinical phenotype of interest. The correlation between single-cell and bulk gene expression profiles is determined. The *Scissor* algorithm then selects phenotype-associated cells based on their predictive ability in correctly assigning clinical labels to bulk RNA-seq samples with the labels held out. These cells have expression profiles with a high degree of correlation with the expression profiles of bulk RNA-seq samples from one phenotypic category relative to the other. In this way, cells with the strongest association with a particular clinical outcome of interest are identified. When used in the binary, logistic regression mode, it chooses subsets of single cells that are associated with one of two clinical outcomes, and leaves the rest as background. By comparing these extremes, the analysis can become better powered to detect signal that differentiates these cell subgroups by ignoring intermediate expression profiles that are not as strongly correlated with one phenotype or the other. DEGs can be identified with greater sensitivity and phenotypically associated cells can be considered in other ways, as described with the sample score approach described below.

**3.2.5.2 Inputs** scRNA-seq data from the combined lung tumor dataset, described above, were used. This dataset contains 9375 tumor cells from eight different patients. Bulk RNA-seq data from NSCLC patient samples collected prior to ICI therapy, and labeled with ICI response (in the RECIST format, quantifying Best Overall Response – see Methods), were also used. There are 152 NSCLC samples in the *Ravi et al., 2023* bulk RNA-seq dataset [145]. 136 of these samples passed quality filtering, and of those, 123 were labeled with one of the 4 major RECIST categories (Complete Response, Partial Response, Stable Disease, Progressive Disease). 11 were not labeled and 2 were labeled “Not Evaluated” (NE). These, along with the 35 samples classified as “Stable Disease,” were removed (see Methods).



**Figure 16: lung tumor cell subpopulations** A) Scissor-positive (ICI response- associated) and Scissor-negative (ICI resistance-associated) tumor cell subpopulations, superimposed on reference cells B) lung tumor cells with the lowest (microenvironment biased toward progenitor exhausted CD8 T cells) and highest (biased toward terminal exhaustion) exhaustion scores in their tumor immune microenvironments C) lung tumor cells from eight patients D) ICI response labels (RECIST) where available

This left samples from 88 patients, with 8 classified as “Complete Response” (CR), 40 as “Partial Response” (PR), and 40 as “Progressive Disease” (PD). These were then binarized into 48 Responders (CR, PR) and 40 Non-Responders (PD). These binary phenotype labels were then used as input into the Scissor algorithm for selection of phenotype-associated lung tumor cells.

**3.2.5.3 Selection of lung tumor cells related to immunotherapy response and resistance** Using these inputs, the Scissor algorithm was run to identify subpopulations of lung tumor cells associated with ICI response and resistance. Steps were taken to optimize the algorithm’s alpha parameter, and the reliability test was successfully run, indicating sufficient similarity between the expression profiles of the bulk and single-cell data for effective selection of phenotype-associated cells (see Methods). 612 Scissor-positive (ICI response-associated) and 615 Scissor-negative (ICI resistance-associated) lung tumor cells were identified (Figure 16a). This represents 15.201% of all tumor cells in the dataset, which meets the Scissor package’s guideline for choosing less than 20% of all cells as phenotype-associated. In order to provide further support for the phenotype associations of these tumor cells, three different strategies were used, described below.

**3.2.5.4 Additional support for phenotype-associated tumor cells – part 1** After ensuring that the single-cell and bulk lung tumor data were sufficiently similar for the use of the Scissor algorithm, we supported this selection of tumor cell subsets using three strategies. The first and primary strategy to support the Scissor results was performed by comparing the differentially expressed genes (DEGs) derived by comparing the Scissor-positive (ICI response-associated) and Scissor-negative (ICI resistance-associated) lung tumor cells with DEGs from an independent bulk RNA-seq dataset comprising human NSCLC tumors sampled prior to immunotherapy treatment (see Methods). The degree of overlap between genes up-regulated in Scissor-positive (compared to Scissor-negative) cells and genes up-regulated in bulk samples from responders (relative to non-responders) was determined. The likelihood of the observed degree of overlap between Scissor DEGs from the NSCLC single-cell data and phenotype-associated DEGs from the NSCLC bulk data occurring by chance was quantified via Fisher's exact test. This yielded a test statistic of 4.34 and a p-value of  $4.26 \times 10^{-18}$ . This indicates strong overlap between the two sets of DEGs and supports the validity of the association between the chosen cells and immunotherapy treatment outcome.

**3.2.5.5 Additional support for phenotype-associated tumor cells – part 2** In order to provide further support for the association of subsets of tumor cells with ICI response, we took advantage of the fact that three out of the eight tumor samples were labeled with immunotherapy outcome. Samples immune01 and immune08 are marked as SD and sample immune05 is marked as PR. Since it is a partial responder to ICI treatment, we expect this sample's tumor cells to be enriched for Scissor-positive cells, relative to Scissor-negative cells. We do not need to adjust for a base rate since an approximately equal number of cells was chosen for each category (612 Scissor-positive vs. 615 Scissor-negative cells). This PR sample contained 3.7 times as many Scissor-positive cells as Scissor-negative cells, relative to an expected baseline ratio of around 1 (Figures 16d, 17). This provides additional support for the validity of the phenotype-association designations of the lung tumor cells.

**3.2.5.6 Additional support for phenotype-associated tumor cells – part 3** In order to further support the association of selected subsets of tumor cells with ICI response, the three pre-treatment samples from the *Hu et al., 2023* NSCLC scRNA-seq dataset were analyzed together with 12 post-treatment samples from the same dataset (Figures 38-41). Working from the assumption that tumor subpopulations that are more responsive to the therapy in question should be enriched for pre-treatment cells and tumor subpopulations more resistant to ICI treatment should be enriched for post-treatment cells, the following comparisons were made. Two groups of Leiden clusters were observed, where group A is made up of 8% pre-treatment cells (and 92% post-treatment cells), and group B comprises 43% pre-treatment cells (Figure 42). Given a baseline level of 20% pre-treatment cells overall, pre-treatment cells are 41% as common in group A and 2.16 times as common in group B as expected by chance.

Given this, we would expect the group with a higher percentage of pre-treatment cells (group B) to be associated with treatment response and contain a higher proportion of scissor positive cells and a lower proportion of scissor negative cells. We would also expect the group with a lower percentage of pre-treatment cells (group A) to be associated with

treatment resistance and contain a lower proportion of scissor positive cells and a higher proportion of scissor negative cells. These expected differences in proportion were, in fact, observed (Figure 43, Methods).

Group A is made up of 2.5% scissor positive cells, as compared to 3.9% scissor positive cells in group B. Group A also contains 1.4% scissor negative cells, relative to 1.0% scissor negative cells in group B. While these are not large differences, they are directionally correct, providing further support for the phenotypic associations of these cell subsets.

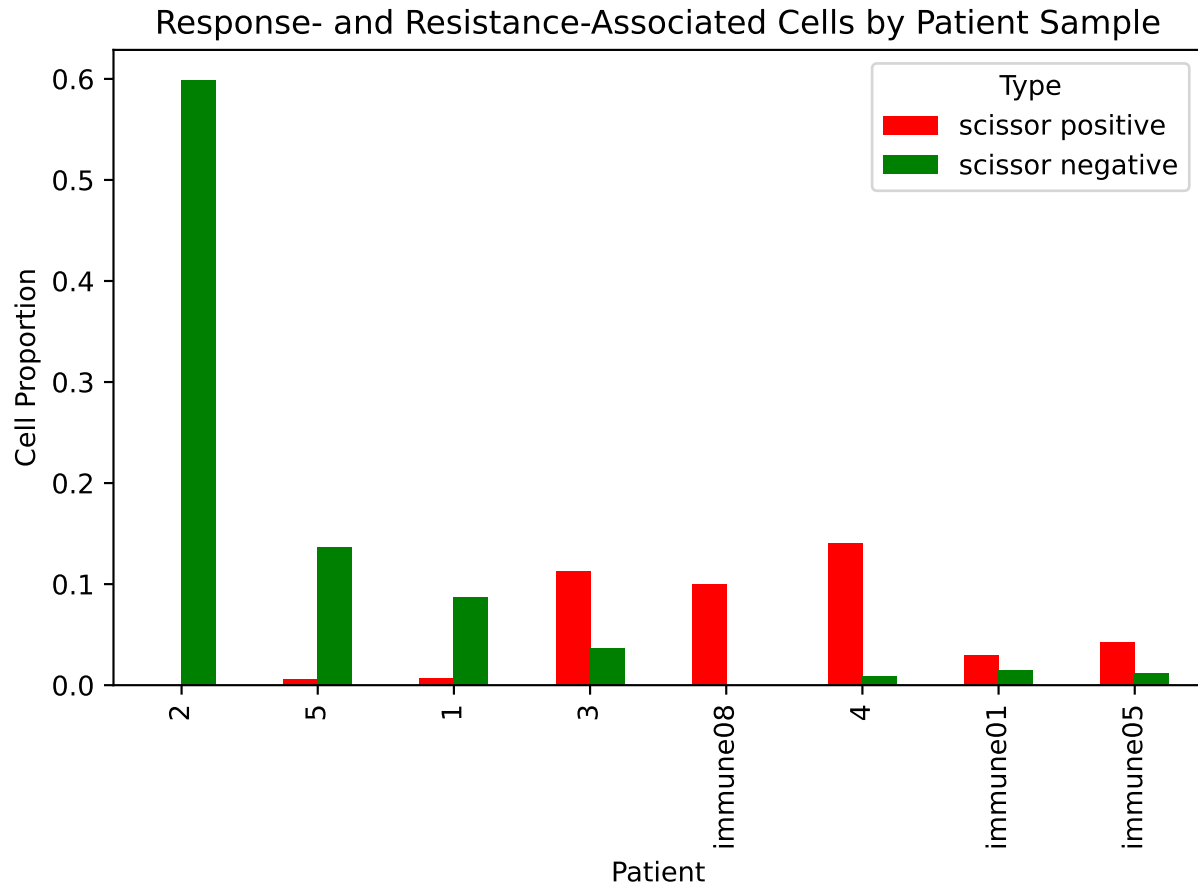
### **3.2.6 Comparison of immunotherapy outcome-associated and T cell exhaustion-associated lung tumor subpopulations**

Using the framework described above, lung tumor cell populations associated with ICI response and resistance have been selected (Figure 16a). Additionally, samples in the single-cell data determined to be from exhaustion-low and exhaustion-high tumor immune microenvironments have been identified (Figure 16b, c). These cell populations can now be compared to each other to further explore the relationship between CD8 T cell exhaustion and patient response to ICI therapy.

**3.2.6.1 Sample-level comparison** These processes can be compared by looking at the relative proportions of response- and resistance-associated cells in each patient sample and determining whether this varies by sample-level exhaustion score. An approximately equal number of each category of phenotype-associated cells was selected, so in absence of a biological effect we expect the proportions to be equal. We observe that patients with the lowest sample-level exhaustion scores have a much greater proportion of resistance-associated cells than response-associated cells (Figures 16b-c, 17). The opposite is observed in samples with intermediate and higher exhaustion scores: an overrepresentation of response-associated cells relative to resistance-associated cells. This means that patients with a greater proportion of progenitor exhausted relative to terminally exhausted CD8 T cell populations contain more cells that match the bulk data of non-responding patients, and patients with a greater proportion of CD8 T cells in the terminally exhausted state contain a greater number of cells that match the responders.

Bearing in mind that in general, the selected phenotype-associated cell subsets make up a minority of the cells within a sample, and tumor cell populations are often heterogeneous, this is somewhat unexpected. Based on previous findings, progenitor exhausted CD8 T cells seem to be associated with more positive outcomes to ICI treatment. In order to quantify this relationship, Spearman rank-based correlation was performed to compare the ranking of exhaustion sample scores with the ratio of the proportion of ICI-response- vs. ICI-resistance-associated cells within each sample. A correlation value ( $\rho$ ) of 0.71 ( $p$ -value=0.047) was calculated. In these data, samples with a higher proportion of terminally exhausted CD8 T cells (relative to progenitor exhausted CD8s) tend to have cells with expression phenotypes that match samples from patients who respond positively to ICI treatment.

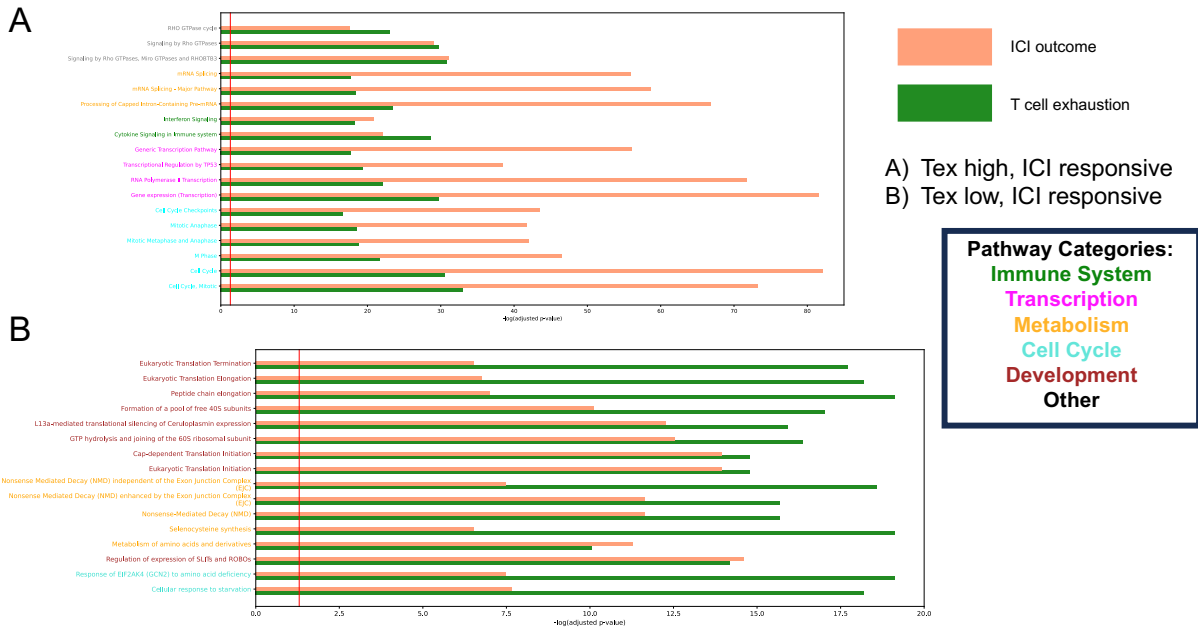
**3.2.6.2 DEG comparison** These two processes of T cell exhaustion and immunotherapy response can also be compared by considering differential gene expression in each. DEG analyses were performed between lung tumor cells



**Figure 17: proportion of cells in each sample that are scissor-positive (ICI response-associated) and scissor-negative (ICI resistance-associated), respectively** Note: samples are ranked from lowest to highest exhaustion score (left to right)

from exhaustion-high and exhaustion-low samples and between ICI-response- and ICI-resistance-associated lung tumor cell populations (from Scissor). These DEGs were then examined for overlap (Table 8, Figures 44-46). The most overlap was seen between exhaustion-high and ICI response-associated DEGs and between exhaustion-low and ICI-resistance-associated DEGs. This trends in the opposite direction from what one may expect from previous work that links a more progenitor exhausted tumor immune microenvironment with better clinical outcomes.

**3.2.6.3 Pathway analysis** After identifying differential gene expression activity shared between T cell exhaustion and ICI responsiveness, these overlapping sets of DEGs were subjected to a pathway enrichment analysis, using Reactome. First, DEGs up-regulated in tumor cells from exhaustion-high samples and also in the ICI-responsive tumor subpopulations were considered (Figure 18a, Figure 47). These genes showed significant shared enrichment in a number of pathways, including transcription and cell-cycle associated pathways. Next, DEGs up-regulated in tumor cells from exhaustion-low samples and in ICI responsive tumor subpopulations were analyzed (Figure 18b). These genes showed significant enrichment in pathways related to development and metabolism.



**Figure 18: Pathway analysis of DEGs in lung tumor cells** Pathways enriched in genes up-regulated in exhaustion and ICI response groups; red line = adjusted p-value cutoff of 0.05 A) pathways up-regulated in exhaustion-high samples and ICI-responsive subpopulations B) pathways up-regulated in exhaustion-low samples and ICI-responsive subpopulations

### 3.2.7 Application of the framework to melanoma data

Initially, use of this framework to analyze NSCLC data and melanoma data was planned. The successful application of the framework to NSCLC is described in this chapter. When searching for datasets, a number of publicly available single-cell RNA-seq datasets from human melanoma tumors were identified. However, none of these datasets included a sufficient number of CD8 T cells and tumor cells from a sufficient number of patients in order to proceed. At this point, the pairing of ICI response phenotype-labeled bulk RNA-seq data from human melanoma tumors with single-cell RNA-seq data from BCC was attempted. Although these are two different cancer types, previous observation of biological similarities in their immune microenvironments indicated that there may be enough similarity in expression

		DEGs – exhaustion score		
		not significant	exhaustion-high	exhaustion-low
DEGs – Scissor cells	not significant	2852	1919	1731
	ICI-responsive	1602	<b>4508</b>	1221
	ICI-resistant	128	190	<b>789</b>

**Table 8: Overlapping DEGs in lung tumor cells – T cell exhaustion vs. immunotherapy outcome**

Genes are categorized based on their expression levels in subsets of tumor cells. Genes are categorized as “exhaustion-high” if their expression in the subset of tumor cells from samples with high CD8 T cell exhaustion scores is significantly up-regulated compared to the subset from low-exhaustion samples. Similarly, genes are categorized as “ICI-responsive” if their expression in the subset of tumor cells associated with ICI response is significantly up-regulated compared to the subset associated with ICI resistance. The inverses are true for “exhaustion-low” and “ICI-resistant.”

profiles for a comparison to be attempted. Scissor’s reliability test was used to evaluate this. The reliability test failed, indicating too low of a concordance between the expression profiles of the bulk and single-cell data to successfully identify phenotype-associated cells.

**3.2.7.1 Application of ICI outcome-related gene signatures to bulk melanoma data** To consider potential biological similarities between NSCLC and melanoma in the context of immunotherapy response, we compared a gene signature derived from DEGs between ICI response and resistance-associated NSCLC tumor cell populations to the ICI response-labeled bulk RNA-seq data from melanoma tumors. A strongly significant overlap was observed here, as quantified by the up-regulation of around six thousand shared genes across the two tissue contexts and a p-value of  $2.83 \times 10^{-187}$  from the overlap analysis (Table 9). This suggests that signatures for ICI response from one tumor type can generalize to other tumor types and that there is likely some shared biology in the process of immunotherapy response between these two types of disease.

### 3.3 Discussion

By developing a framework that combines the capabilities of multiple analysis approaches, we were able to perform comparisons in the tumor immune microenvironment not straightforwardly available with currently existing approaches. The sample score framework (from Chapter 2) enabled the quantification of the degree of CD8 T cell exhaustion in a patient’s tumor immune microenvironment and the subsequent comparison of tumor cells from patients with very different immune microenvironments. By leveraging the Scissor tool and the availability of ICI outcome-labeled bulk expression data, ICI outcome-related tumor subpopulations could be identified. By comparing these subsets of tumor cells, the process of CD8 T cell exhaustion and the clinical course of patients treated with ICI therapy can be more directly compared.

The relationship between the prevalence of more terminally differentiated CD8 T cells and response-associated tumor cell populations was somewhat surprising. There are several explanations for this. First, the biology of T cell exhaustion as it relates to cancer and immunotherapy is not clearly understood. Second, the exhaustion trajectory may be capturing other behavior than purely dysfunction, such as antigen reactivity and normal effector function. Third, this approach relies on clinical phenotype labels that are a snapshot in time and may not capture all of the intricacies of a patient’s disease course.

		<b>DEGs – melanoma data</b>	
		up-regulated	not significant
<b>DEGs – NSCLC data</b>	up-regulated	5999	1332
	not significant	4593	3016

**Table 9: Generalization of ICI outcome DEGs from lung tumor cells to melanoma data –**  
test statistic = 2.96, p-value =  $2.83 \times 10^{-187}$

Another possible explanation arose in another group's work on CD8 T cell exhaustion. A November 2022 paper suggested the possibility that the well-studied transition from progenitor exhausted CD8 T cells to terminally exhausted CD8 T cells may actually be a bifurcating, rather than linear, trajectory [89]. This work identified two endpoints, with one matching the well-known terminally exhausted, TOX-high cell state and another taking the form of a KLR-expressing state with higher anti-tumor effector activity. However, when we applied a gene signature derived from this paper to the NSCLC T cell exhaustion trajectory described above, a bifurcation was not observed. Further investigation is needed, in a tumor-type-specific manner, to better understand this process implicated in the treatment outcomes of so many patients. That said, the generalization of the ICI response expression signature from NSCLC to melanoma lends support to the idea that research findings about ICI treatment in one tumor type may be quite relevant to other tumor types.

This framework facilitated new insights into cancer immunology, but is not restricted to this biological domain. It may be fruitfully applied to other biological processes and disease contexts as well. Any question in which there is a cell state change of interest, interactions with other cell types, and a clearly defined clinical phenotype has the potential to be studied with this computational framework. Such computational investigations have the potential to generate hypotheses and inform bench-top experiments that may confirm new biological relationships related to human health and disease.

### 3.4 Methods

#### 3.4.0.1 Datasets

**3.4.0.2 NSCLC data – single-cell** Three scRNA-seq datasets were used for the lung cancer analysis (Table 1). The *Lambrechts et al., 2018* dataset contains human NSCLC samples from five patients[7]. The *Hu et al., 2023* dataset contains human NSCLC samples from three patients [6]. The *Liu et al., 2022* dataset contains human NSCLC samples from 36 patients[146]. All samples used for the primary analysis are immunotherapy-naïve. Additional post-ICI treatment samples from the *Hu et al., 2023* dataset were used in the section: “Additional support for phenotype-associated tumor cells – part 3.” All datasets were previously published by other groups and are publicly available. Two datasets were downloaded from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) using the following identifiers: GSE207422 (*Hu et al., 2023*), GSE179994 (*Liu et al., 2022*). The third dataset (*Lambrechts et al., 2018*) was downloaded from ArrayExpress using accession numbers E-MTAB-6149 and E-MTAB-6653.

**3.4.0.3 NSCLC data – bulk with ICI response annotation** Additionally, two bulk RNA-seq datasets were used for the lung cancer analysis. The *Ravi et al. 2023* dataset contains human NSCLC samples from 152 patients (88 were used)[145]. The data were obtained from supplemental tables provided with the paper. The *Cho et al., 2020* dataset contains human NSCLC samples from 16 patients. This dataset was downloaded from Gene Expression Omnibus using



the following identifier: GSE126044 [147]. For both of these bulk RNA-seq datasets, samples were collected prior to treatment with ICI therapy and are labeled with immunotherapy response information, scored using RECIST [148].

**3.4.0.4 Melanoma data – bulk with ICI response annotation** A bulk RNA-seq melanoma dataset was used to look for similarity in gene expression between melanoma and NSCLC in terms of immunotherapy response. Human melanoma tumor samples from 18 patients in this dataset were used [149]. The dataset can be downloaded from Gene Expression Omnibus using the following identifier: GSE91061.

### **3.4.1 Pre-processing, merging, cell type annotation, and subsetting of scRNA-seq data**

**3.4.1.1 Merging, batch correction, and pre-processing** Data files were downloaded and converted to the AnnData (v0.10.1) format using in-house Python and Bash scripts and stored as .h5ad files [130]. They were then merged in scanpy (1.9.5)[130]. Doublets were identified and removed using DoubletDetection [132]. Pre-processing was performed using the standard scanpy pipeline for scRNA-seq data, including quality filtering, normalization, and log-transformation steps. Principal Component Analysis was run to reduce dimensionality. In order to account for potential batch effects between samples (and datasets), batch correction was performed using Harmony [150]. A neighbor graph was constructed and clustering was performed using the Leiden algorithm (leidenalg 0.10.1) [134]. The dimensionality was further reduced to two dimensions for visualization using UMAP (umap-learn 0.5.4) [136, 135].

**3.4.1.2 Cell cluster annotation and subsetting** Cell types were manually assigned to clusters based on the following marker genes: CD8 T cells: CD3, CD8, CD4 (absence); NSCLC tumor cells: [151]. For this analysis, only CD8 T cells and tumor cells were of interest. The CD8 T cell clusters and the tumor cell clusters were then subsetted into separate AnnData objects for further analysis.

### **3.4.2 Inference of CD8 T cell exhaustion trajectory with *Monocle3***

The Monocle3 R package was used to infer a pseudotime trajectory from progenitor exhausted to terminally exhausted CD8 T cell states [93]. The top five principal components were used (Figure 48). Root locations were chosen by finding graph principal points with high TCF7 expression and low TOX expression (Figures 49, 50). Pseudotime values were then exported from R to use in sample score calculations. For further detail, see Chapter 2 methods.

### **3.4.3 Calculation of sample-level exhaustion scores**

Sample-level CD8 T cell exhaustion scores, quantifying the distribution of a cell's CD8 T cells along the exhaustion trajectory, were calculated using a previously developed procedure (see Chapter 2). This procedure is analogous to Gene Set Enrichment Analysis (GSEA), which quantifies the over-representation of a pre-defined set of genes in the top-ranked DEGs between two treatment conditions [94]. The procedure here is the same, except cells are substituted for genes, patient samples are substituted for gene sets, and cell ranking comes from the exhaustion pseudotime

trajectory described in the previous step. In this way, samples with a distribution of CD8 T cells biased strongly toward the early end of the exhaustion trajectory will receive a lower score. Samples with CD8 T cells distributed later in the trajectory will receive a higher score. For further detail, see Chapter 2 methods.

#### **3.4.4 Sample clustering and selection of extremes**

Once sample-level exhaustion scores have been calculated, samples are ranked by these scores. The Jenks natural breaks algorithm for clustering in one dimension is used to break the samples into groups. The two extreme groups, representing the “exhaustion-low” and “exhaustion-high” sample groups, are selected for further analysis. For further detail, see Chapter 2 methods.

#### **3.4.5 Support for the exhaustion trajectory – part 1**

In order to evaluate the validity of the exhaustion trajectory, expression of key exhaustion-state-associated genes relative to pseudotime were visualized and their expected trend with increasing pseudotime was confirmed (Figure 51).

#### **3.4.6 Support for the exhaustion trajectory – part 2**

In order to provide further support for the validity of the CD8 T cell exhaustion trajectory inferred with Monocle3, DEGs between the high and low exhaustion sample groups were compared with two datasets. Significance was determined by FDR-corrected p-values and a threshold of 0.05. An overlap analysis between the NSCLC CD8 T cell DEGs and the exhaustion-high and exhaustion-low groups from CD8 T cell trajectories from the melanoma and BCC datasets discussed in Chapter 2 was performed using Fisher’s exact test.

#### **3.4.7 Bulk RNA-seq data – pre-processing and DEG analysis**

Processing and DEG analysis of bulk RNA-seq data were performed with *edgeR* (3.36.0) and *limma* (3.50.0)[152, 153, 154, 155].

#### **3.4.8 Derivation of a binary response phenotype from RECIST scores**

In all of the NSCLC and melanoma datasets used, clinical response to immunotherapy is scored using the RECIST standard [148]. Briefly, the RECIST standard divides patients into one of four categories based on the following criteria. Before and after treatment, all observable tumor lesions are measured, and the sum of the longest diameter of all of these lesions is calculated. If all lesions disappear, the patient is classified as having Complete Response (CR). If the dimensions sum decreases by at least 30% but some lesions remain, the patient falls into the Partial Response (PR) category. If the dimension sum increases by at least 20% or new lesions appear, the patient is said to have Progressive Disease (PD). Patients falling between the criteria for Partial Response and Progressive Disease are said to have Stable Disease (SD).

In order to enrich for the most clear Responder and Non-Responder phenotypes, and thereby select subsets of tumor cells most clearly associated with immunotherapy response or resistance, samples in the Stable Disease category were excluded from the analysis. These phenotypes were then binarized: patients with Complete Response or Partial Response were considered Responders and patients with Progressive Disease were considered Non-Responders. This binary clinical phenotype was then used for selection of phenotype-associated tumor cells and DEG analysis.

### **3.4.9 Selection of ICI response- and resistance-associated tumor cells using *Scissor***

The *Scissor* R package (2.0.0) was used to select subpopulations of lung tumor cells associated with ICI response or resistance [144]. The *Scissor* algorithm uses clinically labeled bulk RNA-seq data to select subsets of cells in scRNA-seq data associated with those clinical labels. When run in binary classification mode, *Scissor* is based on logistic regression. The degree of correlation between the expression profiles of each bulk sample and each single cell are determined. Then, single cells whose expression profiles are most useful in predicting whether an unlabeled bulk sample falls into one phenotypic category or the other are selected as the most phenotype-associated. They are labeled as “*Scissor*-positive” or “*Scissor*-negative”, corresponding to the two phenotypic categories (in this case, ICI response or resistance), and the rest of the cells are left as background. Once identified, the *Scissor*-positive and *Scissor*-negative cell populations can be compared to each other, often giving DEG results that are more sensitive than those derived from bulk RNA-seq. This is particularly useful in our use case of tumor samples, as we can restrict the analysis to just tumor cells and get differential expression results from this one cell type without noise from the expression profiles of other cell types. See the original *Scissor* paper, as well as the corresponding GitHub repository and tutorial, for further details. The Seurat (4.3.0) R package was used for pre-processing [156, 157, 158, 159, 160].

### **3.4.10 Optimization of alpha parameter**

The *Scissor* algorithm contains a parameter, alpha, which imposes a sparsity constraint. Larger values of alpha will lead the algorithm to select a smaller number of phenotype-associated cells. There is a tradeoff here, since we want a clearly defined subset of the tumor cells, the noise from which will start to be diluted if too many tumor cells are chosen. That being said, a sufficiently high cell count is also desirable for downstream analysis. In order to balance these two priorities, *Scissor* was run on the NSCLC data with a number of different alpha values. The resulting number of phenotype-associated cells was determined for each alpha value. Additionally, the mean (and median) pairwise distance across all chosen *Scissor*-positive and *Scissor*-negative cells were calculated. These values were plotted and a value of alpha that best optimized these two metrics was selected. For the lung analysis, an alpha value of 0.275 was chosen (Figures 52, 53).

### **3.4.11 *Scissor* reliability test**

In order to ensure the validity of our selection of ICI response- and resistance-associated cells using *Scissor*, we first ran the built-in reliability test. This reliability test runs permutations of the *Scissor* algorithm with shuffled phenotype

labels in order to determine how much more often than chance the selected phenotype-associated cells correctly predict phenotype labels.

When the reliability test was run on the combined NSCLC tumor single-cell data and the NSCLC bulk RNA-seq training dataset, a p-value of 0.000 was obtained. The Scissor package rounds down to zero after three decimal points. This indicates a successful reliability test and that the bulk and single-cell NSCLC data are sufficiently similar to be used together to identify phenotype-associated single cells.

### **3.4.12 Additional support for phenotype-associated tumor cells – part 1**

In order to provide more support for the association of subsets of tumor cells with ICI response, an additional bulk RNA-seq dataset not used in the selection of Scissor cells was utilized for comparison. The top 500 most significant differentially expressed genes (DEGs) between Scissor-positive and Scissor-negative cells were identified. The top 500 most significant DEGs between Responders (CR, PR) and Non-Responders (PD) in the independent bulk RNA-seq dataset were also identified. A Fisher's exact test was then used to determine whether the top 500 DEGs from the bulk and single-cell data contained more overlapping genes than expected by chance. As above, response categories for the bulk data were defined by RECIST scores, with patients designated as CR or PR labeled as responders and patients marked PD labeled as non-responders. Patients determined to have Stable Disease were omitted.

### **3.4.13 Additional support for phenotype-associated tumor cells – part 2**

In order to provide further support for the selection of ICI outcome-related lung tumor cells, clinical labels for the scRNA-seq samples from the *Hu et al., 2023* NSCLC dataset were used. Relative proportions of response- versus resistance-associated cells were considered in the context of what would be expected for each sample.

### **3.4.14 Additional support for phenotype-associated tumor cells – part 3**

The third piece of support for the phenotype associations of subsets of lung tumor cells was performed by making use of the fact that the *Hu et al., 2023* NSCLC scRNA-seq dataset also contained a number of post-ICI treatment samples. Relative proportions of pre- and post-treatment cells were compared to response- versus resistance-associated cells, working from the assumption that cells from pre-treatment samples should be overrepresented in cell populations associated with treatment response and post-treatment cells should be overrepresented in cell populations associated with treatment resistance (as they are expected to remain following treatment). Note that the Scissor selection of phenotype-associated cells was run only on tumor cells from pre-treatment samples, so post-treatment cells could not be labeled as Scissor-positive or Scissor-negative. Accordingly, the denominator used to calculate the percentage of phenotype-associated cells in a group is the number of pre-treatment cells in that group, rather than the total cell count.

### **3.4.15 Differential gene expression analysis – exhaustion scores vs. ICI outcome**

Once exhaustion-associated and clinical phenotype-associated lung tumor subpopulations were identified, the resulting cell subsets were then compared. First, two sets of differentially expressed genes were calculated. Expression differences between exhaustion-high and exhaustion-low tumor cell populations, based on the distribution of CD8 T cells from the corresponding samples along the exhaustion trajectory, were quantified. Differential expression between ICI response- and resistance-associated lung tumor cells was also calculated. For each, DEG calculations were performed with the 'sc.tl.rank\_genes\_groups()' function in scanpy, with "method=t-test". P-values were then corrected using the Benjamini-Hochberg method, and significance was determined based on an adjusted p-value threshold of 0.05. An overlap analysis was then performed via Fisher's exact test to determine how many genes are associated with one or several of these four categories (exhaustion-low, exhaustion high, ICI resistance, and ICI response).

### **3.4.16 Sample representation analysis – exhaustion scores vs. ICI outcome**

Next, the relationship between sample-level exhaustion scores and the relative proportions of ICI response- and resistance-associated cells within each sample was considered. First, this was examined visually via a barplot. Next, Spearman correlation was used to determine the strength of association between the CD8 T cell exhaustion score and the ratio of the proportion of Scissor-positive (response-associated) to Scissor-negative (resistance-associated) cells in each sample, across samples. This was performed using spearmanr from scipy.stats.

### **3.4.17 Pathway enrichment analysis with Reactome**

Using Reactome (v85) pathway gene sets and pathway hierarchy, we performed a pathway enrichment analysis on DEGs from the exhaustion score comparison and the phenotype association comparison [32, 33, 34, 35]. Each Reactome pathway's enrichment for DEGs was quantified via Fisher's exact test (stats.fisher\_exact()). The resulting p-values were then corrected using the Benjamini Hochberg procedure (statsmodels.stats.multitest, mode='fdr\_bh').

### **3.4.18 Generalization of NSCLC gene signature to other tumor types**

In order to determine whether response- and resistance-associated gene expression activity signatures derived from NSCLC tumor cell populations generalized to other cancer types in their relatedness to immunotherapy response, these gene sets were compared to DEGs calculated for bulk RNA-seq data from human melanoma tumors. In this melanoma data, immunotherapy response was scored with RECIST for each patient; two patients showed Complete Response, four showed Partial Response, and 12 had Progressive Disease. These patients were split into responders and non-responders as described above. The *edgeR* and *limma* R packages were used for processing of bulk RNA-seq data and calculation of DEGs. An overlap analysis was conducted via Fisher's exact test to quantify the similarity between the DEGs from the lung tumor single-cell data and the melanoma bulk data as they each relate to ICI response.

### **3.4.19 Other package versions:**

*matplotlib=3.5.1; matplotlib-base=3.5.1; networkx=2.8.5; numpy=1.22.3; numpy-base=1.22.3; pandas=1.4.2;  
python=3.10.4; scikit-learn=1.1.2; seaborn=0.11*

## 4 Conclusion

By developing novel analysis frameworks, new questions can be asked of publicly available bulk and single-cell transcriptomics data. We have developed a sample-level exhaustion scoring framework and integrated it with an existing algorithm in a new workflow to identify ICI outcome-related tumor subpopulations. This work was motivated by a need to consider gene expression changes within certain cell types as they relate to cell state change in a different cell type, and further, to investigate the links between these expression changes and clinical phenotypes of interest, such as treatment response. This approach yielded new findings that better elucidate the relationship between T cell exhaustion and immunotherapy response.

The development of this framework and its application to data from human tumor samples enabled the identification of immune cell activity in the tumor immune microenvironment associated with T cell exhaustion. Notably, strong associations between macrophage activity and CD8 T cell exhaustion were found. Additionally, some exhaustion-associated immune cell activity was determined to be shared across the tumor and chronic viral infection contexts, while some was either tumor- or viral-specific. Surprising relationships between exhaustion-associated and immunotherapy response-associated gene expression activity in tumor cell subpopulations were also identified.

Going beyond this, this framework adds to the general scRNA-seq analysis toolkit, allowing researchers to pursue answers to a broader range of questions in their analyses. At this point, pseudotemporal ordering is commonly used in the analysis of scRNA-seq data. Any time other cell types of interest are present within a dataset with an inferred cell state change trajectory, this approach can also be applied. In addition, bulk expression data labeled with phenotypes of interest can be used to consider clinical phenotypes in relation to that cell state change. In light of the prevalence of publicly available, phenotype-labeled bulk RNA-seq data, this opens up new lines of analysis for researchers studying a wide range of biological and clinical topics.

This analytical approach has several limitations. First, as with any computational biology research, any results must be experimentally validated before they can be regarded as sound. Computational research can be highly effective in reducing the number of possible relationships of interest by orders of magnitude, prioritizing a subset of molecular interactions when the total number could not feasibly be evaluated experimentally in an exhaustive manner. Extensions of this research direction would include working with experimental biologists to validate the relationships that were identified, starting with signaling from macrophages to CD8 T cells in the tumor immune microenvironment.

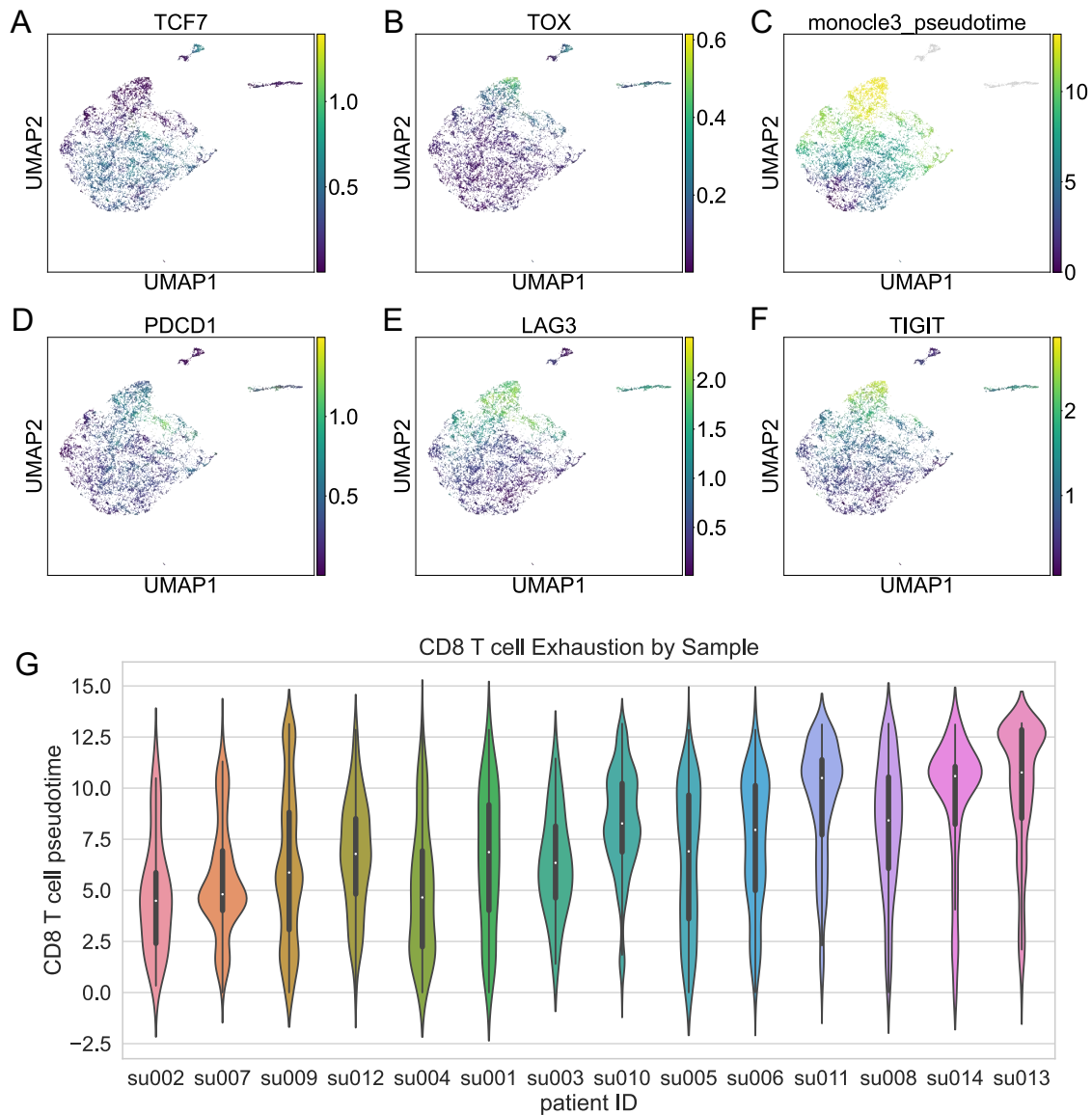
Additionally, this framework is set up to identify statistical correlations, rather than mechanistic relationships. Possible extensions to this work include an approach to track from gene expression of a cytokine, to cytokine activity in the extracellular space from the sending cell type, through to receptor signaling and down to gene regulation in the receiving cell type. This extension, which has already been partially developed, would allow for hypothesis generation regarding potential molecular mechanisms that drive the strong associations that were observed.

The application of this framework is also limited by the datasets used and the characteristics of the patients from whom the tumor samples were derived. Patients may exhibit different disease progression and respond differently to a given treatment based on clinical features such as age, sex, medical history, or haplogroup. In the ideal case, when clinical metadata are available for all samples used, researchers may be able to appropriately caveat the results; they can indicate where results may not generalize to a particular group of patients (and actively seek out more representative datasets in future investigations). However, particularly when publicly available datasets are used, complete clinical metadata are often not available for all samples. In these cases, the degree to which results may generalize to a specific patient cohort remains unclear. The lack of complete clinical annotation of the samples used in this research present this challenge, and more data with more complete clinical labeling will be required to better interpret the generalizability of these results in the context of heterogeneous patient populations. This is a limitation of the datasets used, but is not intrinsic to the framework itself.

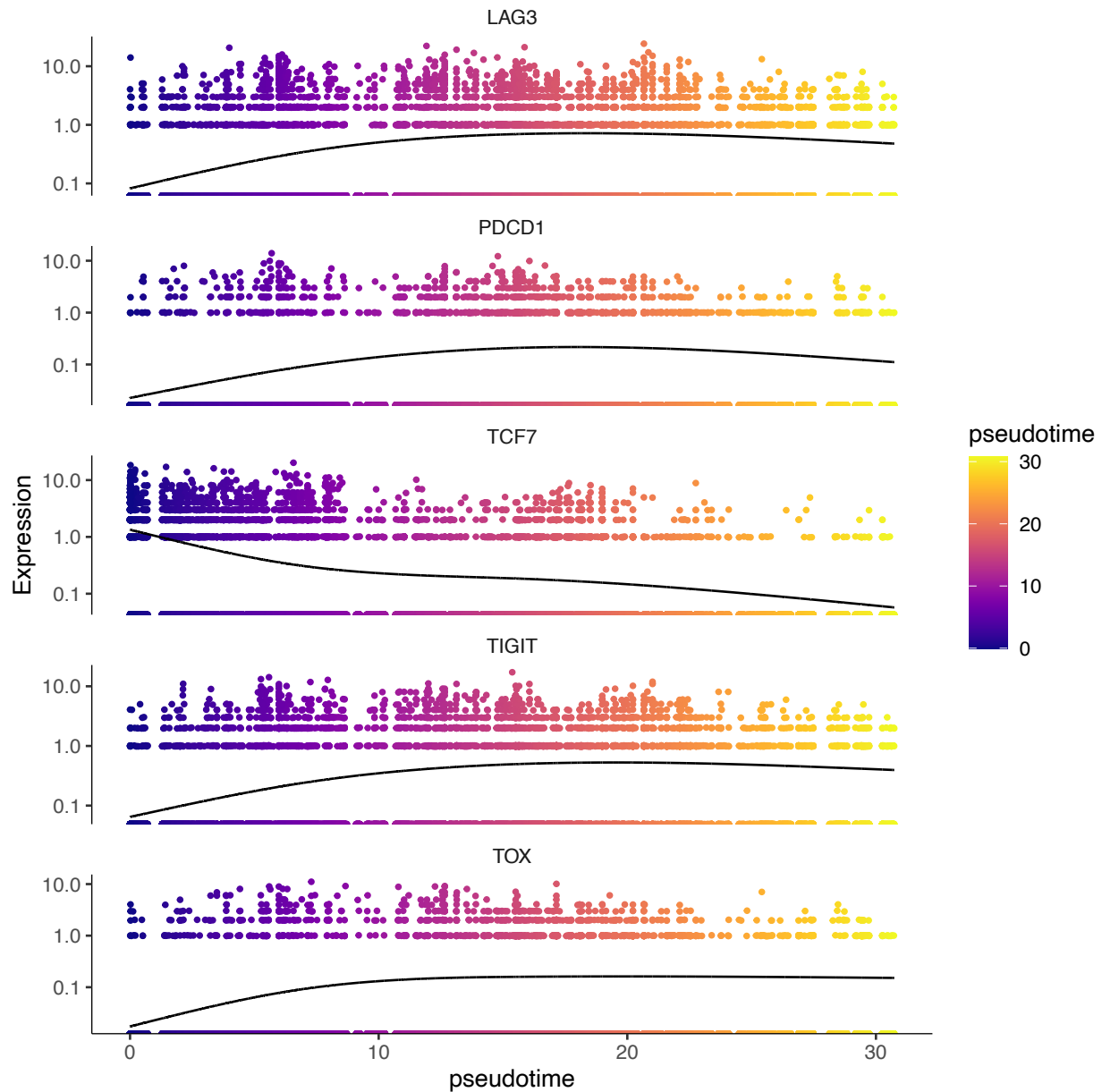
In the coming years, the field of computational biology will continue to evolve rapidly. With the proliferation of multi-omics assays, spatial transcriptomics, and other new data types, new computational techniques will be required to push the limits of gene expression analyses. Greater integration will likely be possible between datasets quantifying gene expression, gene regulation, chromatin accessibility, DNA methylation, proteomics, metabolomics, and other aspects of molecular activity that determine the character of cellular behavior. Computational frameworks like this one will need to be extended, modified, and eventually replaced in order to continue to deepen our understanding of biological systems and to drive translational impact for a wider range of patients.



## 5 Appendix

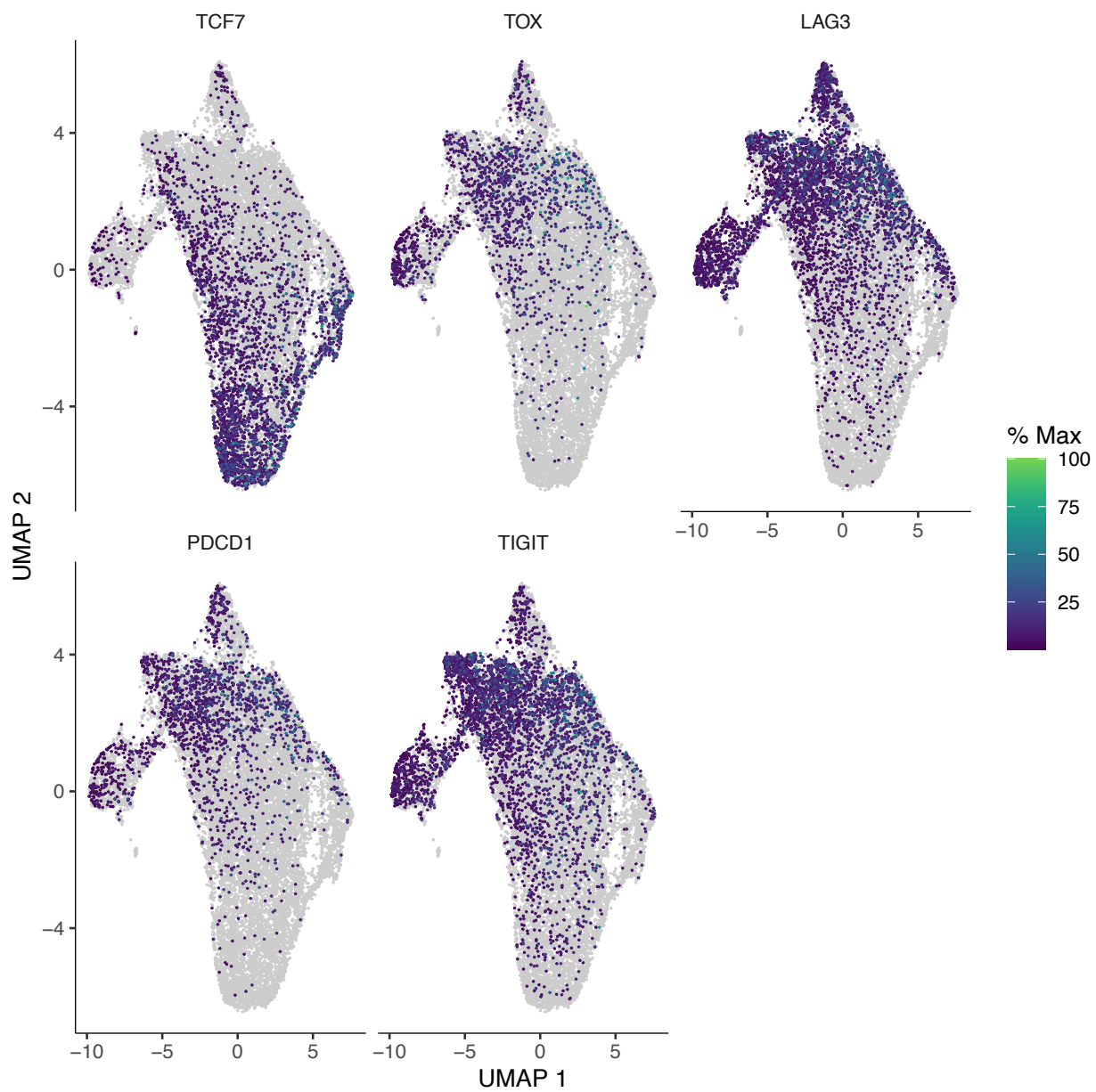


**Figure 19: CD8 T cell exhaustion in human skin tumor (basal cell and squamous cell carcinoma) samples**  
**a, b, d-f)** imputed gene expression of progenitor exhausted marker TCF7, terminal exhaustion marker TOX, and immune checkpoints LAG3, TIGIT, and PDCD1 **c)** Monocle3 pseudotime, characterizing progression from progenitor exhausted to terminally exhausted CD8 T cells **g)** exhaustion pseudotime of CD8 T cells, ordered by sample-level exhaustion score

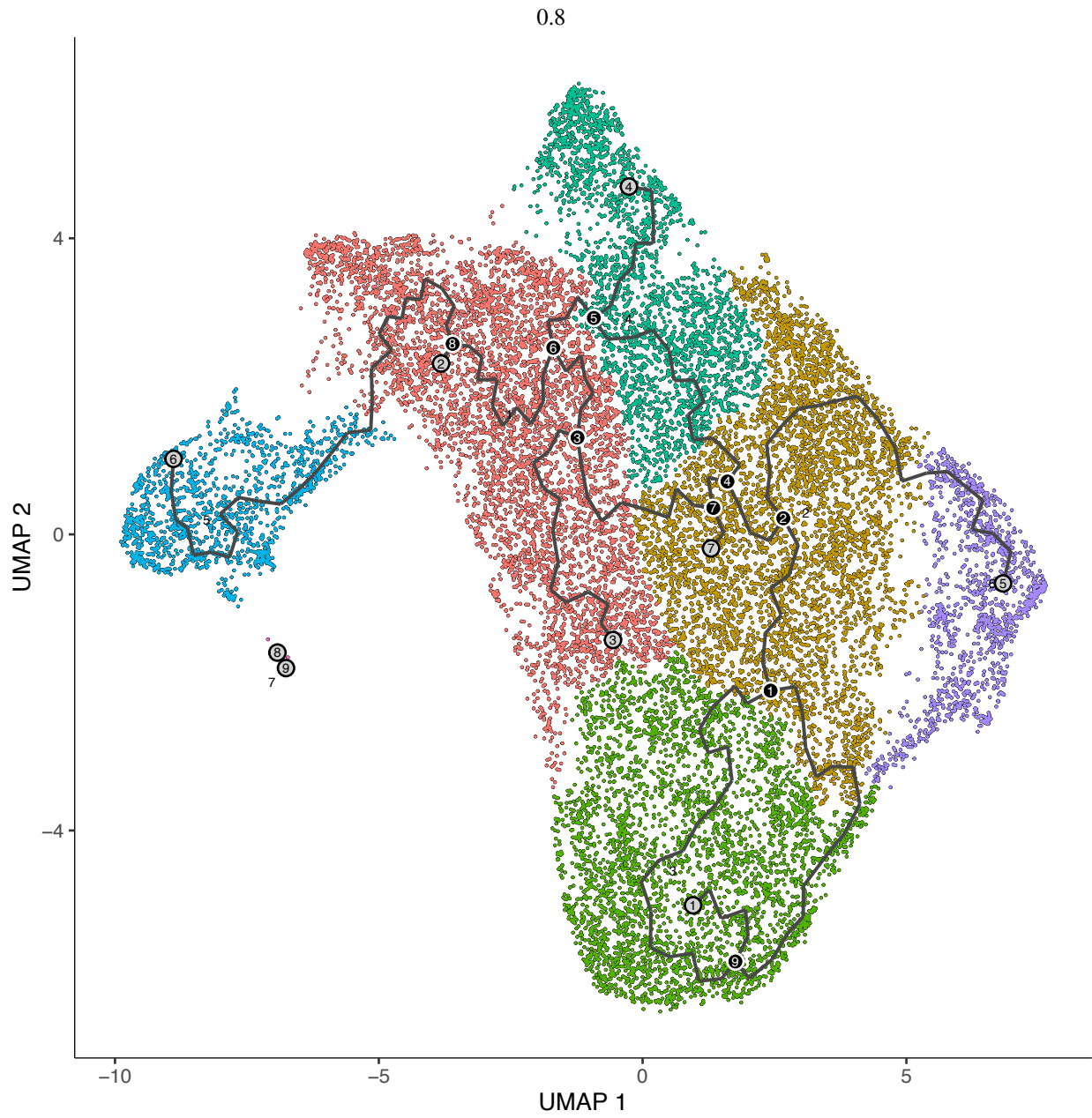


**Figure 20: gene expression vs. pseudotime in melanoma dataset**

Expression of TCF7, the primary marker of the progenitor exhausted CD8 T cell state, starts high and is monotonically decreasing across pseudotime. TOX, the primary marker of the terminally exhausted CD8 T cell state, shows the opposite relationship. Immune checkpoint genes LAG3, PDCD1, and TIGIT show broadly similar expression patterns to TOX relative to pseudotime, although they decrease slightly for the highest pseudotime values.

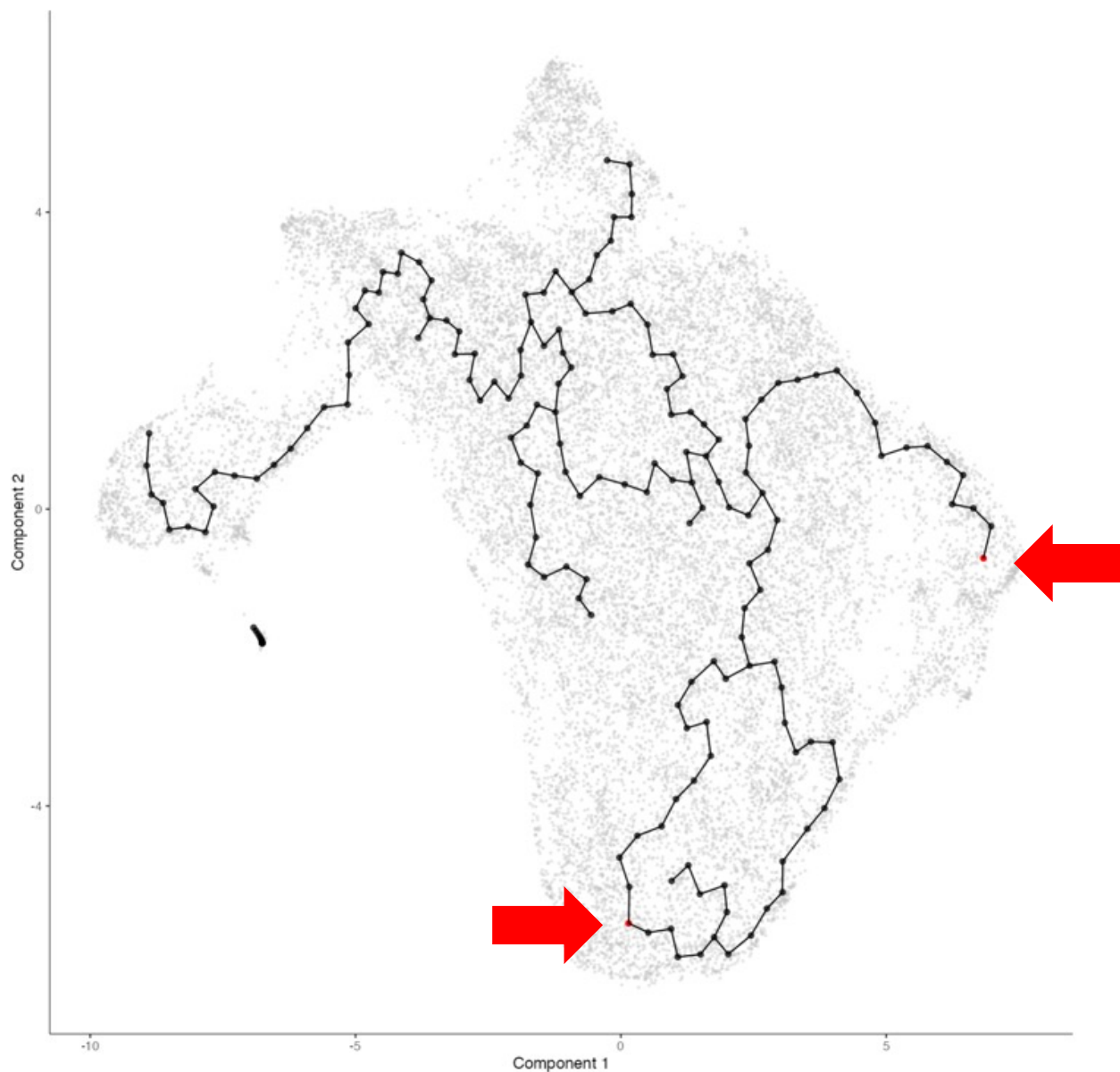


**Figure 21: CD8 T cell UMAP plot – expression of key genes in melanoma dataset**



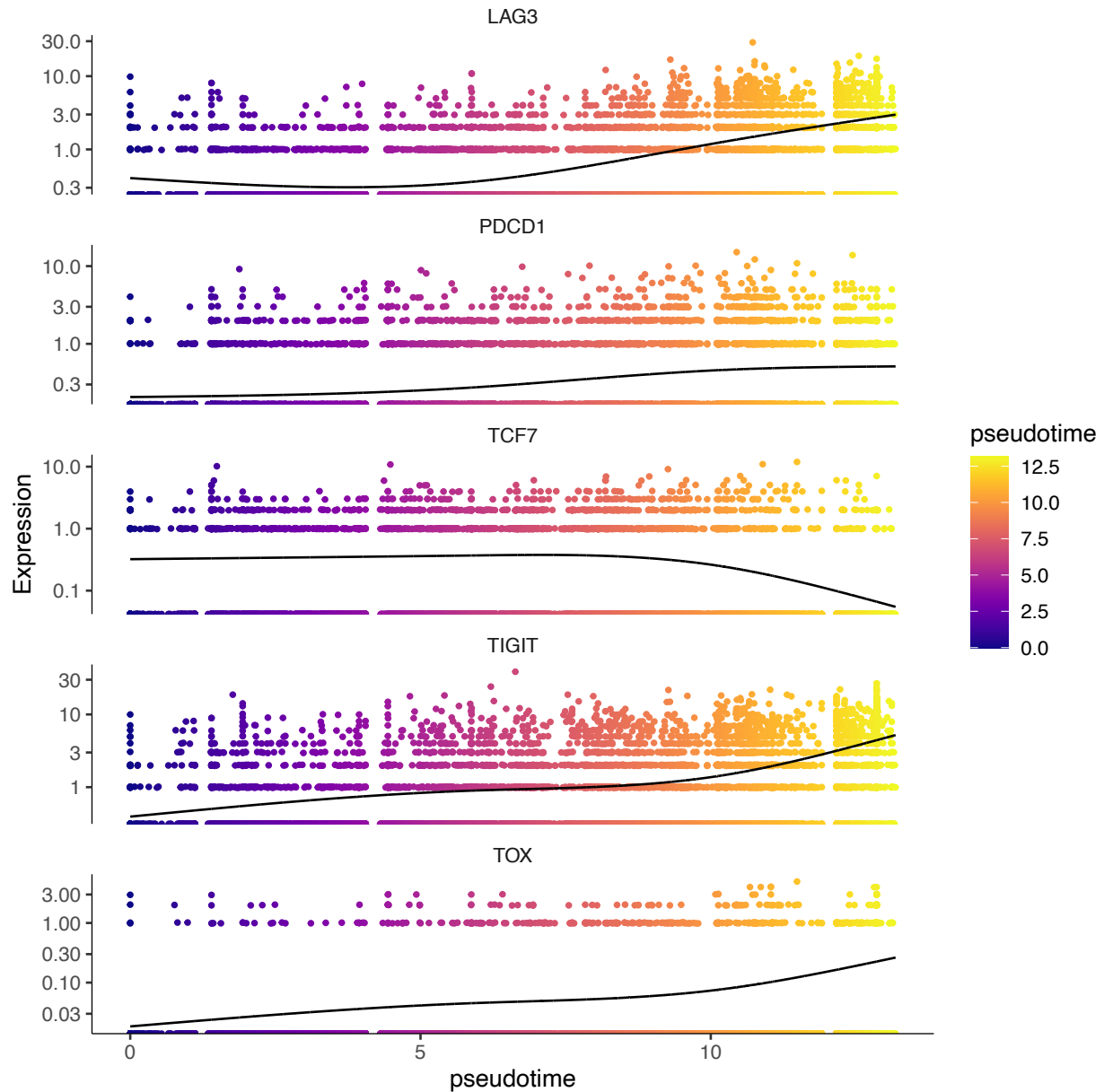
**Figure 22: CD8 T cell UMAP plot – leiden clusters for melanoma dataset**

0.8



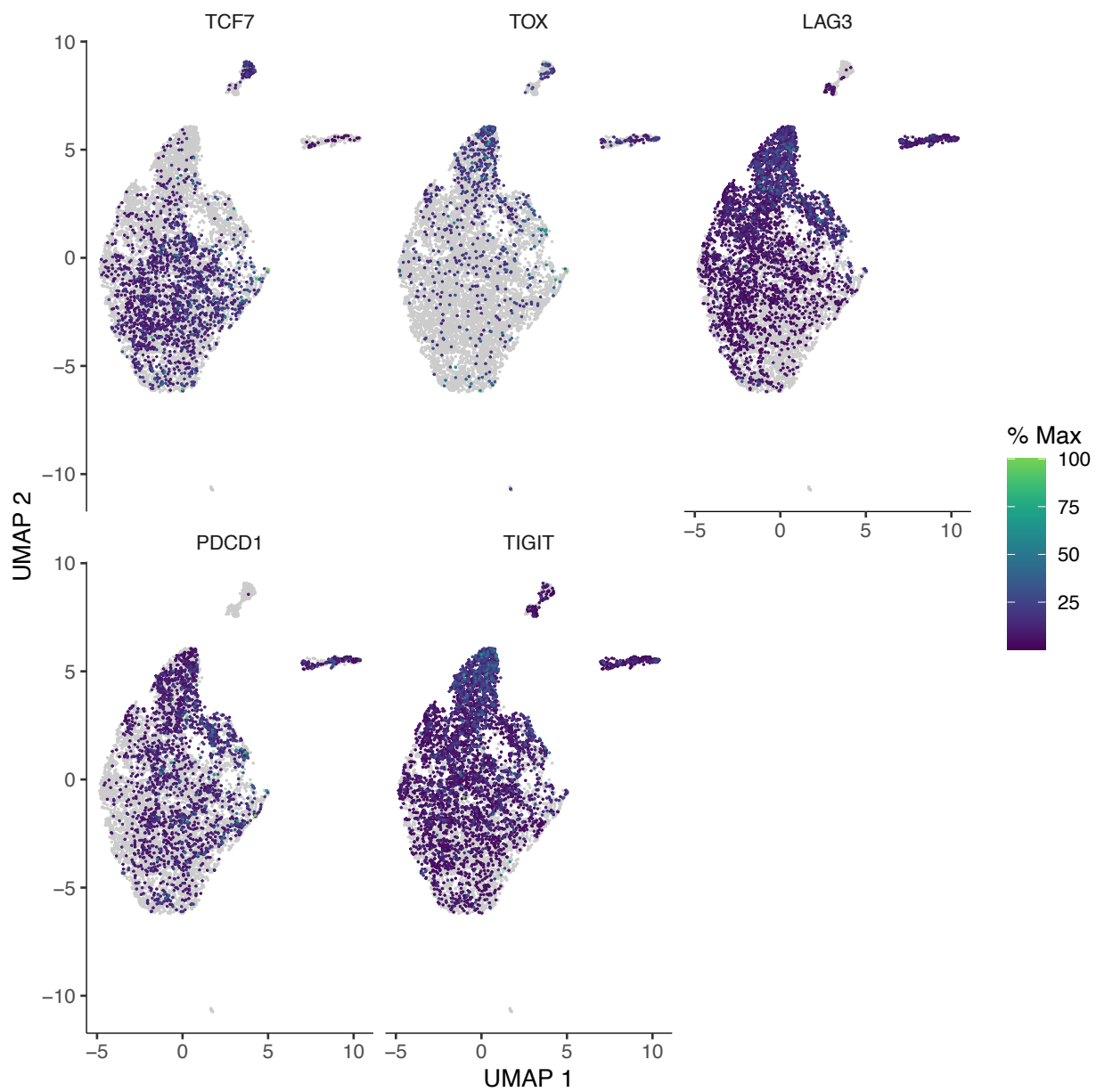
**Figure 22: CD8 T cell UMAP plot – Monocle3 principal graph and clusters for melanoma dataset**  
top: clusters; bottom: principal graph and root cells

**Parameters used:** Number of principal components: 18; Clustering: k=20, resolution=0.0001



**Figure 23: gene expression vs. pseudotime in basal cell carcinoma dataset**

Expression of TCF7, the primary marker of the progenitor exhausted CD8 T cell state, starts high, increases very slightly and then decreases relative to pseudotime. TOX, the primary marker of the terminally exhausted CD8 T cell state, is monotonically increasing across pseudotime. Immune checkpoint genes LAG3, PDCD1, and TIGIT increase across pseudotime (with a slight decrease in LAG3 at the beginning).



**Figure 24: CD8 T cell UMAP plot – expression of key genes – BCC dataset**



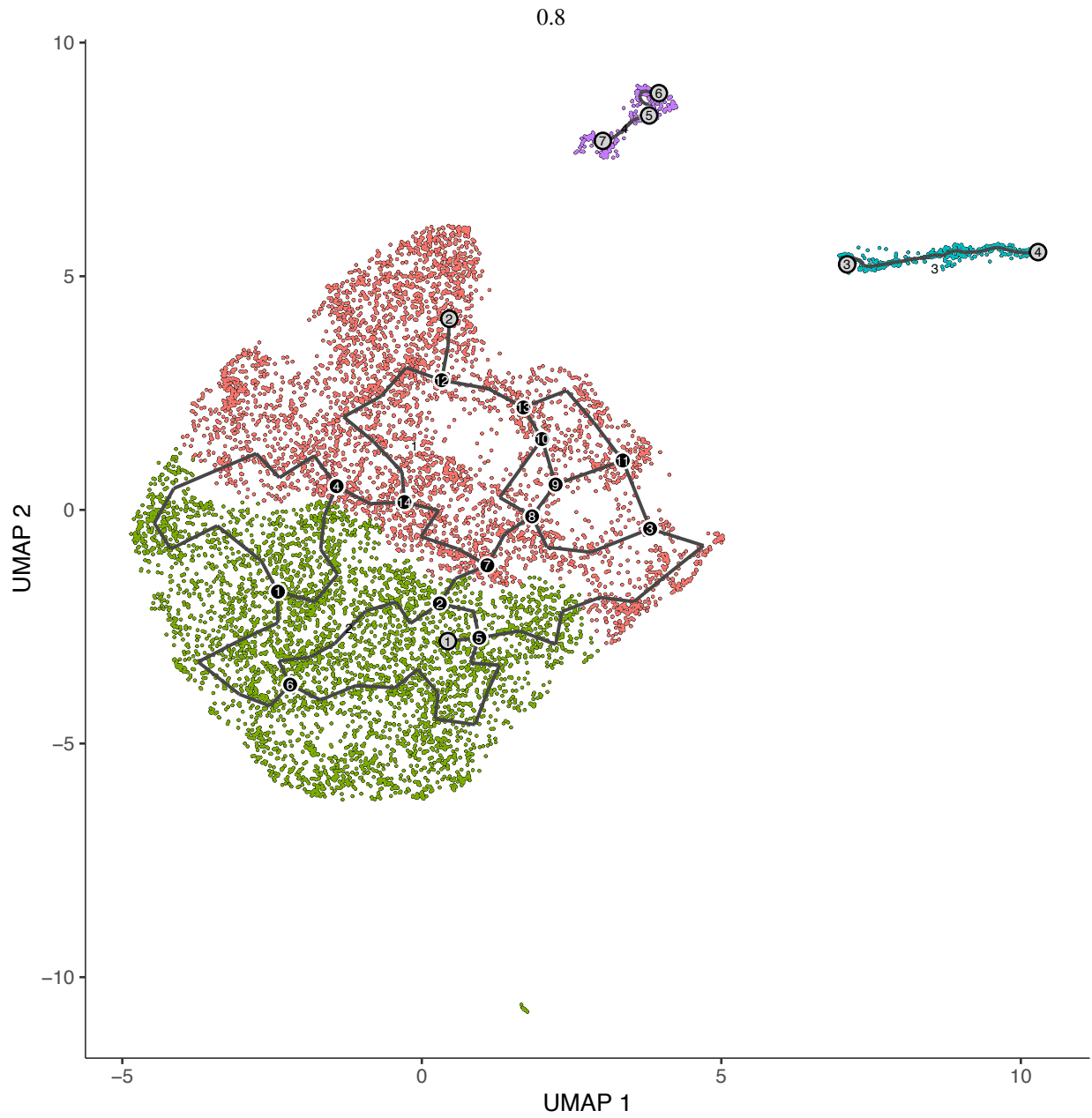
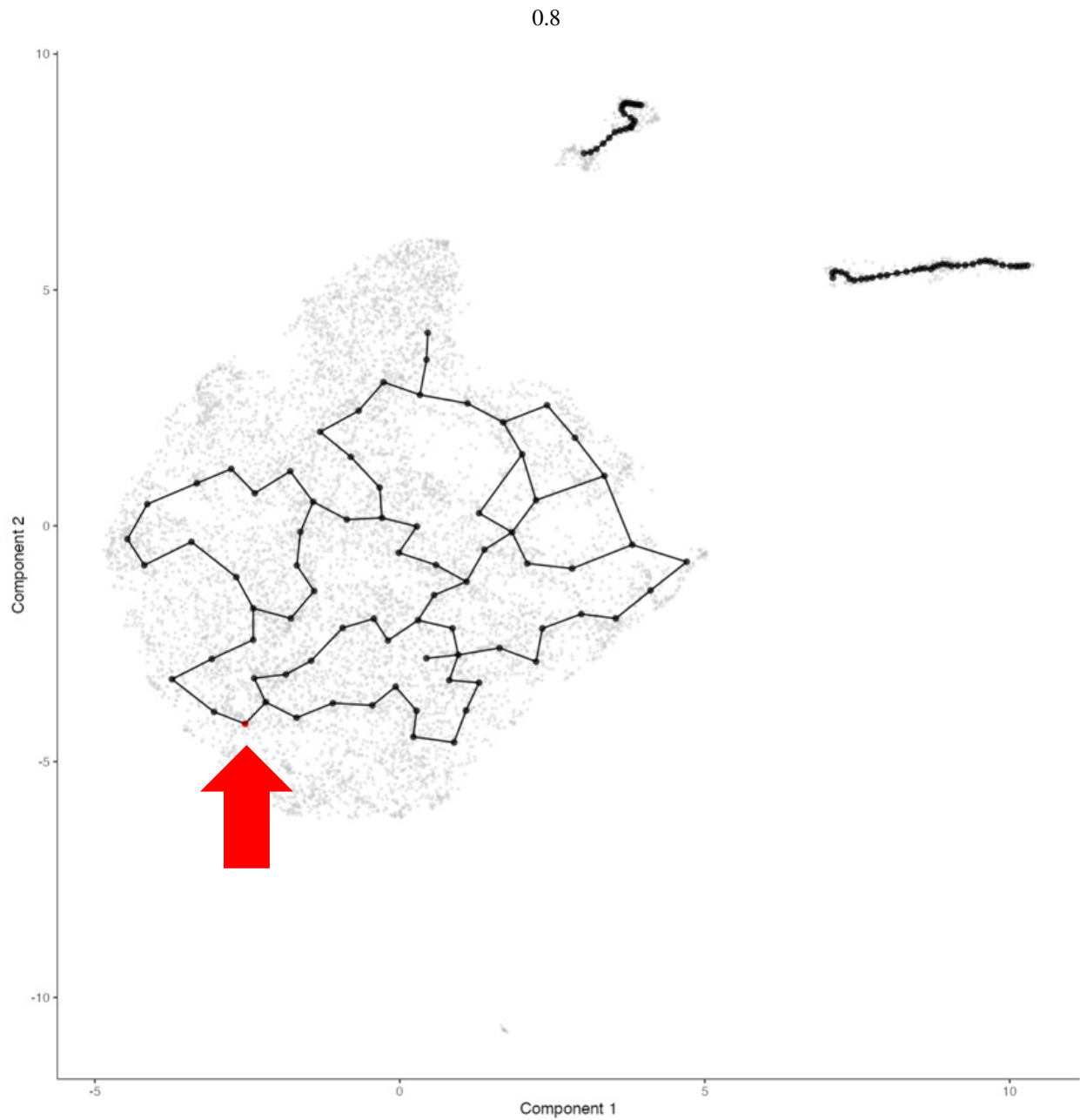
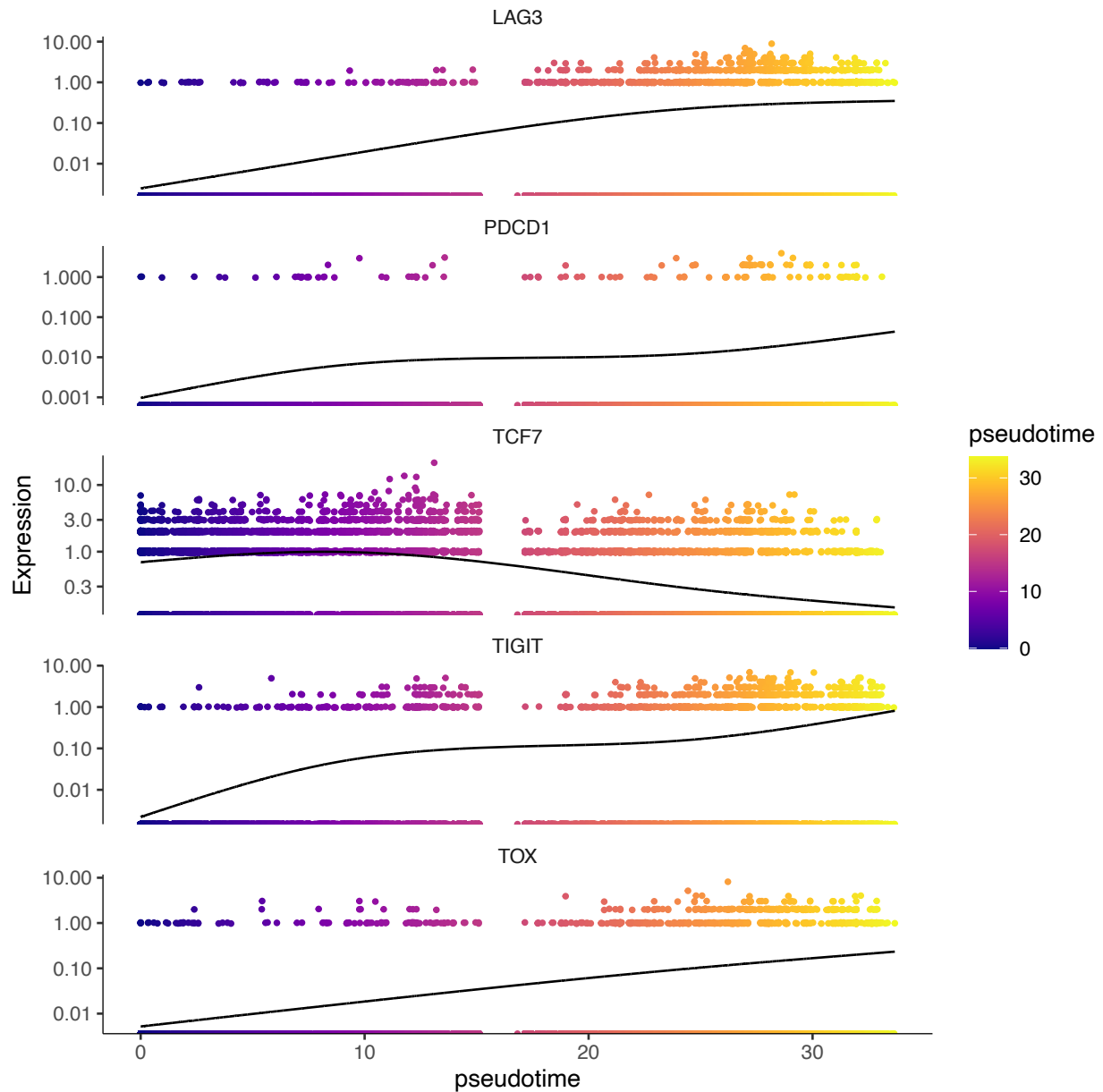


Figure 25: CD8 T cell UMAP plot – Leiden clusters – BCC dataset

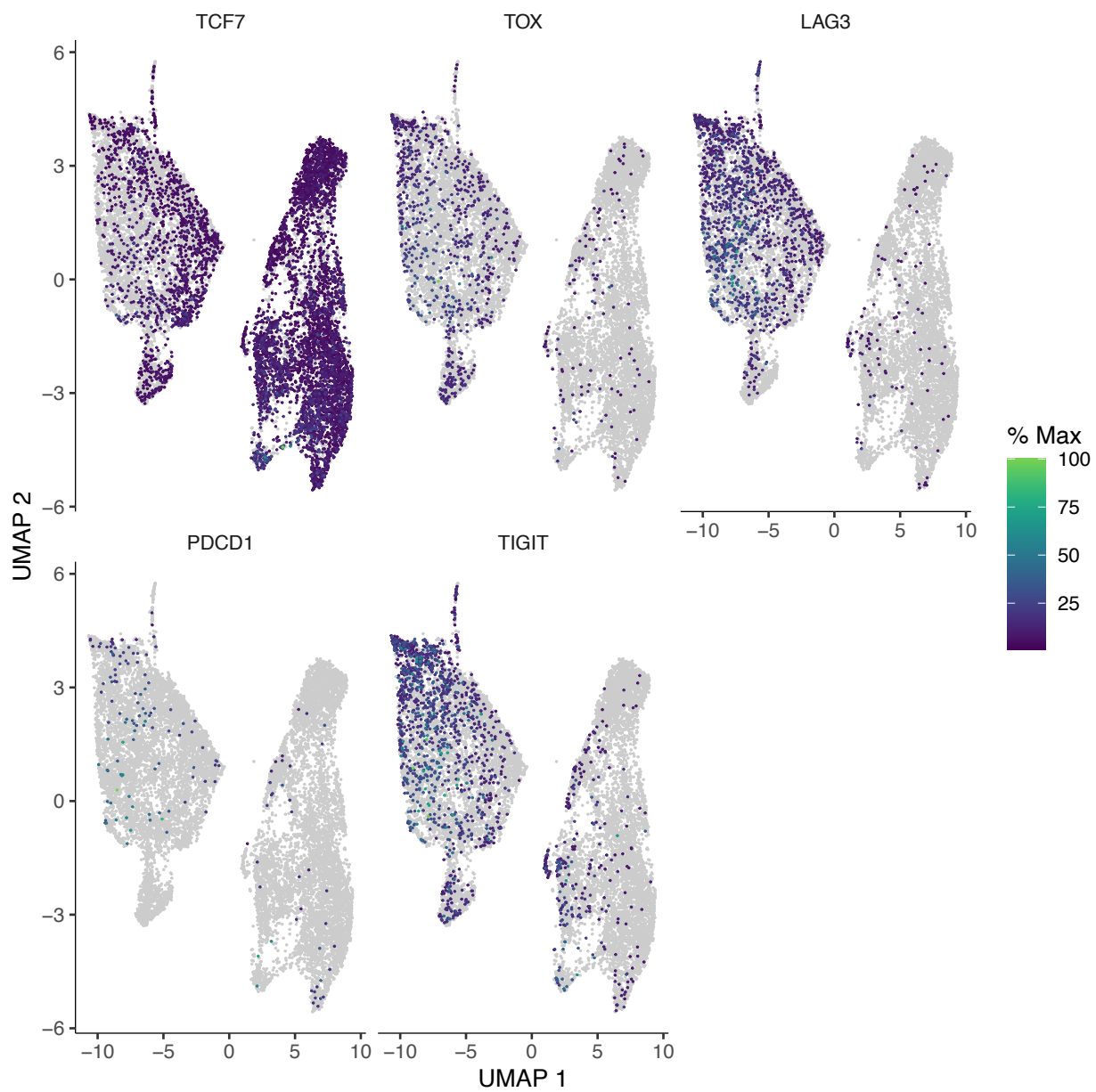


**Figure 25: CD8 T cell UMAP plot – Monocle3 principal graph and clusters – BCC dataset**  
top: clusters; bottom: principal graph and root cells  
**Parameters used:** Number of principal components: 15; Clustering: k=20, resolution=0.0001

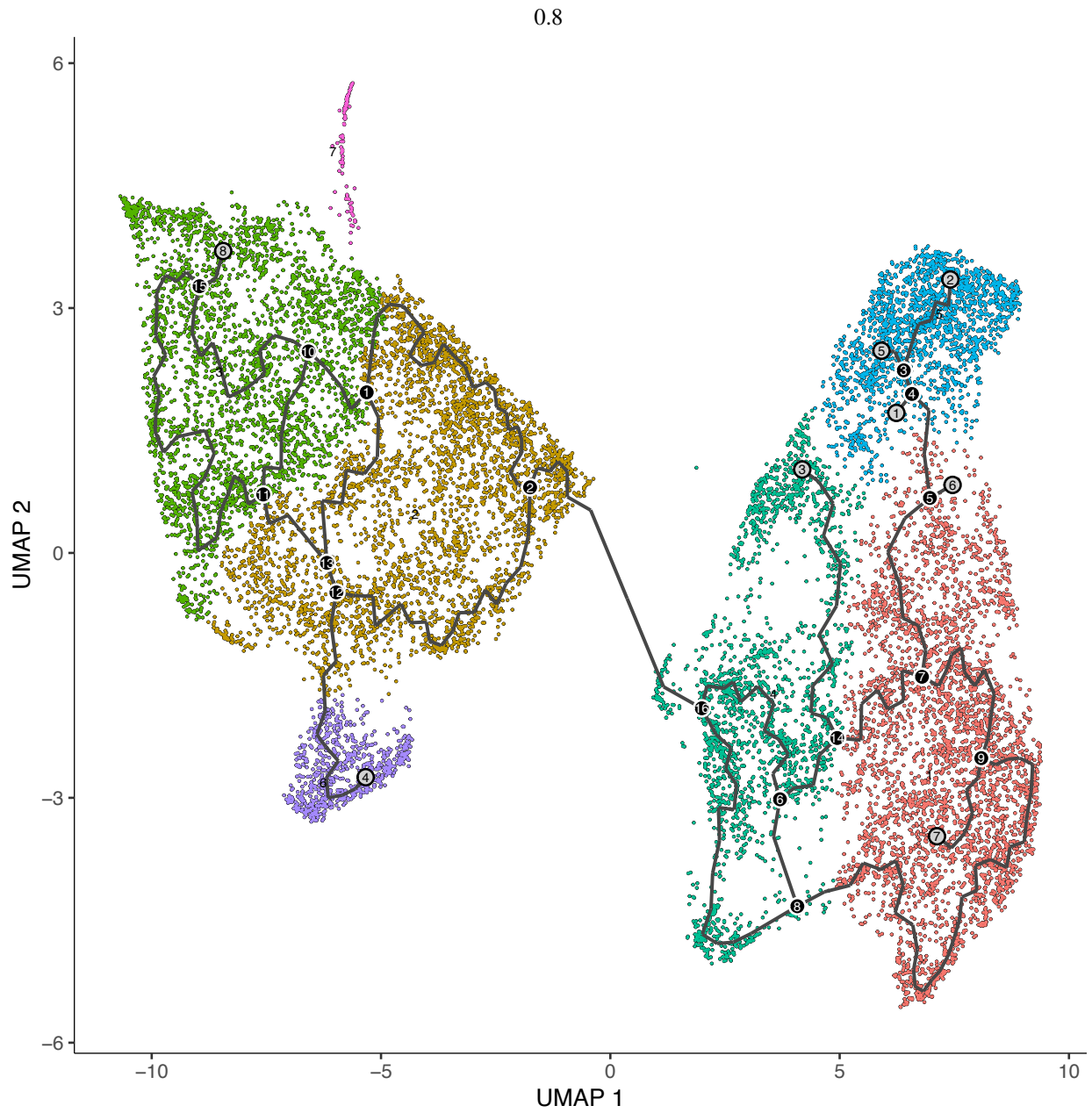


**Figure 26: gene expression vs. pseudotime in chronic HIV dataset**

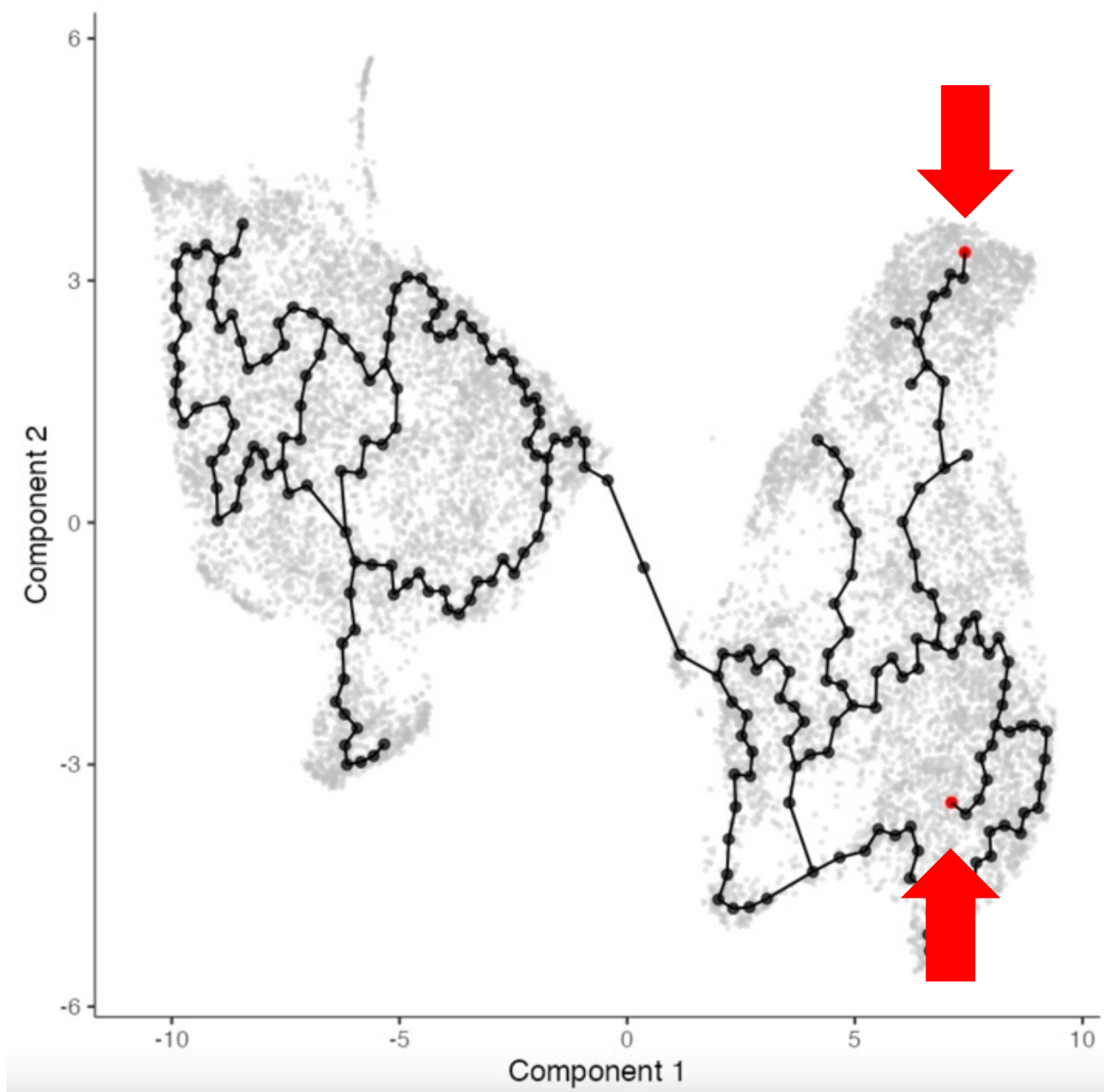
Expression of TCF7, the primary marker of the progenitor exhausted CD8 T cell state, starts high before increasing slightly and then decreasing across pseudotime. TOX, the primary marker of the terminally exhausted CD8 T cell state, is monotonically increasing across pseudotime. Immune checkpoint genes LAG3, PDCD1, and TIGIT are also monotonically increasing across pseudotime.



**Figure 27: CD8 T cell UMAP plot – expression of key genes – HIV dataset**



**Figure 28: CD8 T cell UMAP plot – Leiden clusters – HIV dataset**


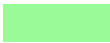



**Figure 28: CD8 T cell UMAP plot – Monocle3 graph and clusters – HIV dataset**

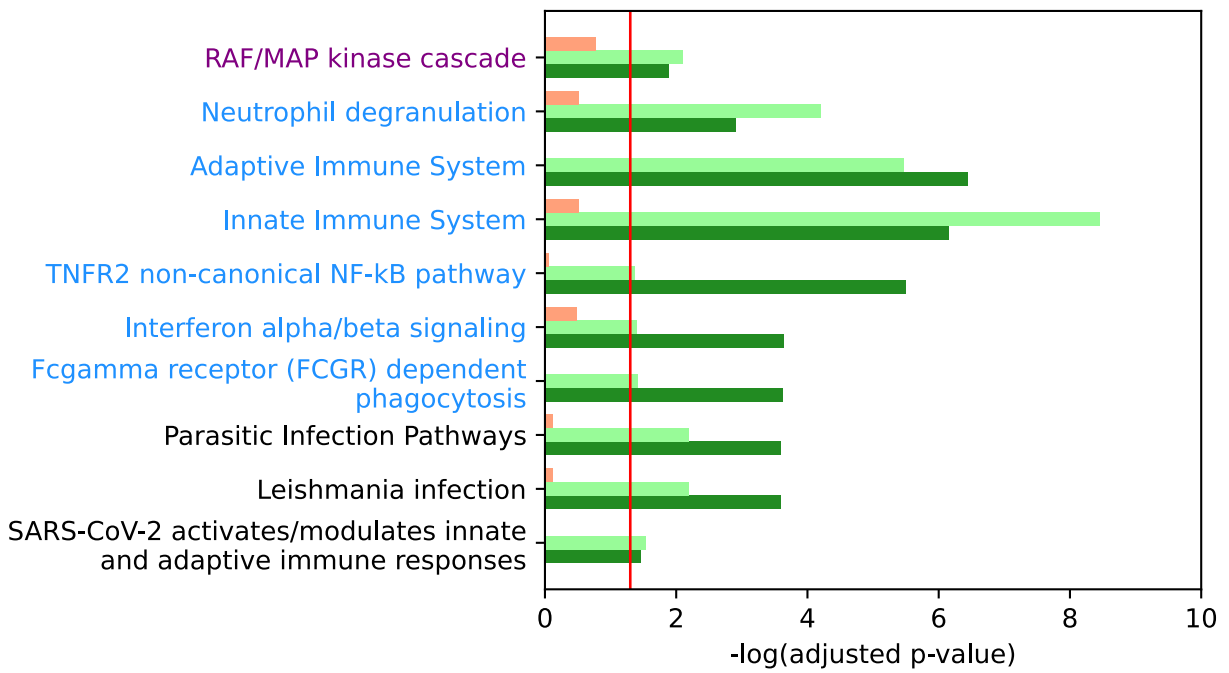
top: clusters; bottom: principal graph and root cells

**Parameters used:** Number of principal components: 10; Clustering: k=20, resolution=0.0001; 'learn\_graph()': use\_partition=FALSE

**PATHWAY FIGURES – KEYS:**

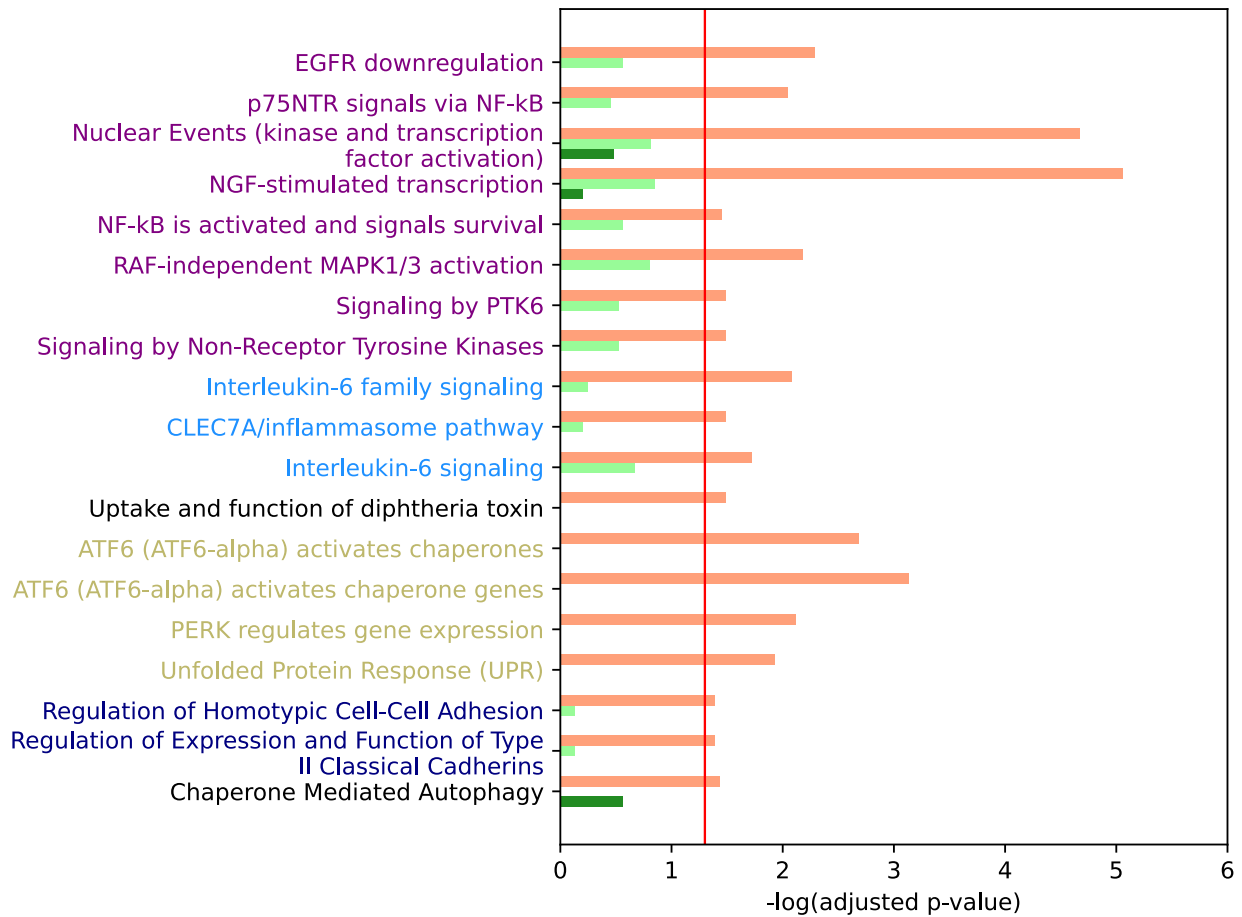
	tumor -- melanoma
	tumor -- BCC
	viral -- HIV

<b>Pathway Categories:</b>
<b>Immune System</b>
<b>Signal Transduction</b>
<b>Cell-Cell</b>
<b>Communication</b>
<b>Responses to Stimuli</b>
<b>Other</b>

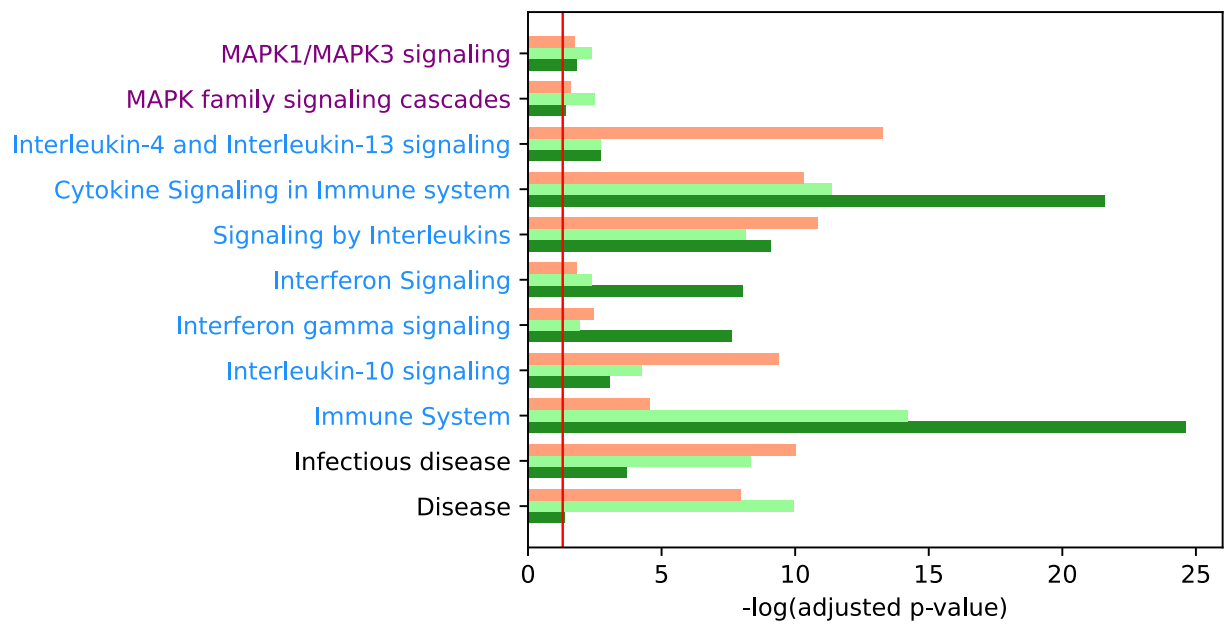


**Figure 29: Pathway enrichment of genes up-regulated in the most exhausted macrophage samples – unabridged, tumor-specific**

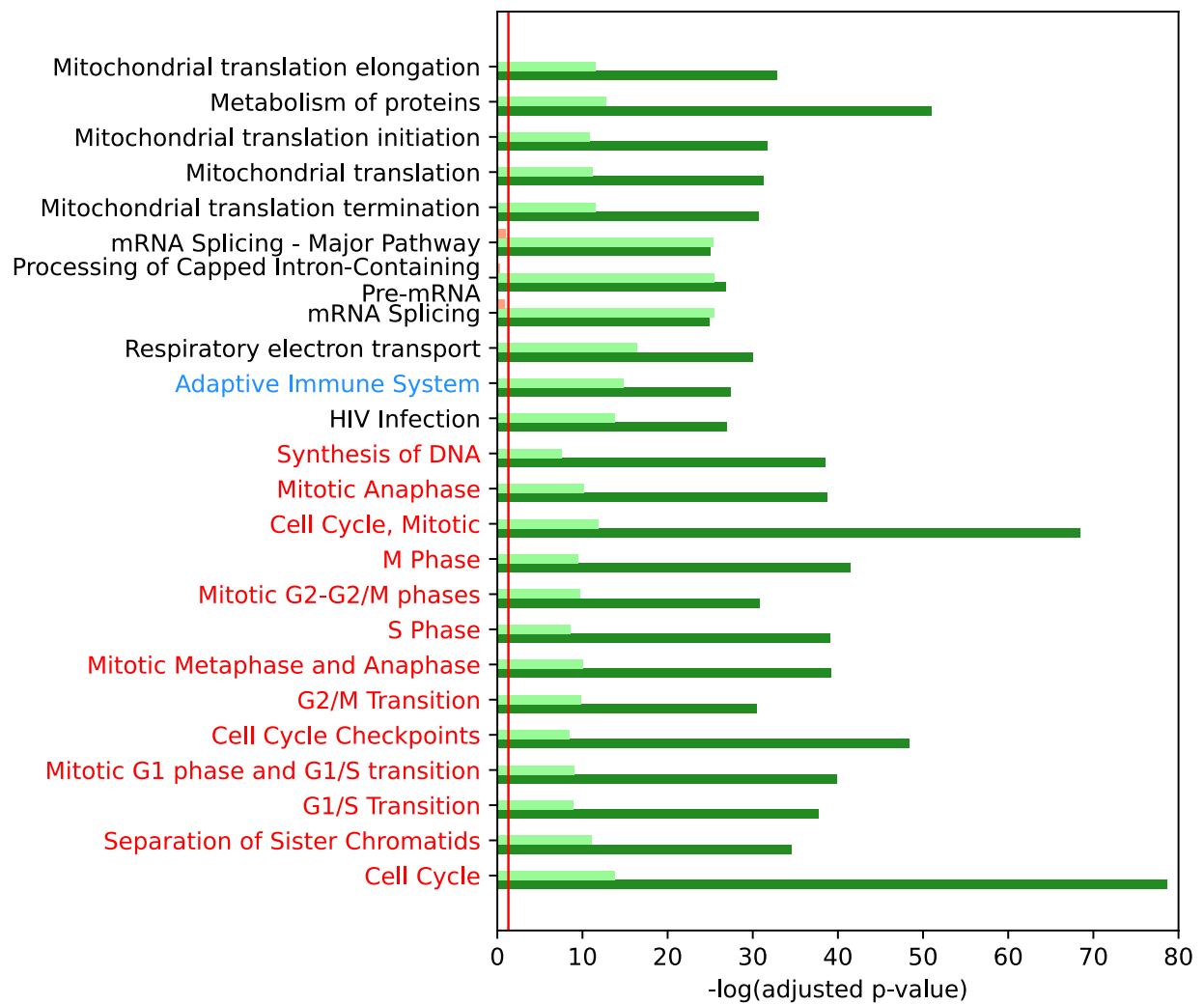




**Figure 30: Pathway enrichment of genes up-regulated in the most exhausted macrophage samples – unabridged, viral-specific**



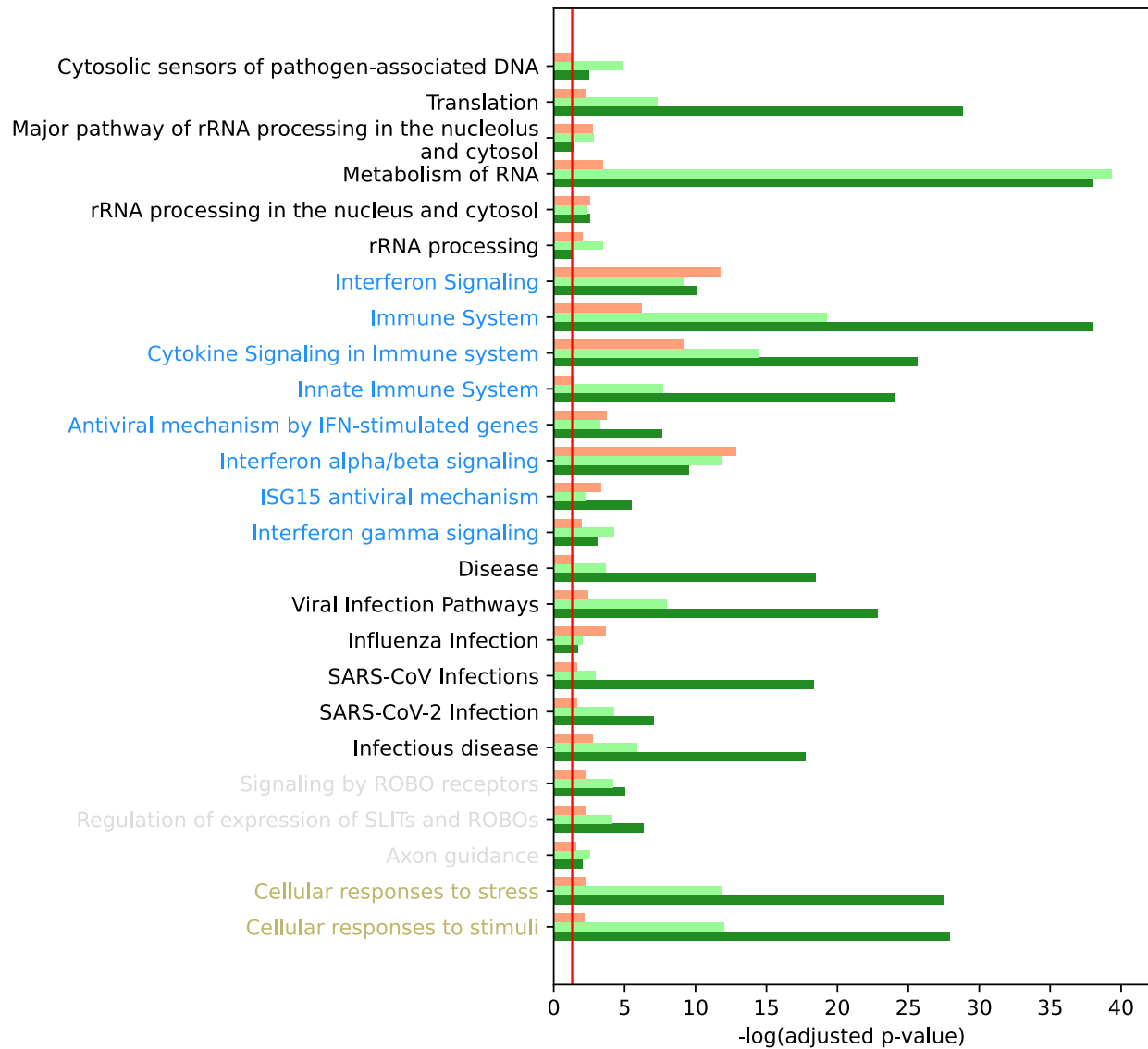
**Figure 31: Pathway enrichment of genes up-regulated in the most exhausted macrophage samples – unabridged, shared**



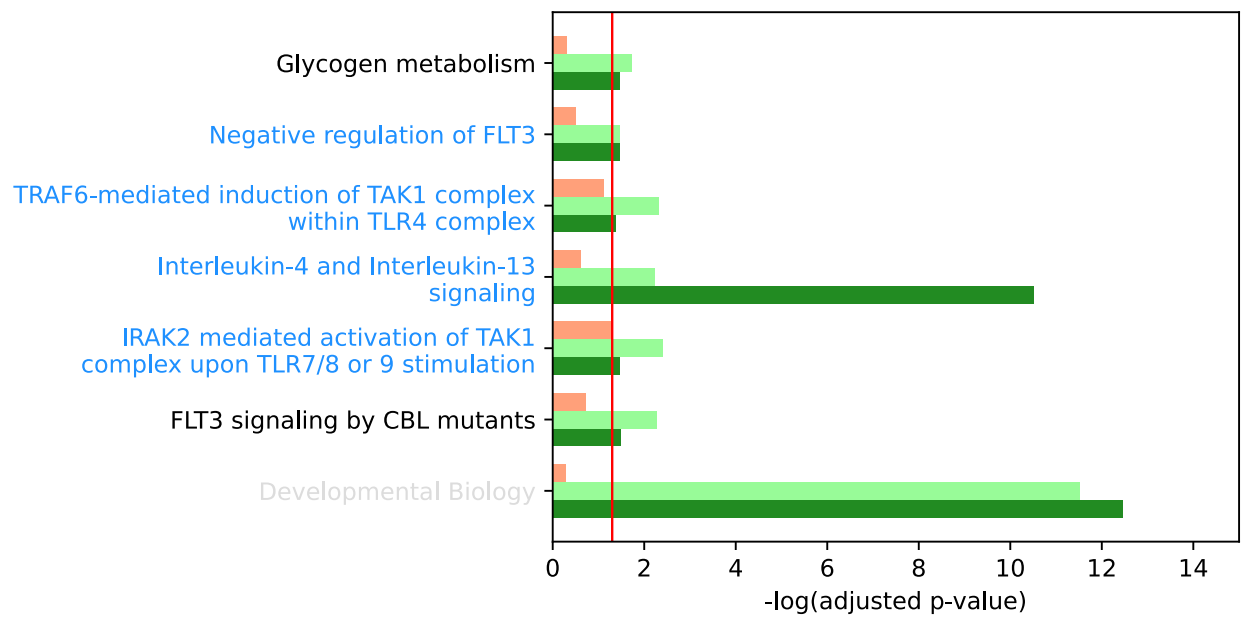
**Figure 32: Pathway enrichment of genes up-regulated in the most exhausted CD8 T cell samples – unabridged, tumor-specific**



**Figure 33: Pathway enrichment of genes up-regulated in the most exhausted CD8 T cell samples – unabridged, viral-specific**

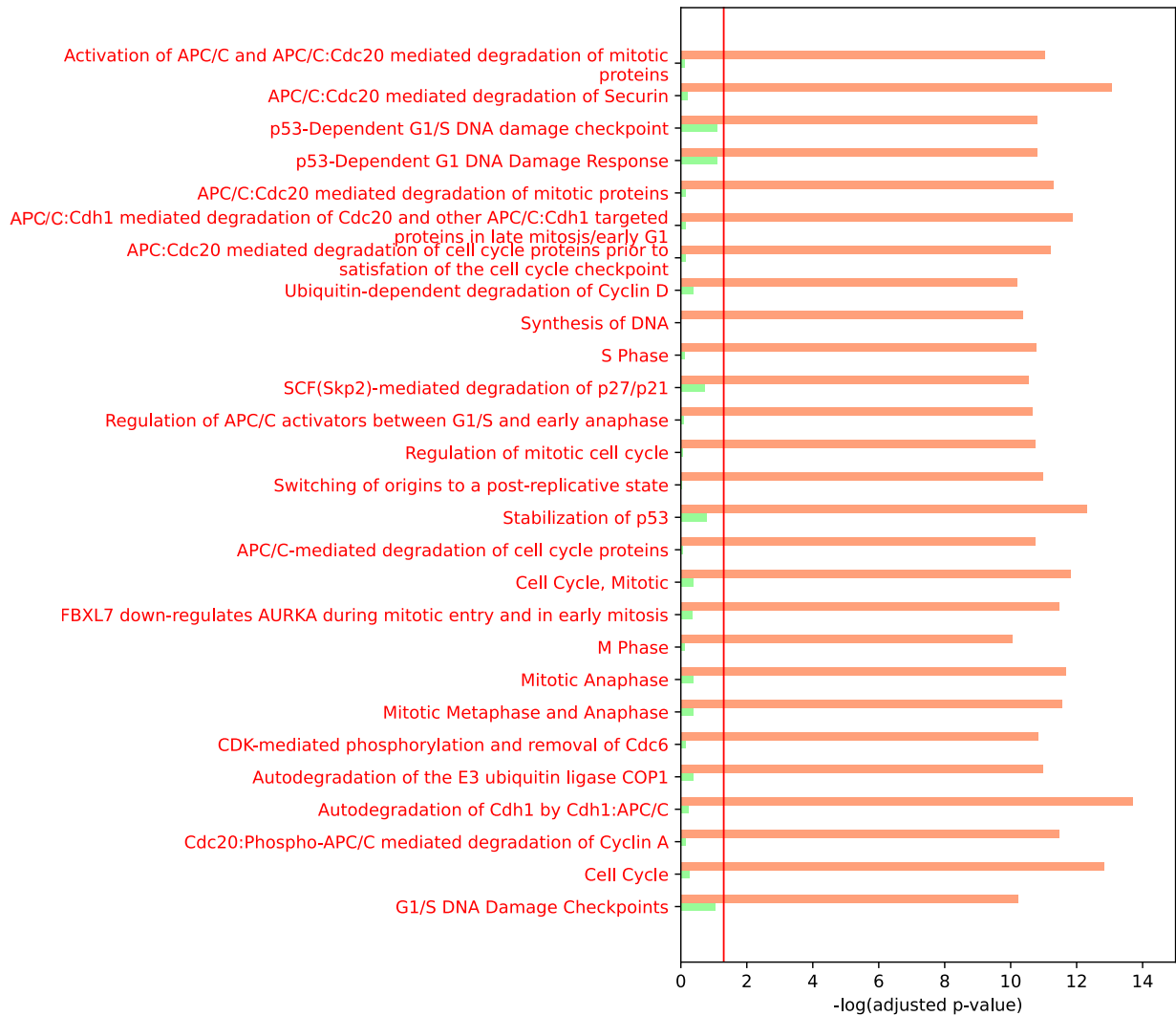


**Figure 34: Pathway enrichment of genes up-regulated in the most exhausted CD8 T cell samples – unabridged, shared**



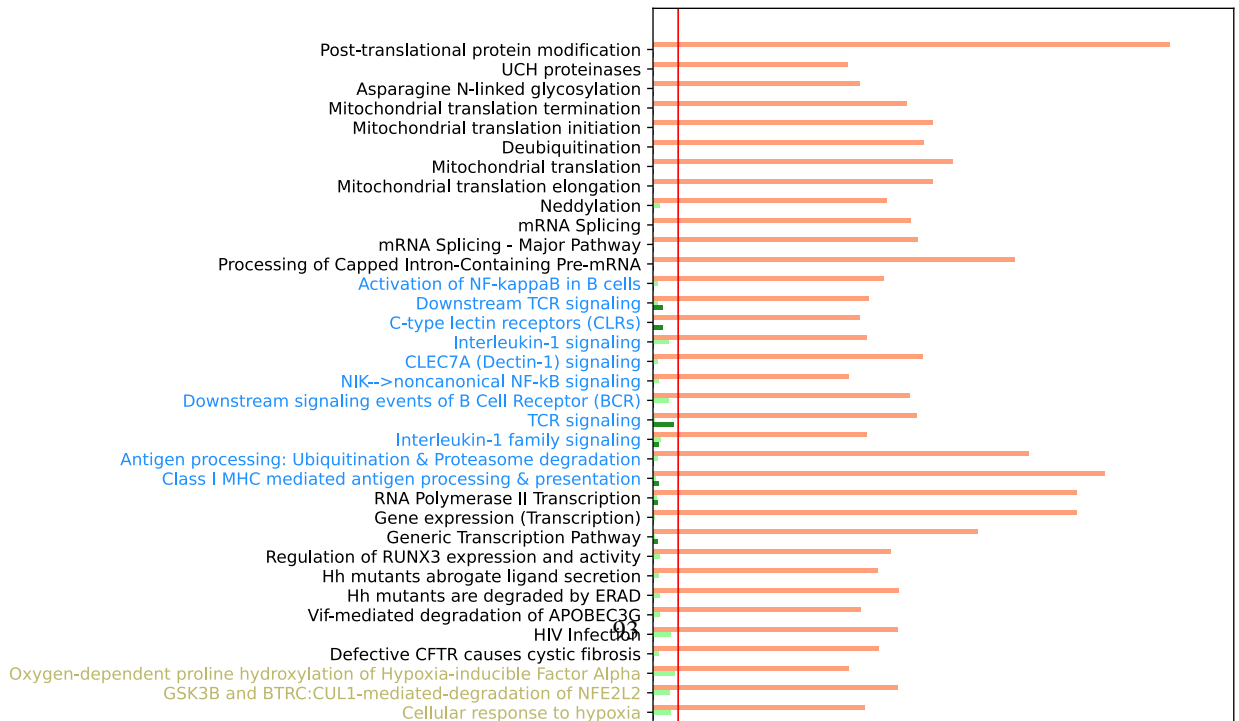
**Figure 35: Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, tumor-specific**

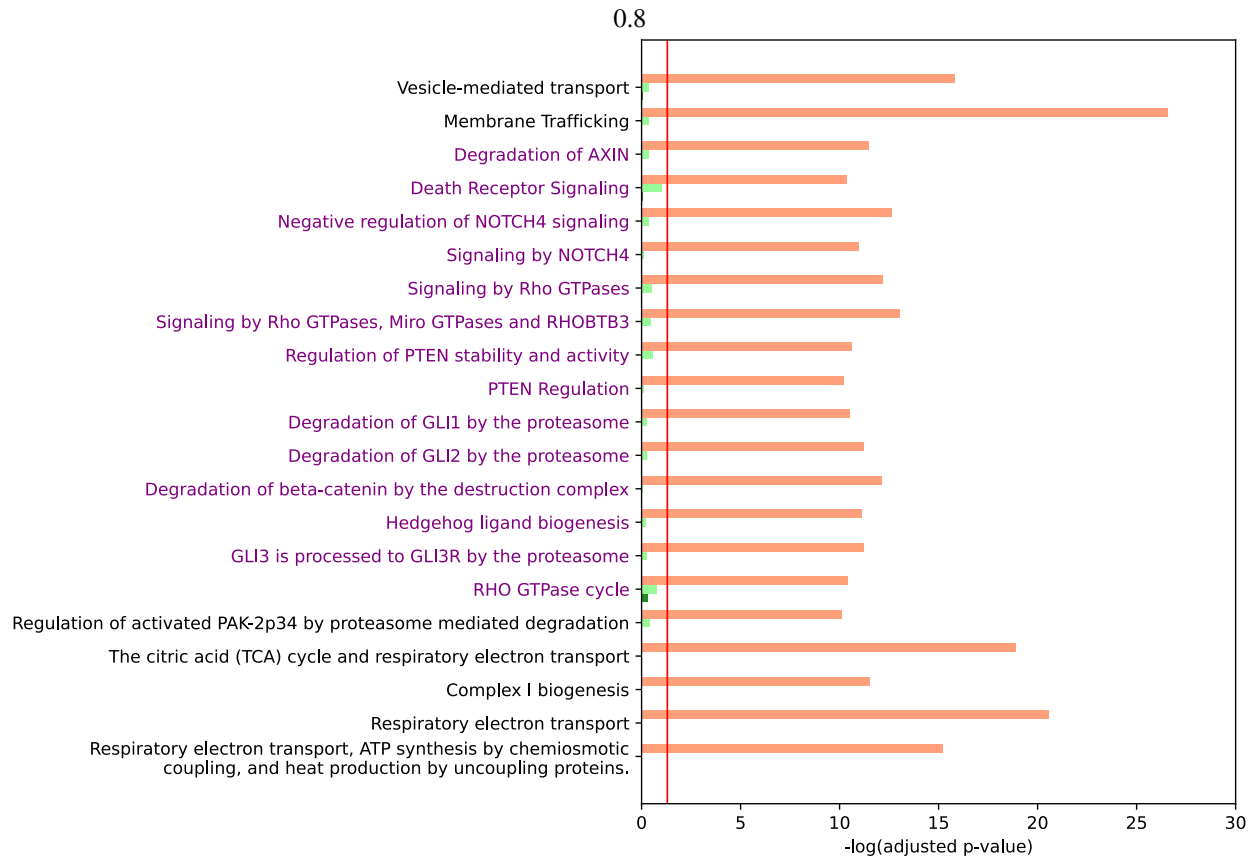
0.8



(part 1 of 3)

0.8

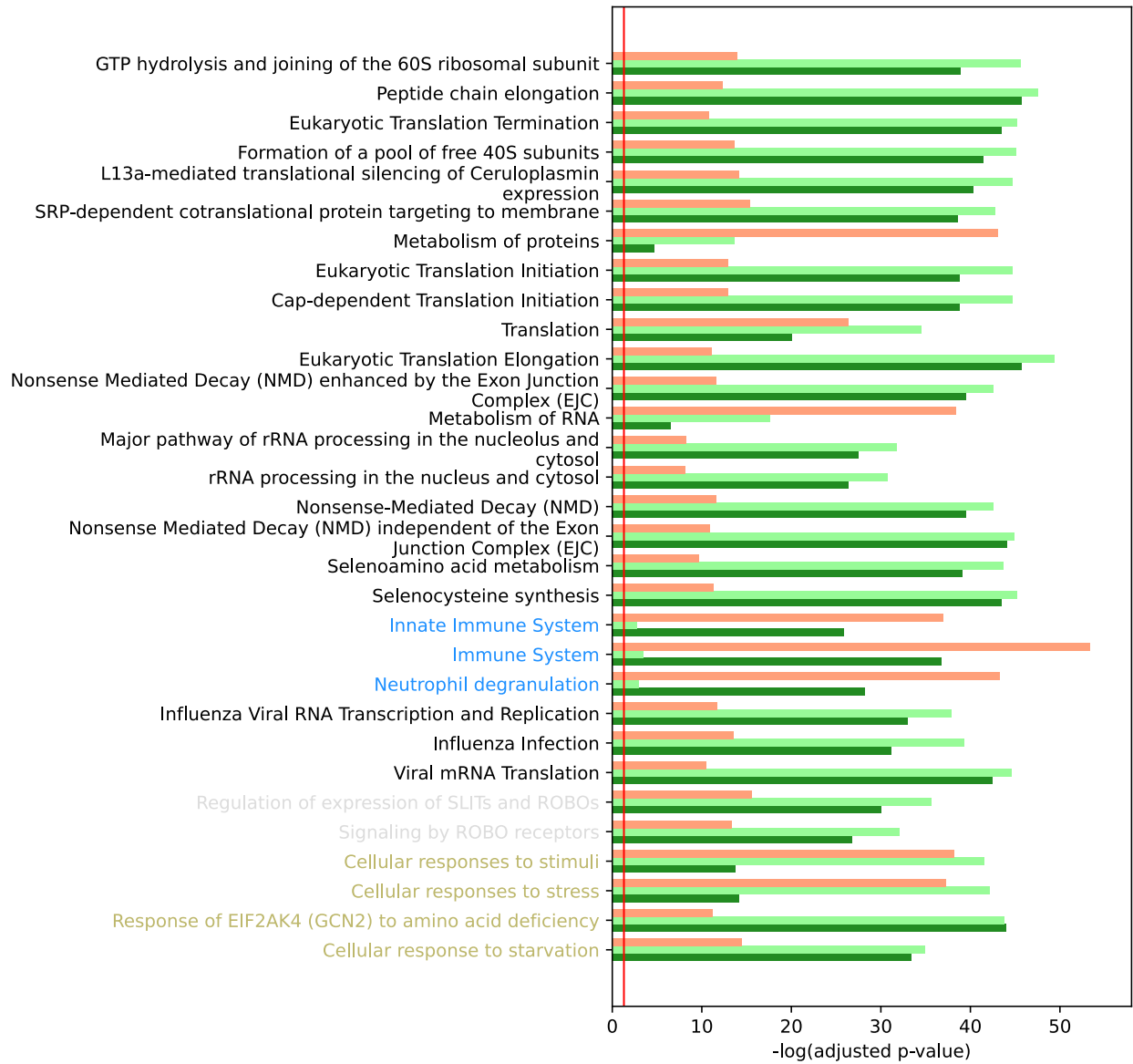




(part 3 of 3)

**Figure 36: Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, viral-specific**

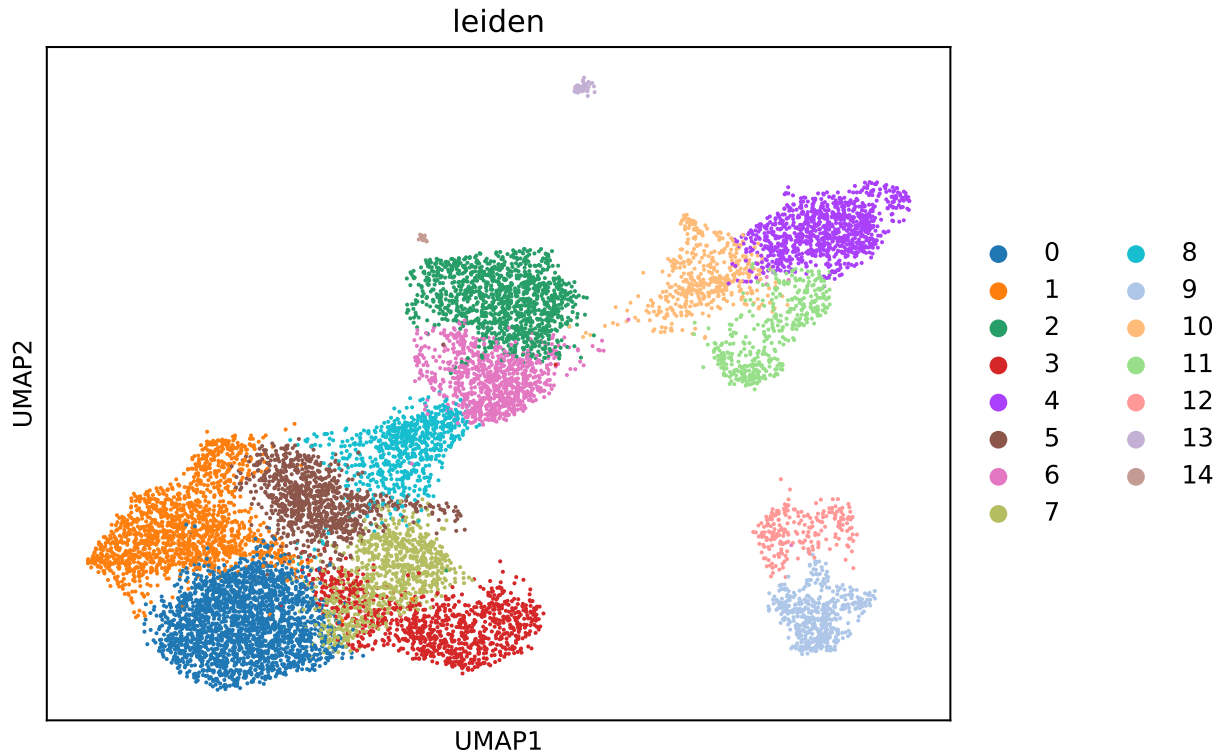




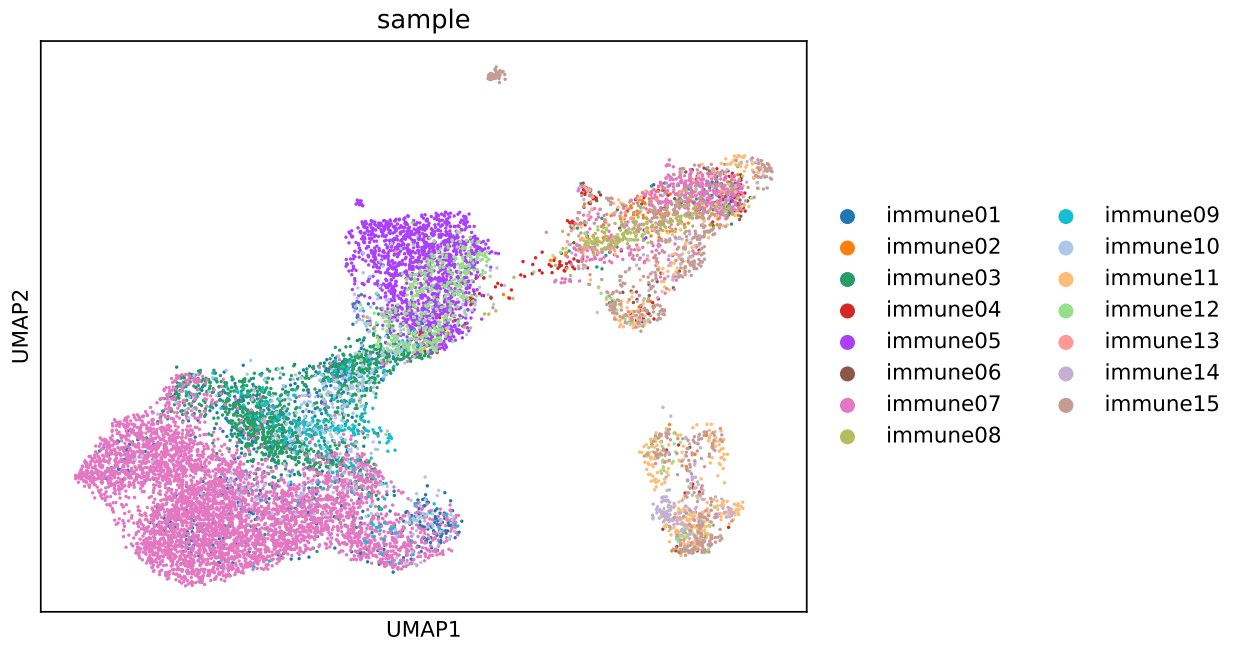
**Figure 37: Pathway enrichment of genes down-regulated in the most exhausted CD8 T cell samples – unabridged, shared**

cell type	Dataset		
	Li et al., 2019	Yost et al., 2019	Wang et al., 2020
CD8 T cells	19741	8136	15202
macrophages	5298	1144	1895
natural killer cells	5523	196	-
B cells	3999	111	4440
plasma cells	1283	1710	404
CD4 T cells	-	3619	-
T regulatory cells	-	2039	-
tumor cells	-	2492	-
cancer-associated fibroblasts	-	626	-
dendritic cells	-	396	-
endothelial cells	-	235	-
melanocytes	-	88	-
myofibroblasts	-	162	-
unknown	1717	-	-
total	37561	20954	21941

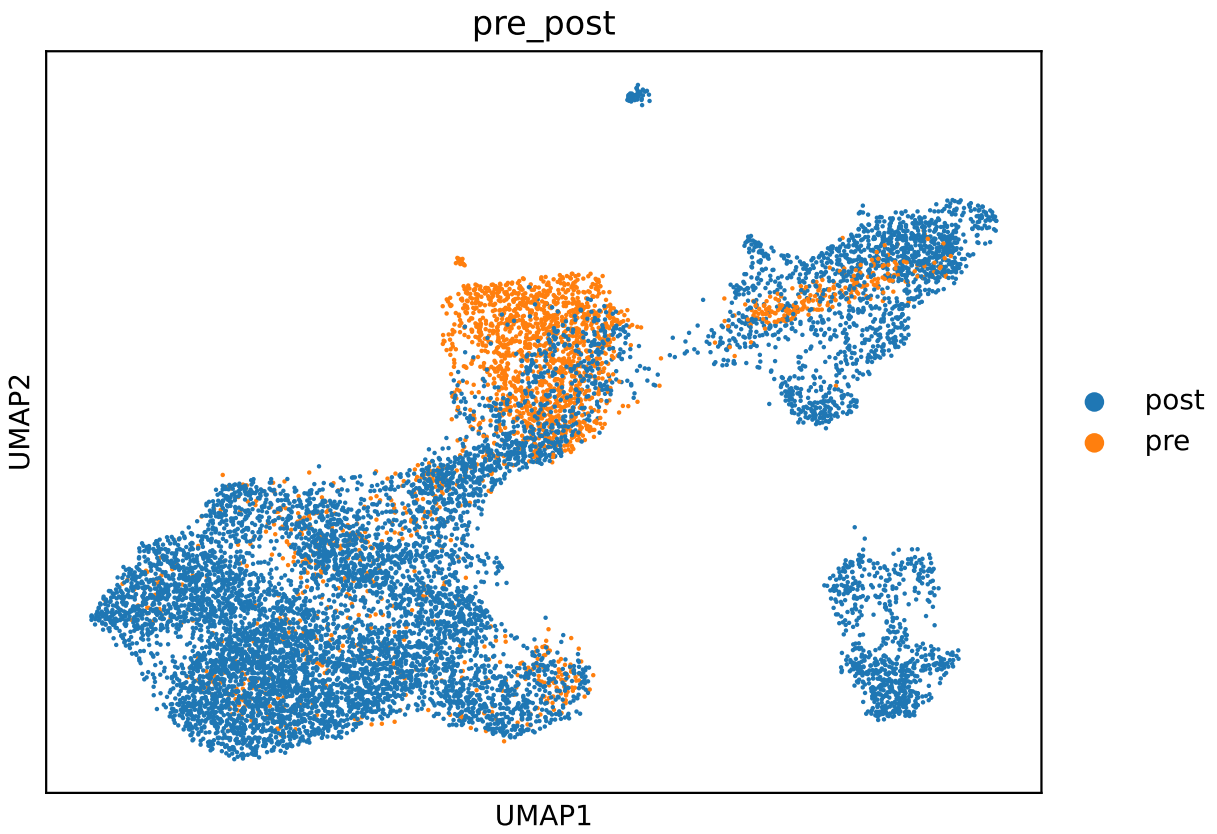
**Table 10: Cell counts by cell type for each of three datasets**



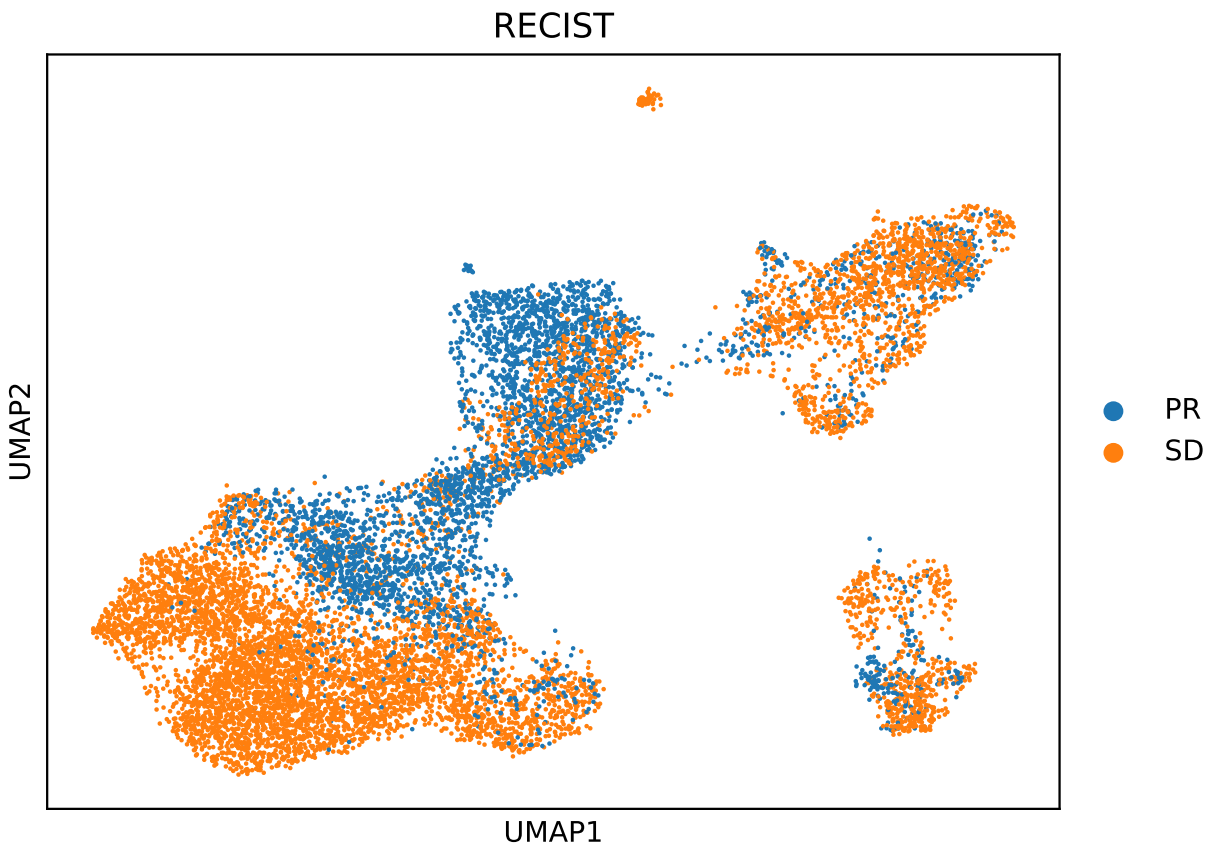
**Figure 38: Leiden clusters**  
*Hu et al., 2023* dataset (pre- and post-treatment tumor cells)



**Figure 39: Patients**  
*Hu et al., 2023* dataset (pre- and post-treatment tumor cells)

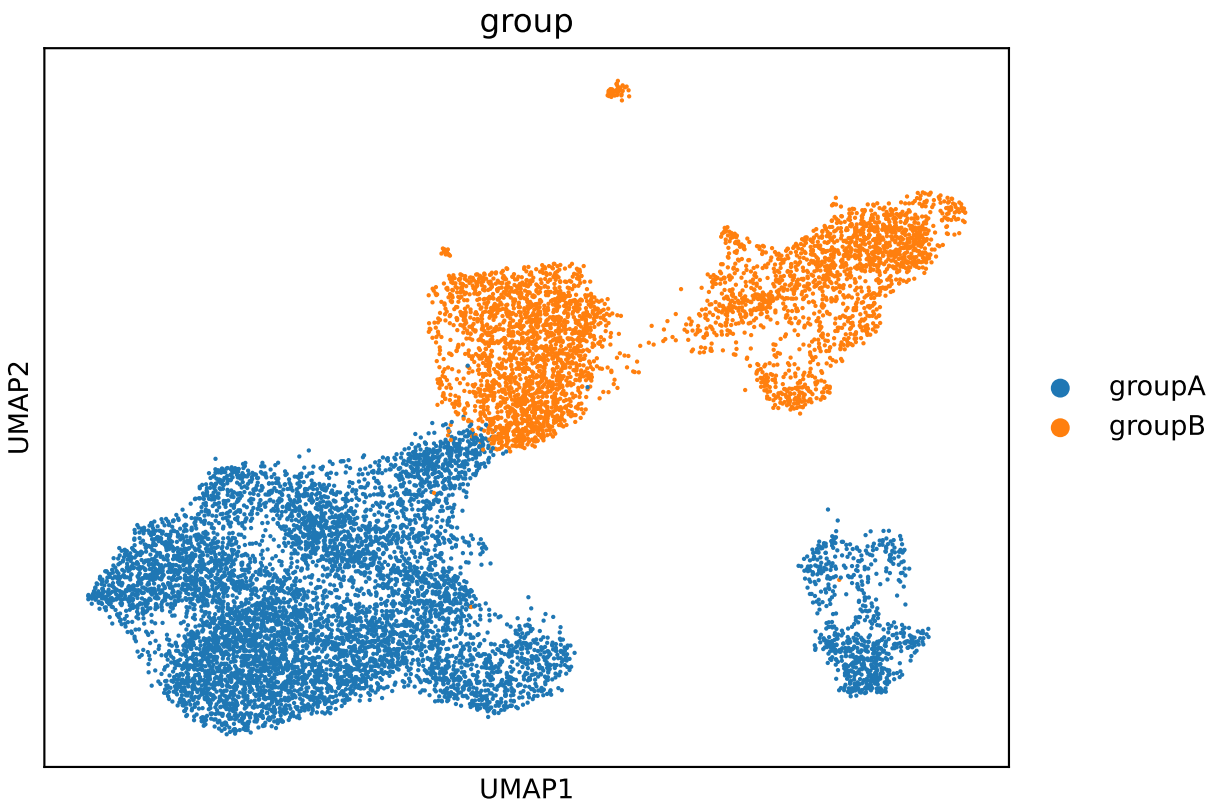


**Figure 40: Pre- vs. post-treatment**  
*Hu et al., 2023* dataset (pre- and post-treatment tumor cells)

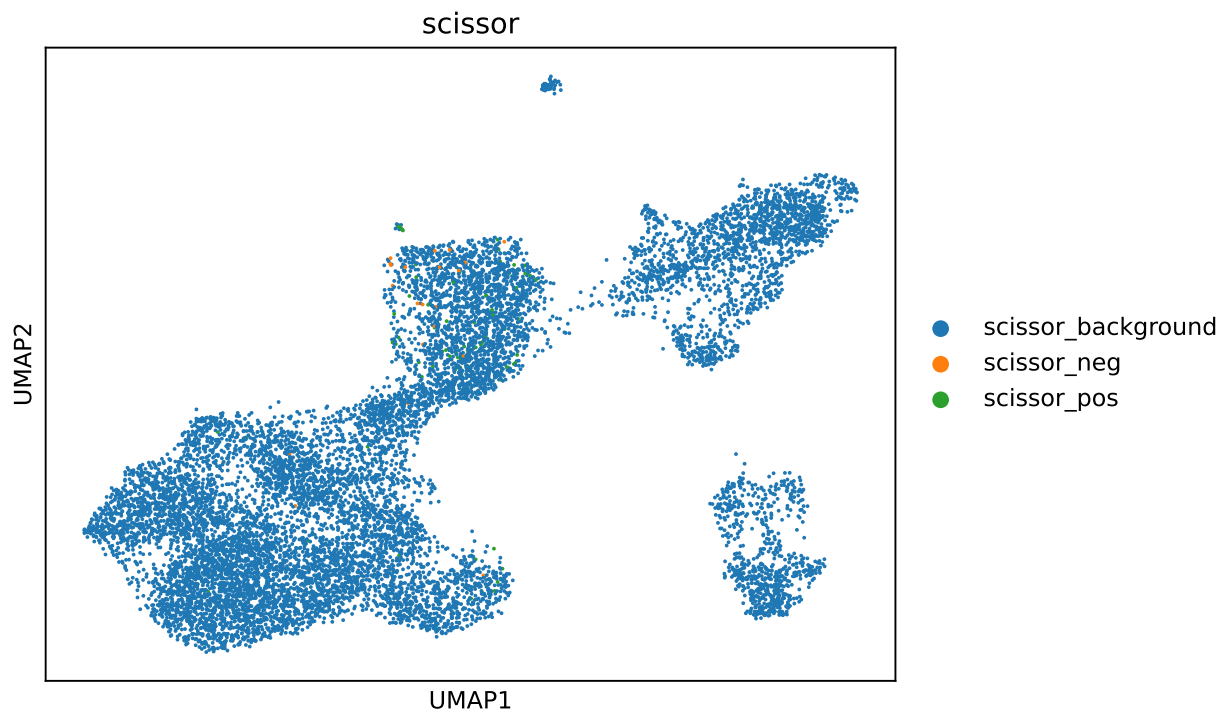


**Figure 41: Treatment outcome (RECIST)**

*Hu et al., 2023* dataset (pre- and post-treatment tumor cells); PR: Partial Response; SD: Stable Disease

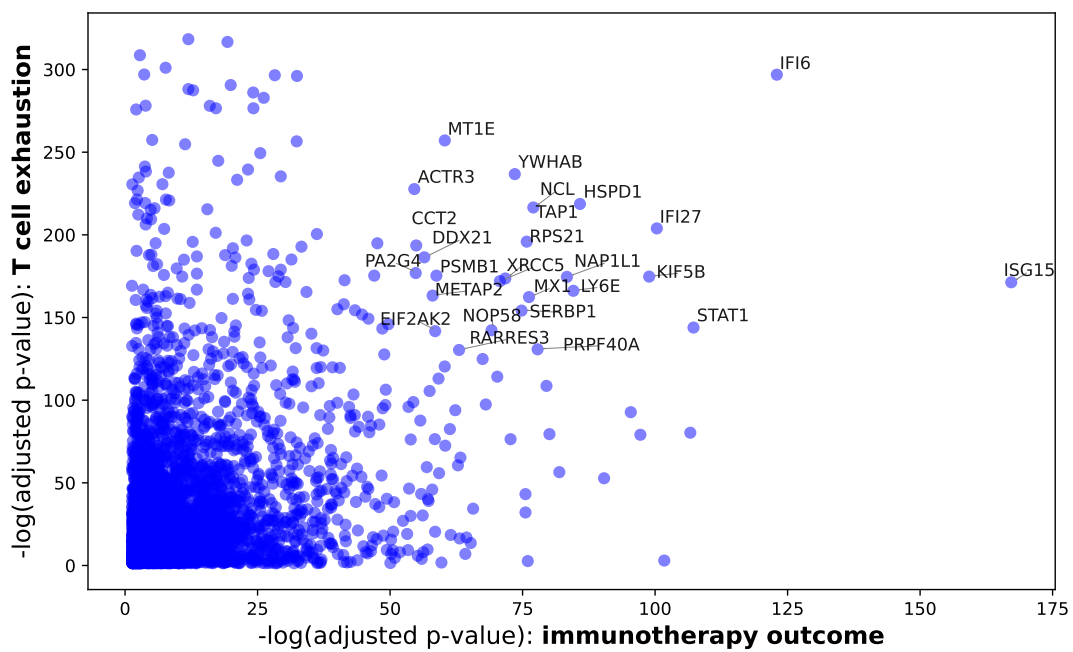


**Figure 42: Group**  
*Hu et al., 2023 dataset (pre- and post-treatment tumor cells)*



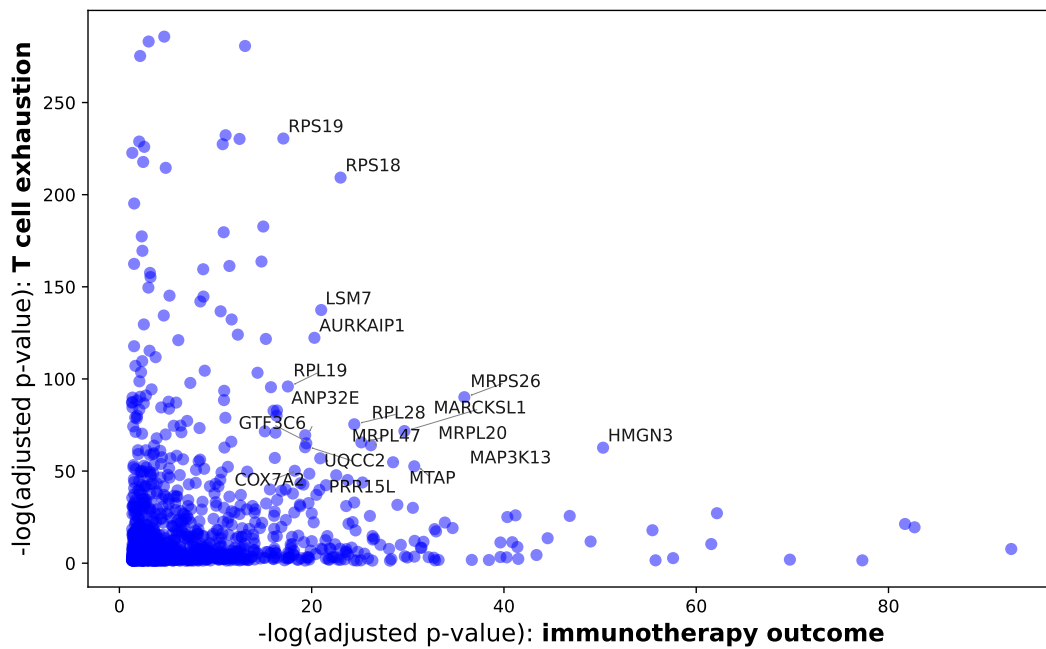
**Figure 43: Scissor status**

*Hu et al., 2023* dataset (pre- and post-treatment tumor cells); phenotype-associated cells selected by Scissor; Note: any Scissor-positive or Scissor-negative cells not from the Hu dataset not pictured here. Additionally, only pre-treatment cells were used for phenotype association analysis



**Figure 44: Significant DEG overlap – Scissor-positive and exhaustion-high**

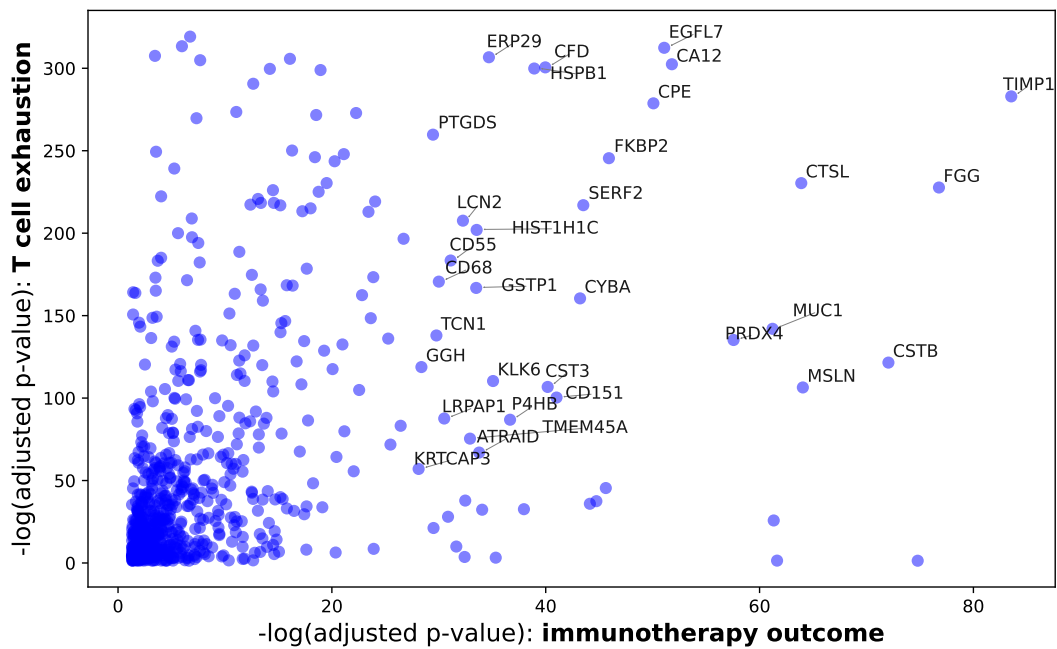
x-axis: genes whose expression is up-regulated in immunotherapy response-related tumor cells relative to immunotherapy resistance-related tumor cells; y-axis: genes whose expression is up-regulated in high CD8 T cell exhaustion-associated tumor cells relative to low exhaustion-associated tumor cells; significance quantified by  $-\log(\text{adjusted } p\text{-value})$



**Figure 45: Significant DEG overlap – Scissor-positive and exhaustion-low**

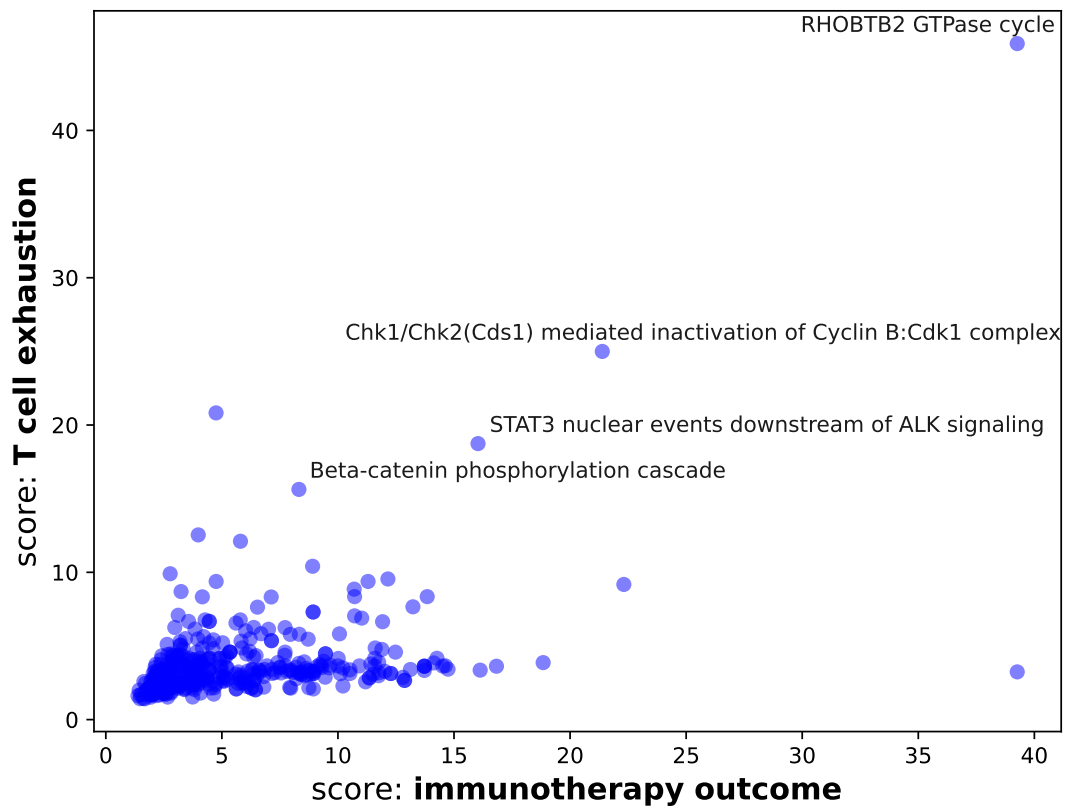
x-axis: genes whose expression is up-regulated in immunotherapy response-related tumor cells; y-axis: genes whose expression is up-regulated in low CD8 T cell exhaustion-associated tumor cells relative to high exhaustion-associated tumor cells; significance quantified by  $-\log(\text{adjusted p-value})$





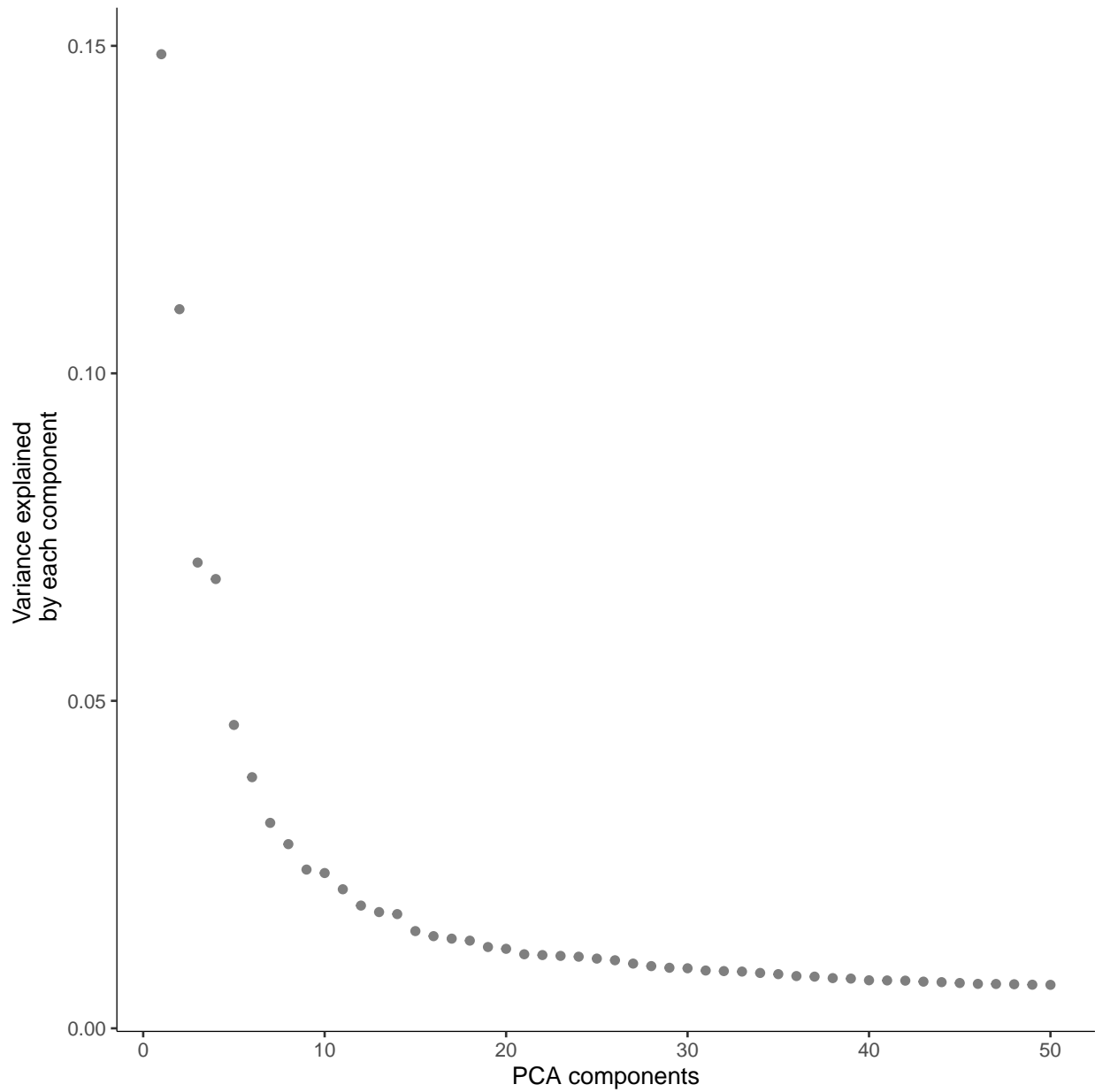
**Figure 46: Significant DEG overlap – Scissor-negative and exhaustion-low**

x-axis: genes whose expression is up-regulated in immunotherapy resistance-related tumor cells; y-axis: genes whose expression is up-regulated in low CD8 T cell exhaustion-associated tumor cells relative to high exhaustion-associated tumor cells; significance quantified by  $-\log(\text{adjusted p-value})$

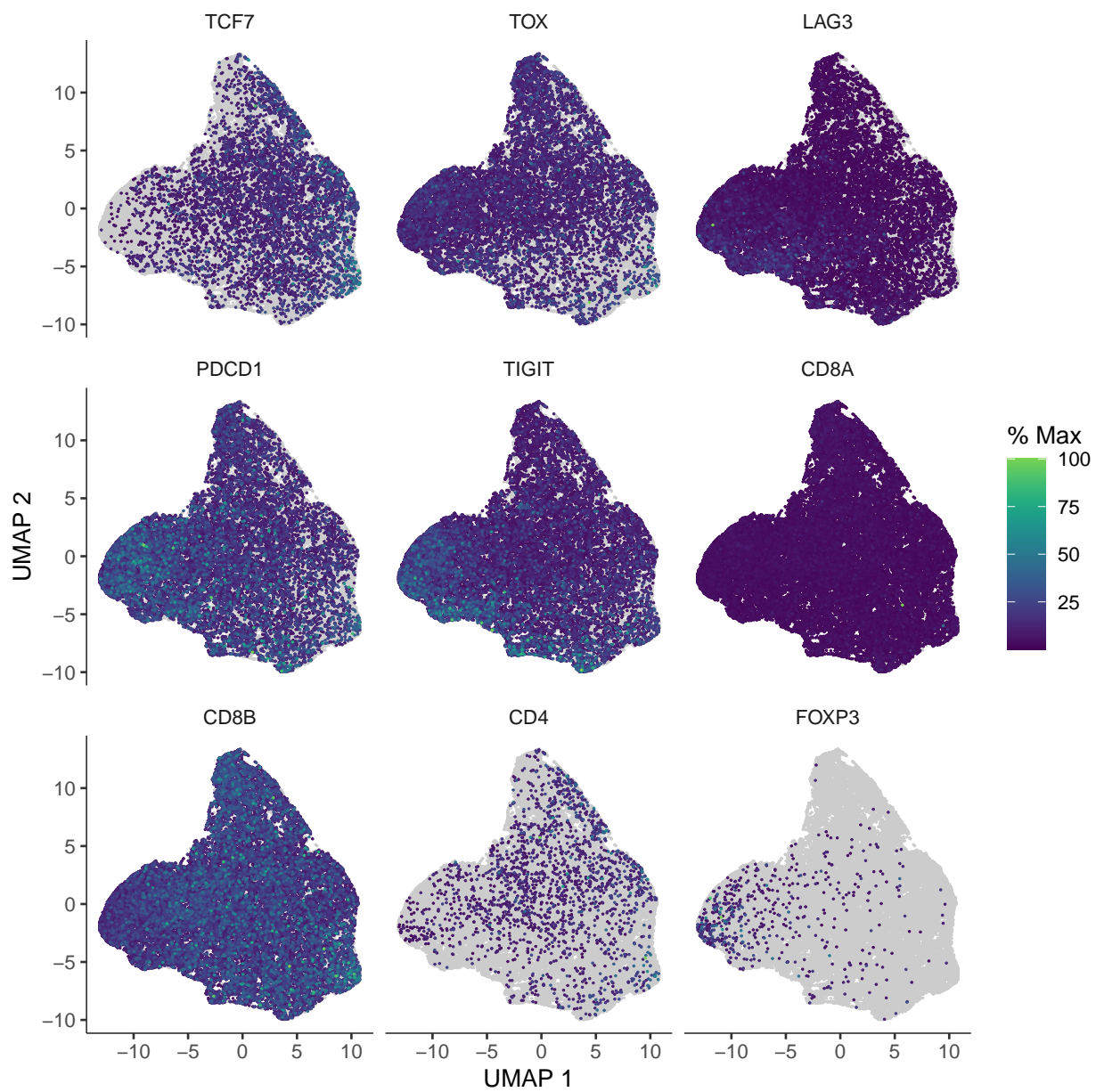


**Figure 47: Significant pathway overlap – Scissor-positive and exhaustion-high**

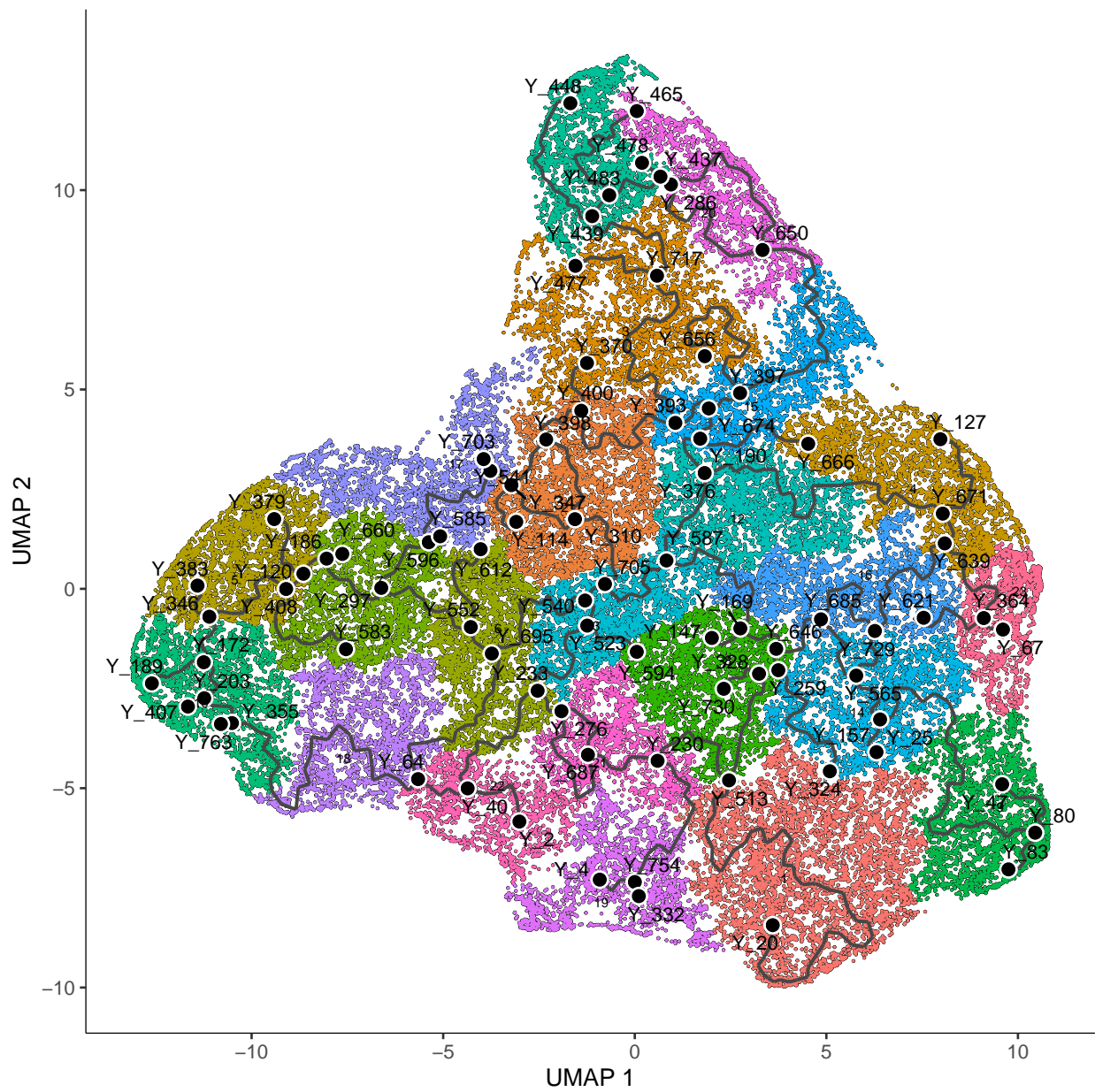
x-axis: pathways whose expression is up-regulated in immunotherapy response-related tumor cells relative to immunotherapy resistance-related tumor cells; y-axis: pathways whose expression is up-regulated in high CD8 T cell exhaustion-associated tumor cells relative to low exhaustion-associated tumor cells



**Figure 48: Percentage of variance explained by principal components – NSCLC CD8 T cells**  
five principal components were used for CD8 T cell trajectory



**Figure 49: Marker genes for setup of CD8 T cell exhaustion trajectory – CD8 T cells from NSCLC tumors**



**Figure 50: Leiden clusters and Monocle 3 principal graph – CD8 T cells from NSCLC tumors**  
 principal point Y\_83 used as root for initialization of Monocle 3 trajectory inference

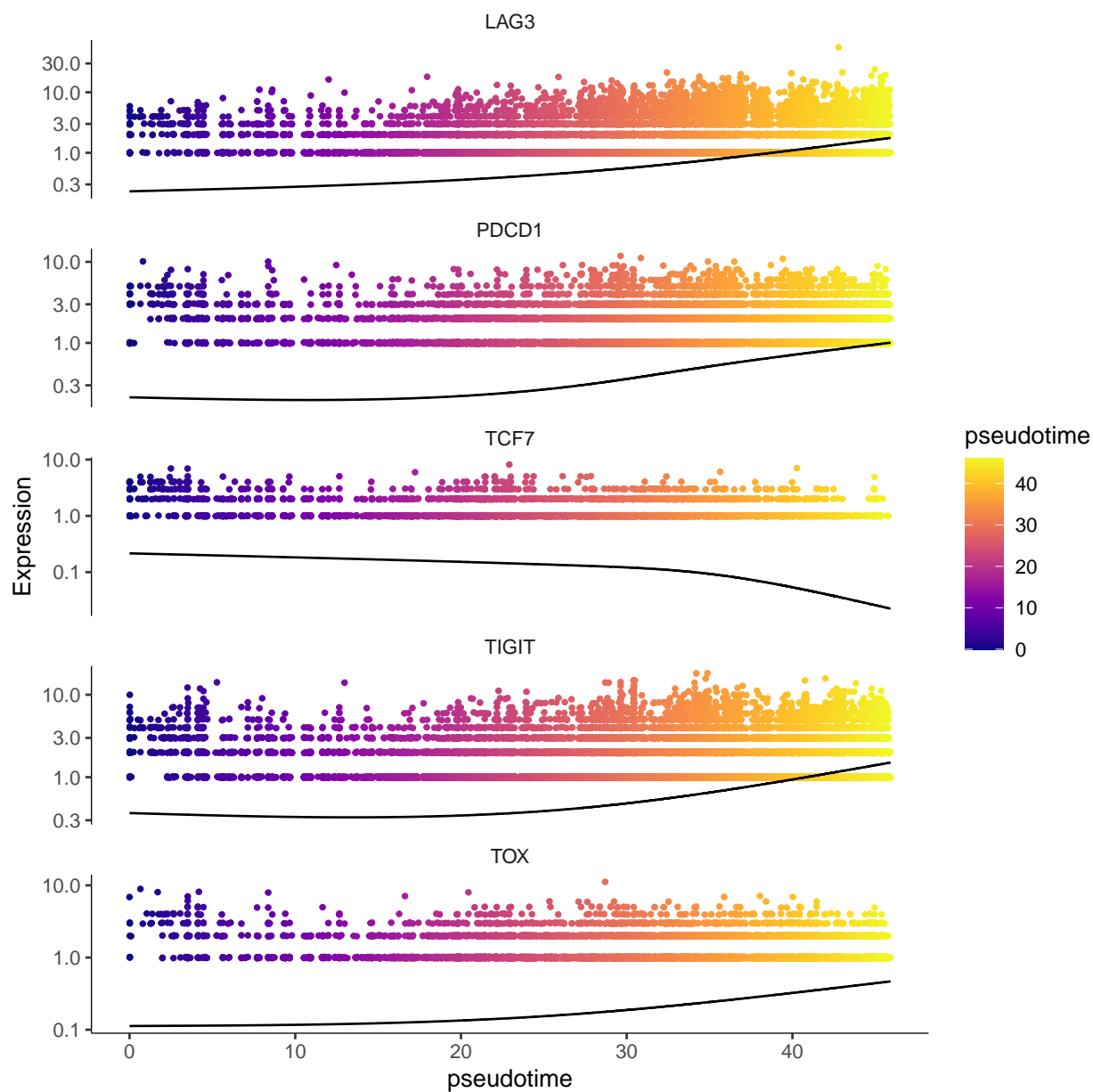
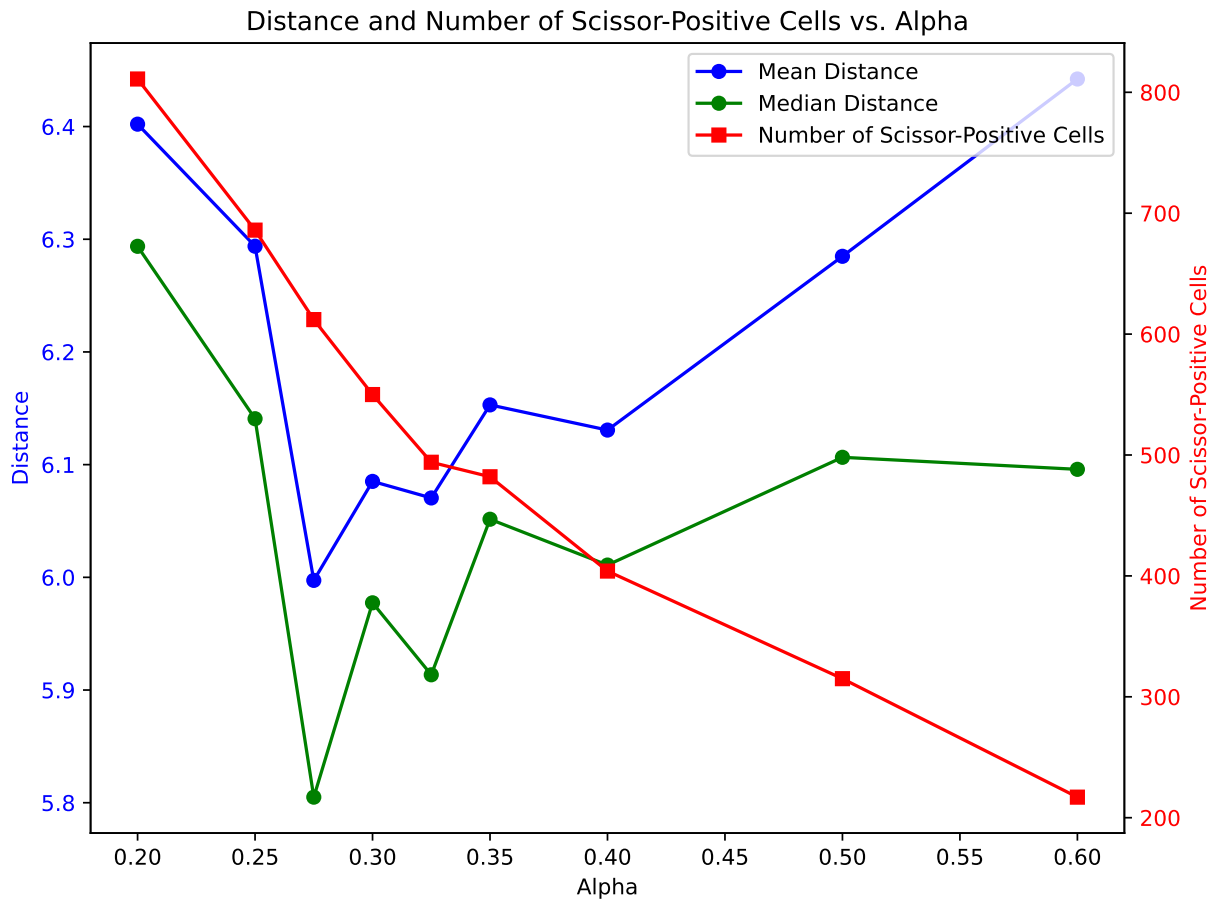
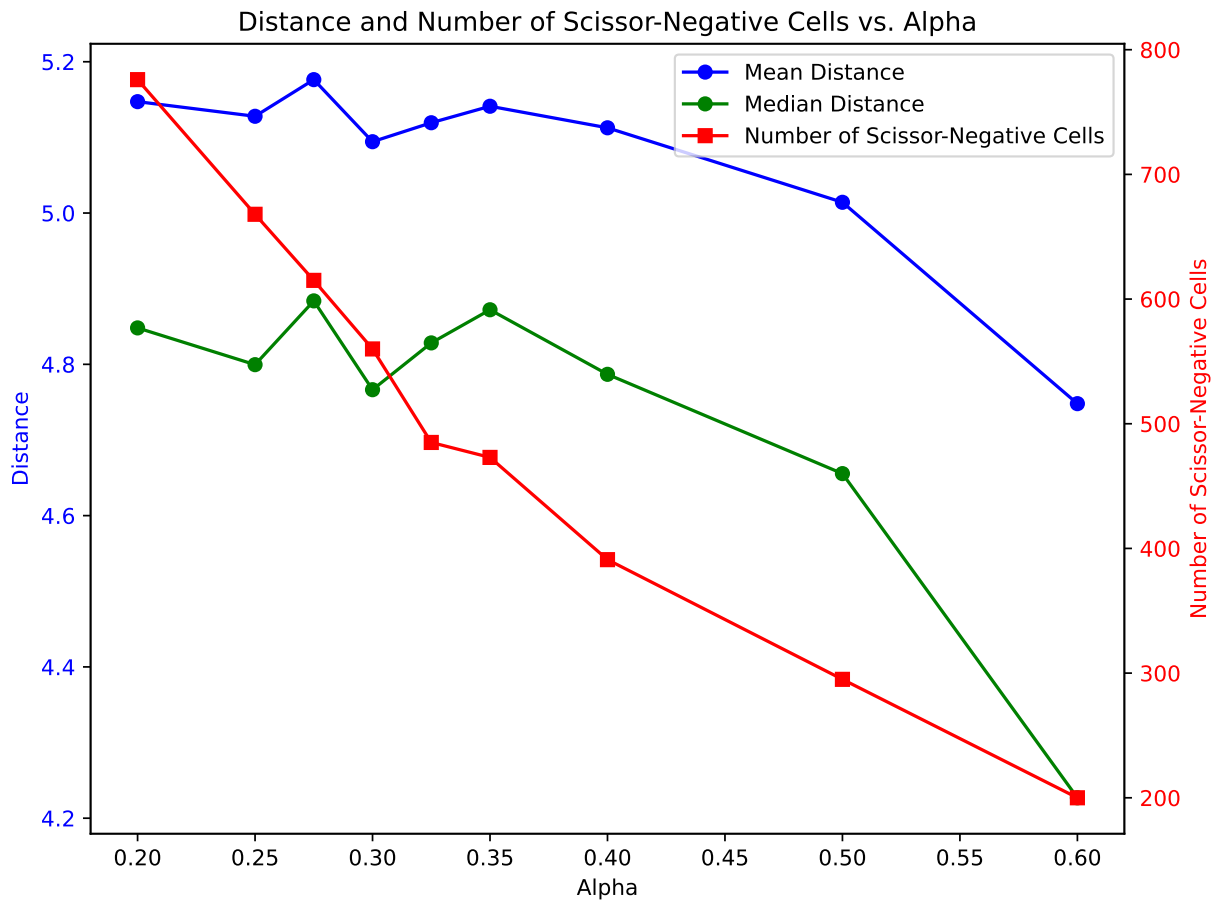


Figure 51: Expression of key genes vs. pseudotime – CD8 T cells from NSCLC tumors



**Figure 52: Optimization of Scissor’s alpha parameter, part 1**

An optimal value of Scissor’s alpha parameter was chosen for the lung tumor cells by considering the tradeoff between a higher value, which selects fewer phenotype-associated cells for downstream analysis, and a lower value, which may select more heterogeneous cells with a wider range of expression profiles. In order to choose a clearly defined subset of the tumor cells, in sufficient quantity, a value of 0.275 was determined to be optimal.



**Figure 53: Optimization of Scissor's alpha parameter, part 2**  
See description of previous figure.





## Confirmation of Publication and Licensing Rights - Open Access

September 10th, 2024

**Subscription Type:** *Individual - Academic*  
**Agreement number:** *RB27A15X39*  
**Name of Publication:** *OHSU dissertation*

**Figure Title:** *CD8 T cell destroying tumor cell*

**Citation to Use:** *Created in BioRender. Anand, S. (2023) [BioRender.com/b97u537](https://www.biorender.com/b97u537)*

To whom this may concern,

This document ("Confirmation") hereby confirms that Science Suite Inc. dba BioRender ("BioRender") has granted the following BioRender user: S Anand ("User") a BioRender Academic Publication License in accordance with BioRender's [Terms of Service](#) and [Academic License Terms](#) ("License Terms") to permit such User to do the following on the condition that all requirements in this Confirmation are met:

- 1) publish their Completed Graphics created in the BioRender Services containing both User Content and BioRender Content (as both are defined in the License Terms) in publications (journals, textbooks, websites, etc.); and
- 2) sublicense such Completed Graphics under "open access" publication sublicensing models such as CC-BY 4.0 and more restrictive models, so long as the conditions set forth herein are fully met.

Requirements of User:

- 1) All Completed Graphics to be published in any publication (journals, textbooks, websites, etc.) must be accompanied by the following citation either as a caption, footnote or reference for each figure that includes a Completed Graphic:  
"Created in BioRender. Anand, S. (2023) [BioRender.com/b97u537](https://www.biorender.com/b97u537)".
- 2) All terms of the License Terms including all Prohibited Uses are fully complied with. E.g. For Academic License Users, no commercial uses (beyond publication in journals, textbooks or websites) are permitted without obtaining or switching to a BioRender Industry Plan.
- 3) A Reader (defined below) may request that the User allow their figure to be a public template for Readers to view, copy, and modify the figure. It is up to the User to determine what level of access to grant.

Open-Access Journal Readers:

Open-Access journal readers ("Reader") who wish to view and/or re-use a particular Completed Graphic in an Open-Access journal subject to CC-BY sublicensing may do so by clicking on the URL link in the applicable citation for the subject Completed Graphic.

The re-use/modification options below are available after the Reader requests the User to adapt their

### Figure 54: biorender publication license for Figure 2



## Confirmation of Publication and Licensing Rights - Open Access

September 10th, 2024

**Subscription Type:** *Individual - Academic*  
**Agreement number:** *AN27A17FHJ*  
**Name of Publication:** *OHSU dissertation*

**Figure Title:** *Immunosurveillance*

**Citation to Use:** *Created in BioRender. Anand, S. (2024) [BioRender.com/g83z800](https://www.biorender.com/g83z800)*

To whom this may concern,

This document ("Confirmation") hereby confirms that Science Suite Inc. dba BioRender ("BioRender") has granted the following BioRender user: S Anand ("User") a BioRender Academic Publication License in accordance with BioRender's [Terms of Service](#) and [Academic License Terms](#) ("License Terms") to permit such User to do the following on the condition that all requirements in this Confirmation are met:

- 1) publish their Completed Graphics created in the BioRender Services containing both User Content and BioRender Content (as both are defined in the License Terms) in publications (journals, textbooks, websites, etc.); and
- 2) sublicense such Completed Graphics under "open access" publication sublicensing models such as CC-BY 4.0 and more restrictive models, so long as the conditions set forth herein are fully met.

Requirements of User:

- 1) All Completed Graphics to be published in any publication (journals, textbooks, websites, etc.) must be accompanied by the following citation either as a caption, footnote or reference for each figure that includes a Completed Graphic:  
"Created in BioRender. Anand, S. (2024) [BioRender.com/g83z800](https://www.biorender.com/g83z800) ".
- 2) All terms of the License Terms including all Prohibited Uses are fully complied with. E.g. For Academic License Users, no commercial uses (beyond publication in journals, textbooks or websites) are permitted without obtaining or switching to a BioRender Industry Plan.
- 3) A Reader (defined below) may request that the User allow their figure to be a public template for Readers to view, copy, and modify the figure. It is up to the User to determine what level of access to grant.

Open-Access Journal Readers:

Open-Access journal readers ("Reader") who wish to view and/or re-use a particular Completed Graphic in an Open-Access journal subject to CC-BY sublicensing may do so by clicking on the URL link in the applicable citation for the subject Completed Graphic.

The re-use/modification options below are available after the Reader requests the User to adapt their

### Figure 55: biorender publication license for Figure 3



## Confirmation of Publication and Licensing Rights - Open Access

September 10th, 2024

**Subscription Type:** *Individual - Academic*  
**Agreement number:** *PS27A17P0K*  
**Name of Publication:** *OHSU dissertation*

**Figure Title:** *Relationship between immune cell activity and the degree of CD8 T cell exhaustion in the tumor immune microenvironment*

**Citation to Use:** *Created in BioRender. Anand, S. (2024) [BioRender.com/n35q609](https://www.biorender.com/n35q609)*

To whom this may concern,

This document ("Confirmation") hereby confirms that Science Suite Inc. dba BioRender ("BioRender") has granted the following BioRender user: [S Anand](#) ("User") a BioRender Academic Publication License in accordance with BioRender's [Terms of Service](#) and [Academic License Terms](#) ("License Terms") to permit such User to do the following on the condition that all requirements in this Confirmation are met:

- 1) publish their Completed Graphics created in the BioRender Services containing both User Content and BioRender Content (as both are defined in the License Terms) in publications (journals, textbooks, websites, etc.); and
- 2) sublicense such Completed Graphics under "open access" publication sublicensing models such as CC-BY 4.0 and more restrictive models, so long as the conditions set forth herein are fully met.

Requirements of User:

- 1) All Completed Graphics to be published in any publication (journals, textbooks, websites, etc.) must be accompanied by the following citation either as a caption, footnote or reference for each figure that includes a Completed Graphic:  
"Created in BioRender. Anand, S. (2024) [BioRender.com/n35q609](https://www.biorender.com/n35q609)".
- 2) All terms of the License Terms including all Prohibited Uses are fully complied with. E.g. For Academic License Users, no commercial uses (beyond publication in journals, textbooks or websites) are permitted without obtaining or switching to a BioRender Industry Plan.
- 3) A Reader (defined below) may request that the User allow their figure to be a public template for Readers to view, copy, and modify the figure. It is up to the User to determine what level of access to grant.

Open-Access Journal Readers:

Open-Access journal readers ("Reader") who wish to view and/or re-use a particular Completed Graphic in an Open-Access journal subject to CC-BY sublicensing may do so by clicking on the URL link in the applicable citation for the subject Completed Graphic.

### Figure 56: biorender publication license for Figure 4



## Confirmation of Publication and Licensing Rights - Open Access

September 10th, 2024

**Subscription Type:** *Individual - Academic*  
**Agreement number:** *NC27A17YWF*  
**Name of Publication:** *OHSU dissertation*

**Figure Title:** *Heterogeneity in tumor subpopulation modulation of and response to anti-tumor immune activity*

**Citation to Use:** *Created in BioRender. Anand, S. (2024) [BioRender.com/e28j663](https://www.biorender.com/e28j663)*

To whom this may concern,

This document ("Confirmation") hereby confirms that Science Suite Inc. dba BioRender ("BioRender") has granted the following BioRender user: [S Anand](#) ("User") a BioRender Academic Publication License in accordance with BioRender's [Terms of Service](#) and [Academic License Terms](#) ("License Terms") to permit such User to do the following on the condition that all requirements in this Confirmation are met:

- 1) publish their Completed Graphics created in the BioRender Services containing both User Content and BioRender Content (as both are defined in the License Terms) in publications (journals, textbooks, websites, etc.); and
- 2) sublicense such Completed Graphics under "open access" publication sublicensing models such as CC-BY 4.0 and more restrictive models, so long as the conditions set forth herein are fully met.

Requirements of User:

- 1) All Completed Graphics to be published in any publication (journals, textbooks, websites, etc.) must be accompanied by the following citation either as a caption, footnote or reference for each figure that includes a Completed Graphic:  
"Created in BioRender. Anand, S. (2024) [BioRender.com/e28j663](https://www.biorender.com/e28j663)".
- 2) All terms of the License Terms including all Prohibited Uses are fully complied with. E.g. For Academic License Users, no commercial uses (beyond publication in journals, textbooks or websites) are permitted without obtaining or switching to a BioRender Industry Plan.
- 3) A Reader (defined below) may request that the User allow their figure to be a public template for Readers to view, copy, and modify the figure. It is up to the User to determine what level of access to grant.

Open-Access Journal Readers:

Open-Access journal readers ("Reader") who wish to view and/or re-use a particular Completed Graphic in an Open-Access journal subject to CC-BY sublicensing may do so by clicking on the URL link in the applicable citation for the subject Completed Graphic.

### Figure 57: biorender publication license for Figure 12

## 6 References

- [1] Y. Shiravand, F. Khodadadi, S. M. A. Kashani, S. R. Hosseini-Fard, S. Hosseini, H. Sadeghirad, R. Ladwa, K. O'Byrne, and A. Kulasinghe. Immune checkpoint inhibitors in cancer therapy. *Curr Oncol*, 29(5):3044–3060, 2022. Shiravand, Yavar Khodadadi, Faezeh Kashani, Seyyed Mohammad Amin Hosseini-Fard, Seyed Reza Hosseini, Shadi Sadeghirad, Habib Ladwa, Rahul O'Byrne, Ken Kulasinghe, Arutha eng Research Support, Non-U.S. Gov't Review Switzerland 2022/05/28 *Curr Oncol*. 2022 Apr 24;29(5):3044-3060. doi: 10.3390/curroncol29050247.
- [2] A. Kalbasi and A. Ribas. Tumour-intrinsic resistance to immune checkpoint blockade. *Nat Rev Immunol*, 20(1):25–39, 2020. Kalbasi, Anusha Ribas, Antoni eng K08 CA245181/CA/NCI NIH HHS/ L30 CA220808/CA/NCI NIH HHS/ R35 CA197633/CA/NCI NIH HHS/ Research Support, Non-U.S. Gov't Review England 2019/10/02 *Nat Rev Immunol*. 2020 Jan;20(1):25-39. doi: 10.1038/s41577-019-0218-4. Epub 2019 Sep 30.
- [3] K. E. Yost, A. T. Satpathy, D. K. Wells, Y. Qi, C. Wang, R. Kageyama, K. L. McNamara, J. M. Granja, K. Y. Sarin, R. A. Brown, R. K. Gupta, C. Curtis, S. L. Bucktrout, M. M. Davis, A. L. S. Chang, and H. Y. Chang. Clonal replacement of tumor-specific t cells following pd-1 blockade. *Nat Med*, 25(8):1251–1259, 2019. Yost, Kathryn E Satpathy, Ansuman T Wells, Daniel K Qi, Yanyan Wang, Chunlin Kageyama, Robin McNamara, Katherine L Granja, Jeffrey M Sarin, Kavita Y Brown, Ryanne A Gupta, Rohit K Curtis, Christina Bucktrout, Samantha L Davis, Mark M Chang, Anne Lynn S Chang, Howard Y eng RM1 HG007735/HG/NHGRI NIH HHS/ L30 CA134161/CA/NCI NIH HHS/ U19 AI057229/AI/NIAID NIH HHS/ L30 HL134161/HL/NHLBI NIH HHS/ P50 HG007735/HG/NHGRI NIH HHS/ K08 CA230188/CA/NCI NIH HHS/ R35 CA209919/CA/NCI NIH HHS/ K23 CA211793/CA/NCI NIH HHS/ S10 OD018220/OD/NIH HHS/ P30 CA124435/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2019/07/31 *Nat Med*. 2019 Aug;25(8):1251-1259. doi: 10.1038/s41591-019-0522-3. Epub 2019 Jul 29.
- [4] L. Jerby-Arnon, P. Shah, M. S. Cuoco, C. Rodman, M. J. Su, J. C. Melms, R. Leeson, A. Kanodia, S. Mei, J. R. Lin, S. Wang, B. Rabasha, D. Liu, G. Zhang, C. Margolais, O. Ashenberg, P. A. Ott, E. I. Buchbinder, R. Haq, F. S. Hodi, G. M. Boland, R. J. Sullivan, D. T. Frederick, B. Miao, T. Moll, K. T. Flaherty, M. Herlyn, R. W. Jenkins, R. Thummalapalli, M. S. Kowalczyk, I. Canadas, B. Schilling, A. N. R. Cartwright, A. M. Luoma, S. Malu, P. Hwu, C. Bernatchez, M. A. Forget, D. A. Barbie, A. K. Shalek, I. Tirosh, P. K. Sorger, K. Wucherpfennig, E. M. Van Allen, D. Schadendorf, B. E. Johnson, A. Rotem, O. Rozenblatt-Rosen, L. A. Garraway, C. H. Yoon, B. Izar, and A. Regev. A cancer cell program promotes t cell exclusion and resistance to checkpoint blockade. *Cell*, 175(4):984–997 e24, 2018. Jerby-Arnon, Livnat Shah, Parin Cuoco, Michael S Rodman, Christopher Su, Mei-Ju Melms, Johannes C Leeson, Rachel Kanodia, Abhay Mei, Shaolin Lin, Jia-Ren Wang, Shu Rabasha, Bokang Liu, David Zhang, Gao Margolais, Claire Ashenberg, Orr Ott, Patrick A

Buchbinder, Elizabeth I Haq, Rizwan Hodi, F Stephen Boland, Genevieve M Sullivan, Ryan J Frederick, Dennie T Miao, Benchun Moll, Tabea Flaherty, Keith T Herlyn, Meenhard Jenkins, Russell W Thummalapalli, Rohit Kowalczyk, Monika S Canadas, Israel Schilling, Bastian Cartwright, Adam N R Luoma, Adrienne M Malu, Shruti Hwu, Patrick Bernatchez, Chantale Forget, Marie-Andree Barbie, David A Shalek, Alex K Tirosh, Itay Sorger, Peter K Wucherpennig, Kai Van Allen, Eliezer M Schadendorf, Dirk Johnson, Bruce E Rotem, Asaf Rozenblatt-Rosen, Orit Garraway, Levi A Yoon, Charles H Izar, Benjamin Regev, Aviv eng U24 CA180922/CA/NCI NIH HHS/ R01 CA173750/CA/NCI NIH HHS/ P30 CA010815/CA/NCI NIH HHS/ K08 CA222663/CA/NCI NIH HHS/ HHMI/Howard Hughes Medical Institute/ P30 CA014051/CA/NCI NIH HHS/ R33 CA202820/CA/NCI NIH HHS/ T32 CA207021/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2018/11/06 Cell. 2018 Nov 1;175(4):984-997.e24. doi: 10.1016/j.cell.2018.09.006.

- [5] M. Sade-Feldman, K. Yizhak, S. L. Bjorgaard, J. P. Ray, C. G. de Boer, R. W. Jenkins, D. J. Lieb, J. H. Chen, D. T. Frederick, M. Barzily-Rokni, S. S. Freeman, A. Reuben, P. J. Hoover, A. C. Villani, E. Ivanova, A. Portell, P. H. Lizotte, A. R. Aref, J. P. Eliane, M. R. Hammond, H. Vitzthum, S. M. Blackmon, B. Li, V. Gopalakrishnan, S. M. Reddy, Z. A. Cooper, C. P. Paweletz, D. A. Barbie, A. Stemmer-Rachamimov, K. T. Flaherty, J. A. Wargo, G. M. Boland, R. J. Sullivan, G. Getz, and N. Hacohen. Defining t cell states associated with response to checkpoint immunotherapy in melanoma. *Cell*, 175(4):998–1013 e20, 2018. Sade-Feldman, Moshe Yizhak, Keren Bjorgaard, Stacey L Ray, John P de Boer, Carl G Jenkins, Russell W Lieb, David J Chen, Jonathan H Frederick, Dennie T Barzily-Rokni, Michal Freeman, Samuel S Reuben, Alexandre Hoover, Paul J Villani, Alexandra-Chloe Ivanova, Elena Portell, Andrew Lizotte, Patrick H Aref, Amir R Eliane, Jean-Pierre Hammond, Marc R Vitzthum, Hans Blackmon, Shauna M Li, Bo Gopalakrishnan, Vancheswaran Reddy, Sangeetha M Cooper, Zachary A Paweletz, Cloud P Barbie, David A Stemmer-Rachamimov, Anat Flaherty, Keith T Wargo, Jennifer A Boland, Genevieve M Sullivan, Ryan J Getz, Gad Hacohen, Nir eng U54 HG003067/HG/NHGRI NIH HHS/ S10 RR023440/RR/NCRR NIH HHS/ S10 RR020936/RR/NCRR NIH HHS/ U54 CA224068/CA/NCI NIH HHS/ S10 OD016372/OD/NIH HHS/ R01 CA208756/CA/NCI NIH HHS/ T32 CA207021/CA/NCI NIH HHS/ S10 OD012027/OD/NIH HHS/ R01 CA190394/CA/NCI NIH HHS/ F32 AI129249/AI/NIAID NIH HHS/ U01 CA214381/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2018/11/06 Cell. 2018 Nov 1;175(4):998-1013.e20. doi: 10.1016/j.cell.2018.10.038.
- [6] J. Hu, L. Zhang, H. Xia, Y. Yan, X. Zhu, F. Sun, L. Sun, S. Li, D. Li, J. Wang, Y. Han, J. Zhang, D. Bian, H. Yu, Y. Chen, P. Fan, Q. Ma, G. Jiang, C. Wang, and P. Zhang. Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell rna sequencing. *Genome Med*, 15(1):14, 2023. Hu, Junjie Zhang, Lele Xia, Haoran Yan, Yilv Zhu, Xinsheng Sun, Fenghuan Sun, Liangdong Li, Shuangyi Li, Dianke Wang, Jin Han, Ya Zhang, Jing Bian, Dongliang Yu, Huansha Chen, Yan Fan, Pengyu Ma, Qiang Jiang, Gening Wang, Chenfei Zhang, Peng eng Research Support, Non-U.S. Gov't England 2023/03/04 Genome Med. 2023 Mar 3;15(1):14. doi: 10.1186/s13073-023-01164-9.

- [7] D. Lambrechts, E. Wauters, B. Boeckx, S. Aibar, D. Nittner, O. Burton, A. Bassez, H. Decaluwe, A. Pircher, K. Van den Eynde, B. Weynand, E. Verbeken, P. De Leyn, A. Liston, J. Vansteenkiste, P. Carmeliet, S. Aerts, and B. Thienpont. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*, 24(8):1277–1289, 2018. Lambrechts, Diether Wauters, Els Boeckx, Bram Aibar, Sara Nittner, David Burton, Oliver Bassez, Ayse Decaluwe, Herbert Pircher, Andreas Van den Eynde, Kathleen Weynand, Birgit Verbeken, Erik De Leyn, Paul Liston, Adrian Vansteenkiste, Johan Carmeliet, Peter Aerts, Stein Thienpont, Bernard eng Research Support, Non-U.S. Gov't 2018/07/11 Nat Med. 2018 Aug;24(8):1277-1289. doi: 10.1038/s41591-018-0096-5. Epub 2018 Jul 9.
- [8] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000. Hanahan, D Weinberg, R A eng Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Review 2000/01/27 Cell. 2000 Jan 7;100(1):57-70. doi: 10.1016/s0092-8674(00)81683-9.
- [9] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–74, 2011. Hanahan, Douglas Weinberg, Robert A eng 09-0228/AICR\_/Worldwide Cancer Research/United Kingdom Research Support, N.I.H., Extramural Review 2011/03/08 Cell. 2011 Mar 4;144(5):646-74. doi: 10.1016/j.cell.2011.02.013.
- [10] Centers for Disease Control and Prevention. The top 10 causes of death. 2021.
- [11] World Health Organization, 2021.
- [12] World Health Organization. Cancer. 2022.
- [13] National Cancer Institute. Nci budget and appropriations, 2024.
- [14] C. Santucci, G. Carioli, P. Bertuccio, M. Malvezzi, U. Pastorino, P. Boffetta, E. Negri, C. Bosetti, and C. La Vecchia. Progress in cancer mortality, incidence, and survival: a global overview. *Eur J Cancer Prev*, 29(5):367–381, 2020. Santucci, Claudia Carioli, Greta Bertuccio, Paola Malvezzi, Matteo Pastorino, Ugo Boffetta, Paolo Negri, Eva Bosetti, Cristina La Vecchia, Carlo eng Research Support, Non-U.S. Gov't Review England 2020/08/03 Eur J Cancer Prev. 2020 Sep;29(5):367-381. doi: 10.1097/CEJ.0000000000000594.
- [15] C. Demoor-Goldschmidt and F. de Vathaire. Review of risk factors of secondary cancers among cancer survivors. *Br J Radiol*, 92(1093):20180390, 2019. Demoor-Goldschmidt, Charlotte de Vathaire, Florent eng Review England 2018/08/14 Br J Radiol. 2019 Jan;92(1093):20180390. doi: 10.1259/bjr.20180390. Epub 2018 Sep 12.
- [16] M. Cusnir and L. Cavalcante. Inter-tumor heterogeneity. *Hum Vaccin Immunother*, 8(8):1143–5, 2012. Cusnir, Mike Cavalcante, Ludmila eng Review 2012/08/03 Hum Vaccin Immunother. 2012 Aug;8(8):1143-5. doi: 10.4161/hv.21203. Epub 2012 Aug 1.
- [17] H. Easwaran, H. C. Tsai, and S. B. Baylin. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell*, 54(5):716–27, 2014. Easwaran, Hariharan Tsai, Hsing-Chen Baylin, Stephen B eng P30 CA006973/CA/NCI NIH HHS/ HL099775/HL/NHLBI NIH HHS/ U01 HL099775/HL/NHLBI NIH HHS/ ES011858/ES/NIEHS NIH HHS/ R01 ES011858/ES/NIEHS NIH HHS/ R01 CA043318/CA/NCI NIH

- HHS/ CA043318/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review 2014/06/07 Mol Cell. 2014 Jun 5;54(5):716-27. doi: 10.1016/j.molcel.2014.05.015.
- [18] C. Kim, R. Gao, E. Sei, R. Brandt, J. Hartman, T. Hatschek, N. Crosetto, T. Foukakis, and N. E. Navin. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell*, 173(4):879–893 e13, 2018. Kim, Charissa Gao, Ruli Sei, Emi Brandt, Rachel Hartman, Johan Hatschek, Thomas Crosetto, Nicola Foukakis, Theodoros Navin, Nicholas E eng P30 CA016672/CA/NCI NIH HHS/ R01 CA169244/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2018/04/24 Cell. 2018 May 3;173(4):879-893.e13. doi: 10.1016/j.cell.2018.03.041. Epub 2018 Apr 19.
- [19] P. C. Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–8, 1976. Nowell, P C eng Research Support, U.S. Gov't, P.H.S. 1976/10/01 Science. 1976 Oct 1;194(4260):23-8. doi: 10.1126/science.959840.
- [20] D. M. Spagnolo, R. Gyanchandani, Y. Al-Kofahi, A. M. Stern, T. R. Lezon, A. Gough, D. E. Meyer, F. Ginty, B. Sarachan, J. Fine, A. V. Lee, D. L. Taylor, and S. C. Chennubhotla. Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers. *J Pathol Inform*, 7:47, 2016. Spagnolo, Daniel M Gyanchandani, Rekha Al-Kofahi, Yousef Stern, Andrew M Lezon, Timothy R Gough, Albert Meyer, Dan E Ginty, Fiona Sarachan, Brion Fine, Jeffrey Lee, Adrian V Taylor, D Lansing Chennubhotla, S Chakra eng P30 CA047904/CA/NCI NIH HHS/ T32 EB009403/EB/NIBIB NIH HHS/ U01 CA204826/CA/NCI NIH HHS/ U54 HG008540/HG/NHGRI NIH HHS/ 2016/12/21 J Pathol Inform. 2016 Nov 29;7:47. doi: 10.4103/2153-3539.194839. eCollection 2016.
- [21] J. K. Rhee, Y. C. Jung, K. R. Kim, J. Yoo, J. Kim, Y. J. Lee, Y. H. Ko, H. H. Lee, B. C. Cho, and T. M. Kim. Impact of tumor purity on immune gene expression and clustering analyses across multiple cancer types. *Cancer Immunol Res*, 6(1):87–97, 2018. Rhee, Je-Keun Jung, Yu Chae Kim, Kyu Ryung Yoo, Jinseon Kim, Jeeyoon Lee, Yong-Jae Ko, Yoon Ho Lee, Han Hong Cho, Byoung Chul Kim, Tae-Min eng Research Support, Non-U.S. Gov't 2017/11/17 Cancer Immunol Res. 2018 Jan;6(1):87-97. doi: 10.1158/2326-6066.CIR-17-0201. Epub 2017 Nov 15.
- [22] A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, and A. A. Alizadeh. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*, 37(7):773–782, 2019. Newman, Aaron M Steen, Chloe B Liu, Chih Long Gentles, Andrew J Chaudhuri, Aadel A Scherer, Florian Khodadoust, Michael S Esfahani, Mohammad S Luca, Bogdan A Steiner, David Diehn, Maximilian Alizadeh, Ash A eng R00 CA187192/CA/NCI NIH HHS/ R25 CA180993/CA/NCI NIH HHS/ R01 CA229766/CA/NCI NIH HHS/ U01 CA194389/CA/NCI NIH HHS/ U01 CA154969/CA/NCI NIH HHS/ P30 CA124435/CA/NCI NIH HHS/ R01 CA188298/CA/NCI NIH HHS/ UL1 TR003142/TR/NCATS NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2019/05/08 Nat Biotechnol. 2019 Jul;37(7):773-782. doi: 10.1038/s41587-019-0114-2. Epub 2019 May 6.



- [23] S. M. Prakadan, A. K. Shalek, and D. A. Weitz. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nat Rev Genet*, 18(6):345–361, 2017. Prakadan, Sanjay M Shalek, Alex K Weitz, David A eng P01 GM096971/GM/NIGMS NIH HHS/ DP2 GM119419/GM/NIGMS NIH HHS/ P01 HL120839/HL/NHLBI NIH HHS/ P50 HG006193/HG/NHGRI NIH HHS/ R01 EB014703/EB/NIBIB NIH HHS/ U24 AI118672/AI/NIAID NIH HHS/ RM1 HG006193/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review England 2017/04/11 Nat Rev Genet. 2017 Jun;18(6):345-361. doi: 10.1038/nrg.2017.15. Epub 2017 Apr 10.
- [24] A. Rotem, O. Ram, N. Shores, R. A. Sperling, A. Goren, D. A. Weitz, and B. E. Bernstein. Single-cell chip-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol*, 33(11):1165–72, 2015. Rotem, Assaf Ram, Oren Shores, Noam Sperling, Ralph A Goren, Alon Weitz, David A Bernstein, Bradley E eng U54 HG006991/HG/NHGRI NIH HHS/ U01 HL100395/HL/NHLBI NIH HHS/ U54HG006991/HG/NHGRI NIH HHS/ P50 HG006193/HG/NHGRI NIH HHS/ Howard Hughes Medical Institute/ P50HG006193/HG/NHGRI NIH HHS/ U01HL100395/HL/NHLBI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2015/10/13 Nat Biotechnol. 2015 Nov;33(11):1165-72. doi: 10.1038/nbt.3383. Epub 2015 Oct 12.
- [25] P. V. Kharchenko. The triumphs and limitations of computational methods for scRNA-seq. *Nat Methods*, 18(7):723–732, 2021. Kharchenko, Peter V eng R01 HL131768/HL/NHLBI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review 2021/06/23 Nat Methods. 2021 Jul;18(7):723-732. doi: 10.1038/s41592-021-01171-x. Epub 2021 Jun 21.
- [26] G. Carangelo, A. Magi, and R. Semeraro. From multitude to singularity: An up-to-date overview of scRNA-seq data generation and analysis. *Front Genet*, 13:994069, 2022. Carangelo, Giulia Magi, Alberto Semeraro, Roberto eng Review Switzerland 2022/10/21 Front Genet. 2022 Oct 3;13:994069. doi: 10.3389/fgene.2022.994069. eCollection 2022.
- [27] D. Jovic, X. Liang, H. Zeng, L. Lin, F. Xu, and Y. Luo. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med*, 12(3):e694, 2022. Jovic, Dragomirka Liang, Xue Zeng, Hua Lin, Lin Xu, Fengping Luo, Yonglun eng Research Support, Non-U.S. Gov't Review 2022/03/31 Clin Transl Med. 2022 Mar;12(3):e694. doi: 10.1002/ctm2.694.
- [28] National Cancer Institute. Human tumor atlas network, 2024.
- [29] Human Cell Atlas. Human cell atlas.
- [30] H. Y. Chuang, M. Hofree, and T. Ideker. A decade of systems biology. *Annu Rev Cell Dev Biol*, 26:721–44, 2010. Chuang, Han-Yu Hofree, Matan Ideker, Trey eng P41 RR031228/RR/NCRR NIH HHS/ Review 2010/07/08 Annu Rev Cell Dev Biol. 2010;26:721-44. doi: 10.1146/annurev-cellbio-100109-104122.
- [31] P. T. Ram, J. Mendelsohn, and G. B. Mills. Bioinformatics and systems biology. *Mol Oncol*, 6(2):147–54, 2012. Ram, Prahlad T Mendelsohn, John Mills, Gordon B eng R01 CA125109-03/CA/NCI NIH HHS/ P30

CA016672/CA/NCI NIH HHS/ U54 CA112970-08/CA/NCI NIH HHS/ P50 CA098258/CA/NCI NIH HHS/ U54 CA112970/CA/NCI NIH HHS/ R01 CA125109/CA/NCI NIH HHS/ Review 2012/03/02 Mol Oncol. 2012 Apr;6(2):147-54. doi: 10.1016/j.molonc.2012.01.008. Epub 2012 Feb 17.

- [32] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res*, 33(Database issue):D428–32, 2005. Joshi-Tope, G Gillespie, M Vastrik, I D'Eustachio, P Schmidt, E de Bono, B Jassal, B Gopinath, G R Wu, G R Matthews, L Lewis, S Birney, E Stein, L eng R01 HG002639/HG/NHGRI NIH HHS/ U41 HG003751/HG/NHGRI NIH HHS/ Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. England 2004/12/21 Nucleic Acids Res. 2005 Jan 1;33(Database issue):D428-32. doi: 10.1093/nar/gki072.
- [33] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. The reactome pathway knowledgebase. *Nucleic Acids Res*, 46(D1):D649–D655, 2018. Fabregat, Antonio Jupe, Steven Matthews, Lisa Sidiropoulos, Konstantinos Gillespie, Marc Garapati, Phani Haw, Robin Jassal, Bijay Korninger, Florian May, Bruce Milacic, Marija Roca, Corina Duenas Rothfels, Karen Sevilla, Cristoffer Shamovsky, Veronica Shorser, Solomon Varusai, Thawfeek Viteri, Guilherme Weiser, Joel Wu, Guanming Stein, Lincoln Hermjakob, Henning D'Eustachio, Peter eng U41 HG003751/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2017/11/18 Nucleic Acids Res. 2018 Jan 4;46(D1):D649-D655. doi: 10.1093/nar/gkx1132.
- [34] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. The reactome pathway knowledgebase. *Nucleic Acids Res*, 48(D1):D498–D503, 2020. Jassal, Bijay Matthews, Lisa Viteri, Guilherme Gong, Chuqiao Lorente, Pascual Fabregat, Antonio Sidiropoulos, Konstantinos Cook, Justin Gillespie, Marc Haw, Robin Loney, Fred May, Bruce Milacic, Marija Rothfels, Karen Sevilla, Cristoffer Shamovsky, Veronica Shorser, Solomon Varusai, Thawfeek Weiser, Joel Wu, Guanming Stein, Lincoln Hermjakob, Henning D'Eustachio, Peter eng P41 HG003751/HG/NHGRI NIH HHS/ U41 HG003751/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2019/11/07 Nucleic Acids Res. 2020 Jan 8;48(D1):D498-D503. doi: 10.1093/nar/gkz1031.
- [35] M. Gillespie, B. Jassal, R. Stephan, M. Milacic, K. Rothfels, A. Senff-Ribeiro, J. Griss, C. Sevilla, L. Matthews, C. Gong, C. Deng, T. Varusai, E. Ragueneau, Y. Haider, B. May, V. Shamovsky, J. Weiser, T. Brunson, N. Sanati, L. Beckman, X. Shao, A. Fabregat, K. Sidiropoulos, J. Murillo, G. Viteri, J. Cook, S. Shorser, G. Bader, E. Demir, C. Sander, R. Haw, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. The reactome pathway knowledgebase 2022. *Nucleic Acids Res*, 50(D1):D687–D692, 2022. Gillespie, Marc Jassal, Bijay Stephan, Ralf Milacic, Marija Rothfels, Karen Senff-Ribeiro, Andrea Griss, Johannes Sevilla, Cristoffer Matthews, Lisa Gong,

Chunqiao Deng, Chuan Varusai, Thawfeek Ragueneau, Eliot Haider, Yusra May, Bruce Shamovsky, Veronica Weiser, Joel Brunson, Timothy Sanati, Nasim Beckman, Liam Shao, Xiang Fabregat, Antonio Sidiropoulos, Konstantinos Murillo, Julieth Viteri, Guilherme Cook, Justin Shorser, Solomon Bader, Gary Demir, Emek Sander, Chris Haw, Robin Wu, Guanming Stein, Lincoln Hermjakob, Henning D'Eustachio, Peter eng U41HG003751/HG/NHGRI NIH HHS/ U24 HG012198/HG/NHGRI NIH HHS/ U54 GM114833/GM/NIGMS NIH HHS/ U01 CA239069/CA/NCI NIH HHS/ U41 HG003751/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2021/11/18 Nucleic Acids Res. 2022 Jan 7;50(D1):D687-D692. doi: 10.1093/nar/gkab1028.

- [36] Encode Project Consortium. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57–74, 2012. eng R01HG003143/HG/NHGRI NIH HHS/ RC2 HG005573/HG/NHGRI NIH HHS/ U54HG004592/HG/NHGRI NIH HHS/ RC2 HG005679/HG/NHGRI NIH HHS/ R01 HG003143/HG/NHGRI NIH HHS/ ZIA HG200341/ImNIH/Intramural NIH HHS/ U41HG004568/HG/NHGRI NIH HHS/ U54 HG004557/HG/NHGRI NIH HHS/ U01HG004695/HG/NHGRI NIH HHS/ U54 HG004570/HG/NHGRI NIH HHS/ R37 DK044746/DK/NIDDK NIH HHS/ U54 HG004576/HG/NHGRI NIH HHS/ R01HG003988/HG/NHGRI NIH HHS/ U54 HG004592/HG/NHGRI NIH HHS/ U54HG004576/HG/NHGRI NIH HHS/ R01 HG003988/HG/NHGRI NIH HHS/ R01 HG003541/HG/NHGRI NIH HHS/ R01HG004456-03/HG/NHGRI NIH HHS/ R01HG003541/HG/NHGRI NIH HHS/ U54HG004558/HG/NHGRI NIH HHS/ RC2HG005591/HG/NHGRI NIH HHS/ RC2 HG005591/HG/NHGRI NIH HHS/ K99 HG006698/HG/NHGRI NIH HHS/ U54 HG004558/HG/NHGRI NIH HHS/ P30 CA016086/CA/NCI NIH HHS/ R01 DK054369/DK/NIDDK NIH HHS/ ZIA HG200323/ImNIH/Intramural NIH HHS/ U54HG004557/HG/NHGRI NIH HHS/ U54 HG004563/HG/NHGRI NIH HHS/ U01HG004561/HG/NHGRI NIH HHS/ U54HG004563/HG/NHGRI NIH HHS/ R01 HG005085/HG/NHGRI NIH HHS/ U01 HG004571/HG/NHGRI NIH HHS/ U41 HG004568/HG/NHGRI NIH HHS/ R01HG003700/HG/NHGRI NIH HHS/ T32 GM007205/GM/NIGMS NIH HHS/ 095908/WT\_/Wellcome Trust/United Kingdom R01 DK065806/DK/NIDDK NIH HHS/ P30 CA045508/CA/NCI NIH HHS/ RC2HG005679/HG/NHGRI NIH HHS/ WT\_/Wellcome Trust/United Kingdom U54HG004570/HG/NHGRI NIH HHS/ R01 HG003700/HG/NHGRI NIH HHS/ U01 HG004695/HG/NHGRI NIH HHS/ U54HG004555/HG/NHGRI NIH HHS/ U01 HG004561/HG/NHGRI NIH HHS/ U01HG004571/HG/NHGRI NIH HHS/ U54 HG004555/HG/NHGRI NIH HHS/ R01 HG004456/HG/NHGRI NIH HHS/ Research Support, American Recovery and Reinvestment Act Research Support, N.I.H., Extramural Research Support, N.I.H., Intramural Research Support, U.S. Gov't, Non-P.H.S. England 2012/09/08 Nature. 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247.

- [37] M. Wanzel, S. Herold, and M. Eilers. Transcriptional repression by myc. *Trends Cell Biol*, 13(3):146–50, 2003. Wanzel, Michael Herold, Steffi Eilers, Martin eng Review England 2003/03/12 Trends Cell Biol. 2003 Mar;13(3):146-50. doi: 10.1016/s0962-8924(03)00003-5.

- [38] P. J. Park. Chip-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 10(10):669–80, 2009. Park, Peter J eng R01GM082798/GM/NIGMS NIH HHS/ RL1DE019021/DE/NIDCR NIH HHS/ R01 GM082798/GM/NIGMS NIH HHS/ RL1 DE019021/DE/NIDCR NIH HHS/ U01HG004258/HG/NHGRI NIH HHS/ U01 HG004258/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Review England 2009/09/09 *Nat Rev Genet*. 2009 Oct;10(10):669-80. doi: 10.1038/nrg2641. Epub 2009 Sep 8.
- [39] E. H. Davidson. Emerging properties of animal gene regulatory networks. *Nature*, 468(7326):911–20, 2010. Davidson, Eric H eng P01 HD037105/HD/NICHD NIH HHS/ R01 GM061005/GM/NIGMS NIH HHS/ HD-37105/HD/NICHD NIH HHS/ GM-61005/GM/NIGMS NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review England 2010/12/18 *Nature*. 2010 Dec 16;468(7326):911-20. doi: 10.1038/nature09645.
- [40] I. S. Peter and E. H. Davidson. Evolution of gene regulatory networks controlling body plan development. *Cell*, 144(6):970–85, 2011. Peter, Isabelle S Davidson, Eric H eng P01 HD037105/HD/NICHD NIH HHS/ P01 HD037105-12/HD/NICHD NIH HHS/ HD-037105/HD/NICHD NIH HHS/ Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. Review 2011/03/19 *Cell*. 2011 Mar 18;144(6):970-85. doi: 10.1016/j.cell.2011.02.017.
- [41] S. A. Assi, M. R. Imperato, D. J. L. Coleman, A. Pickin, S. Potluri, A. Ptasinska, P. S. Chin, H. Blair, P. Cauchy, S. R. James, J. Zacarias-Cabeza, L. N. Gilding, A. Beggs, S. Clokie, J. C. Loke, P. Jenkin, A. Uddin, R. Delwel, S. J. Richards, M. Raghavan, M. J. Griffiths, O. Heidenreich, P. N. Cockerill, and C. Bonifer. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet*, 51(1):151–162, 2019. Assi, Salam A Imperato, Maria Rosaria Coleman, Daniel J L Pickin, Anna Potluri, Sandeep Ptasinska, Anetta Chin, Paulynn Suyin Blair, Helen Cauchy, Pierre James, Sally R Zacarias-Cabeza, Joaquin Gilding, L Niall Beggs, Andrew Clokie, Sam Loke, Justin C Jenkin, Phil Uddin, Ash Delwel, Ruud Richards, Stephen J Raghavan, Manoj Griffiths, Michael J Heidenreich, Olaf Cockerill, Peter N Bonifer, Constanze eng 23389/CRUK\_/Cancer Research UK/United Kingdom MR/M009157/1/MRC\_/Medical Research Council/United Kingdom MR/P019609/1/MRC\_/Medical Research Council/United Kingdom Research Support, Non-U.S. Gov't 2018/11/14 *Nat Genet*. 2019 Jan;51(1):151-162. doi: 10.1038/s41588-018-0270-1. Epub 2018 Nov 12.
- [42] P. Han, C. Gopalakrishnan, H. Yu, and E. Wang. Gene regulatory network rewiring in the immune cells associated with cancer. *Genes (Basel)*, 8(11), 2017. Han, Pengyong Gopalakrishnan, Chandrasekhar Yu, Haiquan Wang, Edwin eng Switzerland 2017/11/08 *Genes (Basel)*. 2017 Nov 7;8(11):308. doi: 10.3390/genes8110308.
- [43] D. Lahnemann, J. Koster, E. Szczurek, D. J. McCarthy, S. C. Hicks, M. D. Robinson, C. A. Vallejos, K. R. Campbell, N. Beerenwinkel, A. Mahfouz, L. Pinello, P. Skums, A. Stamatakis, C. S. Attolini, S. Aparicio, J. Baaijens, M. Balvert, B. Barbanson, A. Cappuccio, G. Corleone, B. E. Dutilh, M. Florescu, V. Guryev, R. Holmer, K. Jahn, T. J. Lobo, E. M. Keizer, I. Khatri, S. M. Kielbasa, J. O. Korbel, A. M. Kozlov, T. H. Kuo, B. P. F. Lelieveldt, II Mandoiu, J. C. Marioni, T. Marschall, F. Molder, A. Niknejad, A. Raczowska,

- M. Reinders, J. Ridder, A. E. Saliba, A. Somarakis, O. Stegle, F. J. Theis, H. Yang, A. Zelikovsky, A. C. McHardy, B. J. Raphael, S. P. Shah, and A. Schonhuth. Eleven grand challenges in single-cell data science. *Genome Biol*, 21(1):31, 2020. Lahnemann, David Koster, Johannes Szczurek, Ewa McCarthy, Davis J Hicks, Stephanie C Robinson, Mark D Vallejos, Catalina A Campbell, Kieran R Beerenwinkel, Niko Mahfouz, Ahmed Pinello, Luca Skums, Pavel Stamatakis, Alexandros Attolini, Camille Stephan-Otto Aparicio, Samuel Baaijens, Jasmijn Balvert, Marleen Barbanson, Buys de Cappuccio, Antonio Corleone, Giacomo Dutilh, Bas E Florescu, Maria Guryev, Victor Holmer, Rens Jahn, Katharina Lobo, Thamar Jessurun Keizer, Emma M Khatri, Indu Kielbasa, Szymon M Korbek, Jan O Kozlov, Alexey M Kuo, Tzu-Hao Lelieveldt, Boudewijn P F Mandoiu, Ion I Marioni, John C Marschall, Tobias Molder, Felix Niknejad, Amir Raczkowska, Alicja Reinders, Marcel Ridder, Jeroen de Saliba, Antoine-Emmanuel Somarakis, Antonios Stegle, Oliver Theis, Fabian J Yang, Huan Zelikovsky, Alex McHardy, Alice C Raphael, Benjamin J Shah, Sohrab P Schonhuth, Alexander eng R35 HG010717/HG/NHGRI NIH HHS/ R01 HG007069/HG/NHGRI NIH HHS/ P30 CA008748/CA/NCI NIH HHS/ R00 HG009007/HG/NHGRI NIH HHS/ 22231/CRUK\_/Cancer Research UK/United Kingdom C9545/A29580/CRUK\_/Cancer Research UK/United Kingdom R01 EB025022/EB/NIBIB NIH HHS/ R00 HG008399/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review England 2020/02/09 *Genome Biol*. 2020 Feb 7;21(1):31. doi: 10.1186/s13059-020-1926-6.
- [44] G. Chen, B. Ning, and T. Shi. Single-cell rna-seq technologies and related computational data analysis. *Front Genet*, 10:317, 2019. Chen, Geng Ning, Baitang Shi, Tielu eng Review Switzerland 2019/04/27 *Front Genet*. 2019 Apr 5;10:317. doi: 10.3389/fgene.2019.00317. eCollection 2019.
- [45] A. Pratapa, A. P. Jalihal, J. N. Law, A. Bharadwaj, and T. M. Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*, 17(2):147–154, 2020. Pratapa, Aditya Jalihal, Amogh P Law, Jeffrey N Bharadwaj, Aditya Murali, T M eng UH2 CA203768/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. 2020/01/08 *Nat Methods*. 2020 Feb;17(2):147-154. doi: 10.1038/s41592-019-0690-6. Epub 2020 Jan 6.
- [46] S. Chen and J. C. Mar. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*, 19(1):232, 2018. Chen, Shuonan Mar, Jessica C eng Evaluation Study Research Support, Non-U.S. Gov't England 2018/06/20 *BMC Bioinformatics*. 2018 Jun 19;19(1):232. doi: 10.1186/s12859-018-2217-z.
- [47] S. Aibar, C. B. Gonzalez-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, and S. Aerts. Scenic: single-cell regulatory network inference and clustering. *Nat Methods*, 14(11):1083–1086, 2017. Aibar, Sara Gonzalez-Blas, Carmen Bravo Moerman, Thomas Huynh-Thu, Van Anh Imrichova, Hana Hulselmans, Gert Rambow, Florian Marine, Jean-Christophe Geurts, Pierre Aerts, Jan van den Oord, Joost Atak, Zeynep Kalender Wouters, Jasper Aerts,

Stein eng 724226/ERC\_/European Research Council/International 2017/10/11 Nat Methods. 2017 Nov;14(11):1083-1086. doi: 10.1038/nmeth.4463. Epub 2017 Oct 9.

- [48] T. E. Chan, M. P. H. Stumpf, and A. C. Babbie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst*, 5(3):251–267 e3, 2017. Chan, Thalia E Stumpf, Michael P H Babbie, Ann C eng BB/N011597/1/Biotechnology and Biological Sciences Research Council/United Kingdom Research Support, Non-U.S. Gov't 2017/09/29 Cell Syst. 2017 Sep 27;5(3):251-267.e3. doi: 10.1016/j.cels.2017.08.014.
- [49] F. Edfors, F. Danielsson, B. M. Hallstrom, L. Kall, E. Lundberg, F. Ponten, B. Forsstrom, and M. Uhlen. Gene-specific correlation of rna and protein levels in human cells and tissues. *Mol Syst Biol*, 12(10):883, 2016. Edfors, Fredrik Danielsson, Frida Hallstrom, Bjorn M Kall, Lukas Lundberg, Emma Ponten, Fredrik Forsstrom, Bjorn Uhlen, Mathias eng Comparative Study England 2016/12/13 Mol Syst Biol. 2016 Oct 20;12(10):883. doi: 10.15252/msb.20167144.
- [50] D. Hanahan. Hallmarks of cancer: New dimensions. *Cancer Discov*, 12(1):31–46, 2022. Hanahan, Douglas eng Review 2022/01/14 Cancer Discov. 2022 Jan;12(1):31-46. doi: 10.1158/2159-8290.CD-21-1059.
- [51] L. Gibellini, S. De Biasi, C. Porta, D. Lo Tartaro, R. Depenni, G. Pellacani, R. Sabbatini, and A. Cossarizza. Single-cell approaches to profile the response to immune checkpoint inhibitors. *Front Immunol*, 11:490, 2020. Gibellini, Lara De Biasi, Sara Porta, Camillo Lo Tartaro, Domenico Depenni, Roberta Pellacani, Giovanni Sabbatini, Roberto Cossarizza, Andrea eng Research Support, Non-U.S. Gov't Review Switzerland 2020/04/09 Front Immunol. 2020 Mar 20;11:490. doi: 10.3389/fimmu.2020.00490. eCollection 2020.
- [52] Samhita Rao, Karim Gharib, and Arnold Han. Chapter five - cancer immunosurveillance by t cells. In Lorenzo Galluzzi and Nils-Petter Rudqvist, editors, *Biology of T Cells - Part B*, volume 342 of *International Review of Cell and Molecular Biology*, pages 149–173. Academic Press, 2019.
- [53] S.J. Oiseth and Mohamed S. Aziz. Cancer immunotherapy: a brief review of the history, possibilities, and challenges ahead. *J Cancer Metastasis Treat*, 3(10):250–261, 2017.
- [54] C. H. June and M. Sadelain. Chimeric antigen receptor therapy. *N Engl J Med*, 379(1):64–73, 2018. June, Carl H Sadelain, Michel eng P01 CA214278/CA/NCI NIH HHS/ Review 2018/07/05 N Engl J Med. 2018 Jul 5;379(1):64-73. doi: 10.1056/NEJMra1706169.
- [55] J. L. Tanyi, S. Bobisse, E. Ophir, S. Tuyaeerts, A. Roberti, R. Genolet, P. Baumgartner, B. J. Stevenson, C. Iseli, D. Dangaj, B. Czerniecki, A. Semilietof, J. Racle, A. Michel, I. Xenarios, C. Chiang, D. S. Monos, D. A. Torigian, H. L. Nisenbaum, O. Michielin, C. H. June, B. L. Levine, Jr. Powell, D. J., D. Gfeller, R. Mick, U. Dafni, V. Zoete, A. Harari, G. Coukos, and L. E. Kandalaft. Personalized cancer vaccine effectively mobilizes antitumor t cell immunity in ovarian cancer. *Sci Transl Med*, 10(436), 2018. Tanyi, Janos L Bobisse, Sara Ophir, Eran Tuyaeerts, Sandra Roberti, Annalisa Genolet, Raphael Baumgartner, Petra Stevenson, Brian J Iseli, Christian Dangaj, Denarda Czerniecki, Brian Semilietof, Aikaterini Racle, Julien Michel, Alexandra Xenarios, Ioannis

- Chiang, Cheryl Monos, Dimitri S Torigian, Drew A Nisenbaum, Harvey L Michielin, Olivier June, Carl H Levine, Bruce L Powell, Daniel J Jr Gfeller, David Mick, Rosemarie Dafni, Urania Zoete, Vincent Harari, Alexandre Coukos, George Kandalaft, Lana E eng P50 CA083638/NCI NIH HHS/National Cancer Institute/ R21 CA156224/NCI NIH HHS/National Cancer Institute/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2018/04/13 Sci Transl Med. 2018 Apr 11;10(436):eaa05931. doi: 10.1126/scitranslmed.aao5931.
- [56] A. Haslam and V. Prasad. Estimation of the percentage of us patients with cancer who are eligible for and respond to checkpoint inhibitor immunotherapy drugs. *JAMA Netw Open*, 2(5):e192535, 2019. Haslam, Alyson Prasad, Vinay eng 2019/05/06 JAMA Netw Open. 2019 May 3;2(5):e192535. doi: 10.1001/jamanetworkopen.2019.2535.
- [57] D. M. Pardoll. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*, 12(4):252–64, 2012. Pardoll, Drew M eng P30 CA006973/CA/NCI NIH HHS/ P30 CA021765/CA/NCI NIH HHS/ Research Support, Non-U.S. Gov't Review England 2012/03/23 Nat Rev Cancer. 2012 Mar 22;12(4):252-64. doi: 10.1038/nrc3239.
- [58] S. L. Topalian, C. G. Drake, and D. M. Pardoll. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell*, 27(4):450–61, 2015. Topalian, Suzanne L Drake, Charles G Pardoll, Drew M eng R01 CA142779/CA/NCI NIH HHS/ R01 CA154555/CA/NCI NIH HHS/ R01CA142779/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review 2015/04/11 Cancer Cell. 2015 Apr 13;27(4):450-61. doi: 10.1016/j.ccell.2015.03.001. Epub 2015 Apr 6.
- [59] S. L. Topalian, J. M. Taube, R. A. Anders, and D. M. Pardoll. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer*, 16(5):275–87, 2016. Topalian, Suzanne L Taube, Janis M Anders, Robert A Pardoll, Drew M eng R01 CA142779/CA/NCI NIH HHS/ Review England 2016/04/16 Nat Rev Cancer. 2016 May;16(5):275-87. doi: 10.1038/nrc.2016.36. Epub 2016 Apr 15.
- [60] M. A. Postow, M. K. Callahan, and J. D. Wolchok. Immune checkpoint blockade in cancer therapy. *J Clin Oncol*, 33(17):1974–82, 2015. Postow, Michael A Callahan, Margaret K Wolchok, Jedd D eng P30 CA008748/CA/NCI NIH HHS/ Review 2015/01/22 J Clin Oncol. 2015 Jun 10;33(17):1974-82. doi: 10.1200/JCO.2014.59.4358. Epub 2015 Jan 20.
- [61] P. Sharma and J. P. Allison. The future of immune checkpoint therapy. *Science*, 348(6230):56–61, 2015. Sharma, Padmanee Allison, James P eng R01 CA163793/CA/NCI NIH HHS/ 1-R01 CA1633793-01/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review 2015/04/04 Science. 2015 Apr 3;348(6230):56-61. doi: 10.1126/science.aaa8172.
- [62] NCT04693234. Advantig-202: Anti-pd-1 monoclonal antibody tislelizumab (bgb-a317) combined with or without anti-tigit monoclonal antibody ociperlimab (bgb-a1217) in participants with previously treated recurrent or metastatic cervical cancer. Report NCT04693234, 2021.

- [63] NCT04732494. Advantig-203: Anti-pd-1 monoclonal antibody tislelizumab (bgb-a317) combined with or without anti-tigit monoclonal antibody ociperlimab (bgb-a1217) in participants with recurrent or metastatic esophageal squamous cell carcinoma. Report NCT04732494, 2021.
- [64] NCT01968109. An investigational immuno-therapy study to assess the safety, tolerability and effectiveness of anti-lag-3 with and without anti-pd-1 in the treatment of solid tumors. Report NCT01968109, 2013.
- [65] NCT04370704. Study of combination therapy with incmga00012 (anti-pd-1), incagn02385 (anti-lag-3), and incagn02390 (anti-tim-3) in participants with select advanced malignancies. Report NCT04370704, 2020.
- [66] NCT03005782. Study of regn3767 (anti-lag-3) with or without regn2810 (anti-pd1) in advanced cancers. Report NCT03005782, 2016.
- [67] NCT03961971. Trial of anti-tim-3 in combination with anti-pd-1 and srs in recurrent gbm. Report NCT03961971, 2020.
- [68] NCT03680508. Tsr-022 (anti-tim-3 antibody) and tsr-042 (anti-pd-1 antibody) in patients with liver cancer. Report NCT03680508, 2018.
- [69] L. Zhu, Y. Kong, J. Zhang, D. F. Claxton, W. C. Ehmann, W. B. Rybka, N. D. Palmisiano, M. Wang, B. Jia, M. Bayerl, T. D. Schell, R. J. Hohl, H. Zeng, and H. Zheng. Blimp-1 impairs t cell function via upregulation of tigit and pd-1 in patients with acute myeloid leukemia. *J Hematol Oncol*, 10(1):124, 2017. Zhu, Liuluan Kong, Yaxian Zhang, Jianhong Claxton, David F Ehmann, W Christopher Rybka, Witold B Palmisiano, Neil D Wang, Ming Jia, Bei Bayerl, Michael Schell, Todd D Hohl, Raymond J Zeng, Hui Zheng, Hong eng P01 CA171983/CA/NCI NIH HHS/ Research Support, Non-U.S. Gov't England 2017/06/21 *J Hematol Oncol*. 2017 Jun 19;10(1):124. doi: 10.1186/s13045-017-0486-z.
- [70] P. Guruprasad, Y. G. Lee, K. H. Kim, and M. Ruella. The current landscape of single-cell transcriptomics for cancer immunotherapy. *J Exp Med*, 218(1), 2021. Guruprasad, Puneeth Lee, Yong Gu Kim, Ki Hyun Ruella, Marco eng K99 CA212302/CA/NCI NIH HHS/ P01 CA214278/CA/NCI NIH HHS/ P01 CA217805/CA/NCI NIH HHS/ R00 CA212302/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review 2021/02/19 *J Exp Med*. 2021 Jan 4;218(1):e20201574. doi: 10.1084/jem.20201574.
- [71] C. U. Blank, W. N. Haining, W. Held, P. G. Hogan, A. Kallies, E. Lugli, R. C. Lynn, M. Philip, A. Rao, N. P. Restifo, A. Schietinger, T. N. Schumacher, P. L. Schwartzberg, A. H. Sharpe, D. E. Speiser, E. J. Wherry, B. A. Youngblood, and D. Zehn. Defining 't cell exhaustion'. *Nat Rev Immunol*, 19(11):665–674, 2019. Blank, Christian U Haining, W Nicholas Held, Werner Hogan, Patrick G Kallies, Axel Lugli, Enrico Lynn, Rachel C Philip, Mary Rao, Anjana Restifo, Nicholas P Schietinger, Andrea Schumacher, Ton N Schwartzberg, Pamela L Sharpe, Arlene H Speiser, Daniel E Wherry, E John Youngblood, Benjamin A Zehn, Dietmar eng R01 AI040127/AI/NIAID NIH HHS/ R56 AI109842/AI/NIAID NIH HHS/ U01 DE028227/DE/NIDCR NIH HHS/ R01 AI084167/AI/NIAID NIH HHS/ R01 AI109842/AI/NIAID NIH HHS/ Research Support, N.I.H.,



Extramural Research Support, Non-U.S. Gov't Review England 2019/10/02 Nat Rev Immunol. 2019 Nov;19(11):665-674. doi: 10.1038/s41577-019-0221-9. Epub 2019 Sep 30.

- [72] Jr. Aguiar, P. N., L. A. Perry, J. Penny-Dimri, H. Babiker, H. Tadokoro, R. A. de Mello, and Jr. Lopes, G. L. The effect of pd-11 testing on the cost-effectiveness and economic impact of immune checkpoint inhibitors for the second-line treatment of nsclc. *Ann Oncol*, 28(9):2256–2263, 2017. Aguiar, P N Jr Perry, L A Penny-Dimri, J Babiker, H Tadokoro, H de Mello, R A Lopes, G L Jr eng England 2017/06/22 Ann Oncol. 2017 Sep 1;28(9):2256-2263. doi: 10.1093/annonc/mdx305.
- [73] A. Shimabukuro-Vornhagen, P. Godel, M. Subklewe, H. J. Stemmler, H. A. Schlosser, M. Schlaak, M. Kochanek, B. Boll, and M. S. von Bergwelt-Baildon. Cytokine release syndrome. *J Immunother Cancer*, 6(1):56, 2018. Shimabukuro-Vornhagen, Alexander Godel, Philipp Subklewe, Marion Stemmler, Hans Joachim Schlosser, Hans Anton Schlaak, Max Kochanek, Matthias Boll, Boris von Bergwelt-Baildon, Michael S eng Review England 2018/06/17 J Immunother Cancer. 2018 Jun 15;6(1):56. doi: 10.1186/s40425-018-0343-9.
- [74] S. H. Tay, M. M. X. Toh, Y. L. Thian, B. A. Vellayappan, A. M. Fairhurst, Y. H. Chan, F. Aminkeng, L. D. Bharwani, Y. Huang, A. Mak, and A. S. C. Wong. Cytokine release syndrome in cancer patients receiving immune checkpoint inhibitors: A case series of 25 patients and review of the literature. *Front Immunol*, 13:807050, 2022. Tay, Sen Hee Toh, Michelle Min Xuan Thian, Yee Liang Vellayappan, Balamurugan A Fairhurst, Anna-Marie Chan, Yiong Huak Aminkeng, Folefac Bharwani, Lavina D Huang, Yiqing Mak, Anselm Wong, Alvin Seng Cheong eng Research Support, Non-U.S. Gov't Review Switzerland 2022/02/15 Front Immunol. 2022 Jan 28;13:807050. doi: 10.3389/fimmu.2022.807050. eCollection 2022.
- [75] E. J. Wherry. T cell exhaustion. *Nat Immunol*, 12(6):492–9, 2011. Wherry, E John eng AI071309/AI/NIAID NIH HHS/ AI082630/AI/NIAID NIH HHS/ AI083022/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review 2011/07/09 Nat Immunol. 2011 Jun;12(6):492-9. doi: 10.1038/ni.2035.
- [76] K. E. Pauken and E. J. Wherry. Overcoming t cell exhaustion in infection and cancer. *Trends Immunol*, 36(4):265–76, 2015. Pauken, Kristen E Wherry, E John eng AI112521/AI/NIAID NIH HHS/ AI082630/AI/NIAID NIH HHS/ U19 AI082630/AI/NIAID NIH HHS/ P01 AI112521/AI/NIAID NIH HHS/ AI05343/AI/NIAID NIH HHS/ R01 AI105343/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review England 2015/03/24 Trends Immunol. 2015 Apr;36(4):265-76. doi: 10.1016/j.it.2015.02.008. Epub 2015 Mar 18.
- [77] E. J. Wherry and M. Kurachi. Molecular and cellular insights into t cell exhaustion. *Nat Rev Immunol*, 15(8):486–99, 2015. Wherry, E John Kurachi, Makoto eng AI105343/AI/NIAID NIH HHS/ AI112521/AI/NIAID NIH HHS/ AI082630/AI/NIAID NIH HHS/ U19 AI082630/AI/NIAID NIH HHS/ P01 AI112521/AI/NIAID NIH HHS/ AI095608/AI/NIAID NIH HHS/ HHSN266200500030C/PHS HHS/ R01 AI105343/AI/NIAID NIH HHS/ U01 AI095608/AI/NIAID NIH HHS/ HHSN266200500030C/AI/NIAID NIH

- HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review England 2015/07/25 Nat Rev Immunol. 2015 Aug;15(8):486-99. doi: 10.1038/nri3862.
- [78] M. Philip, L. Fairchild, L. Sun, E. L. Horste, S. Camara, M. Shakiba, A. C. Scott, A. Viale, P. Lauer, T. Merghoub, M. D. Hellmann, J. D. Wolchok, C. S. Leslie, and A. Schietinger. Chromatin states define tumour-specific t cell dysfunction and reprogramming. *Nature*, 545(7655):452–456, 2017. Philip, Mary Fairchild, Lauren Sun, Liping Horste, Ellen L Camara, Steven Shakiba, Mojdeh Scott, Andrew C Viale, Agnes Lauer, Peter Merghoub, Taha Hellmann, Matthew D Wolchok, Jedd D Leslie, Christina S Schietinger, Andrea eng U01 HG007893/HG/NHGRI NIH HHS/ K08 CA158069/CA/NCI NIH HHS/ P30 CA008748/CA/NCI NIH HHS/ R00 CA172371/CA/NCI NIH HHS/ U54 CA209975/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2017/05/18 Nature. 2017 May 25;545(7655):452-456. doi: 10.1038/nature22367. Epub 2017 May 17.
- [79] Q. Guan, M. Han, Q. Guo, F. Yan, M. Wang, Q. Ning, and D. Xi. Strategies to reinvigorate exhausted cd8(+) t cells in tumor microenvironment. *Front Immunol*, 14:1204363, 2023. Guan, Qianting Han, Meiwen Guo, Qinghao Yan, Fangfei Wang, Ming Ning, Qin Xi, Dong eng Research Support, Non-U.S. Gov't Review Switzerland 2023/07/03 Front Immunol. 2023 Jun 16;14:1204363. doi: 10.3389/fimmu.2023.1204363. eCollection 2023.
- [80] Y. Jiang, Y. Li, and B. Zhu. T-cell exhaustion in the tumor microenvironment. *Cell Death Dis*, 6(6):e1792, 2015. Jiang, Y Li, Y Zhu, B eng Research Support, Non-U.S. Gov't Review England 2015/06/19 Cell Death Dis. 2015 Jun 18;6(6):e1792. doi: 10.1038/cddis.2015.162.
- [81] D. E. Speiser, D. T. Utzschneider, S. G. Oberle, C. Munz, P. Romero, and D. Zehn. T cell differentiation in chronic infection and cancer: functional adaptation or exhaustion? *Nat Rev Immunol*, 14(11):768–74, 2014. Speiser, Daniel E Utzschneider, Daniel T Oberle, Susanne G Munz, Christian Romero, Pedro Zehn, Dietmar eng England 2014/09/27 Nat Rev Immunol. 2014 Nov;14(11):768-74. doi: 10.1038/nri3740. Epub 2014 Sep 26.
- [82] A. Schietinger and P. D. Greenberg. Tolerance and exhaustion: defining mechanisms of t cell dysfunction. *Trends Immunol*, 35(2):51–60, 2014. Schietinger, Andrea Greenberg, Philip D eng K99 CA172371/CA/NCI NIH HHS/ P01 CA018029/CA/NCI NIH HHS/ R01 CA033084/CA/NCI NIH HHS/ R21 AI107776/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review England 2013/11/12 Trends Immunol. 2014 Feb;35(2):51-60. doi: 10.1016/j.it.2013.10.001. Epub 2013 Nov 6.
- [83] I. Tirosh, B. Izar, S. M. Prakadan, 2nd Wadsworth, M. H., D. Treacy, J. J. Trombetta, A. Rothenberg, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, and L. A. Garraway. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science*, 352(6282):189–96, 2016. Tirosh, Itay Izar, Benjamin Prakadan, Sanjay M Wadsworth, Marc H 2nd Treacy, Daniel Trombetta, John J

Rotem, Asaf Rodman, Christopher Lian, Christine Murphy, George Fallahi-Sichani, Mohammad Dutton-Register, Ken Lin, Jia-Ren Cohen, Ofir Shah, Parin Lu, Diana Genshaft, Alex S Hughes, Travis K Ziegler, Carly G K Kazer, Samuel W Gaillard, Aleth Kolb, Kellie E Villani, Alexandra-Chloe Johannessen, Cory M Andreev, Aleksandr Y Van Allen, Eliezer M Bertagnolli, Monica Sorger, Peter K Sullivan, Ryan J Flaherty, Keith T Frederick, Dennie T Jane-Valbuena, Judit Yoon, Charles H Rozenblatt-Rosen, Orit Shalek, Alex K Regev, Aviv Garraway, Levi A eng U24 CA180922/CA/NCI NIH HHS/ K99 CA194163/CA/NCI NIH HHS/ P50GM107618/GM/NIGMS NIH HHS/ P50 GM107618/GM/NIGMS NIH HHS/ P01 CA163222/CA/NCI NIH HHS/ P30 CA014051/CA/NCI NIH HHS/ T32 GM007753/GM/NIGMS NIH HHS/ P30-CA14051/CA/NCI NIH HHS/ DP2 GM119419/GM/NIGMS NIH HHS/ U54CA112962/CA/NCI NIH HHS/ R35CA197737/CA/NCI NIH HHS/ P01CA163222/CA/NCI NIH HHS/ R35 CA197737/CA/NCI NIH HHS/ R33 CA202820/CA/NCI NIH HHS/ K99CA194163/CA/NCI NIH HHS/ Howard Hughes Medical Institute/ T32 GM008313/GM/NIGMS NIH HHS/ U54 CA112962/CA/NCI NIH HHS/ U24 AI118672/AI/NIAID NIH HHS/ DP2 OD020839/OD/NIH HHS/ 1U24CA180922/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2016/04/29 Science. 2016 Apr 8;352(6282):189-96. doi: 10.1126/science.aad0501.

- [84] H. Li, A. M. van der Leun, I. Yofe, Y. Lubling, D. Gelbard-Solodkin, A. C. J. van Akkooi, M. van den Braber, E. A. Rozeman, Jbag Haanen, C. U. Blank, H. M. Horlings, E. David, Y. Baran, A. Bercovich, A. Lifshitz, T. N. Schumacher, A. Tanay, and I. Amit. Dysfunctional cd8 t cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell*, 176(4):775–789 e18, 2019. Li, Hanjie van der Leun, Anne M Yofe, Ido Lubling, Yaniv Gelbard-Solodkin, Dikla van Akkooi, Alexander C J van den Braber, Marlous Rozeman, Elisa A Haanen, John B A G Blank, Christian U Horlings, Hugo M David, Eyal Baran, Yael Bercovich, Akhiad Lifshitz, Aviezer Schumacher, Ton N Tanay, Amos Amit, Ido eng 724471/ERC\_/European Research Council/International 724824/ERC\_/European Research Council/International 742259/ERC\_/European Research Council/International HHMI/Howard Hughes Medical Institute/ Research Support, Non-U.S. Gov't 2019/01/01 Cell. 2019 Feb 7;176(4):775-789.e18. doi: 10.1016/j.cell.2018.11.043. Epub 2018 Dec 27.
- [85] K. Kim, S. Park, S. Y. Park, G. Kim, S. M. Park, J. W. Cho, D. H. Kim, Y. M. Park, Y. W. Koh, H. R. Kim, S. J. Ha, and I. Lee. Single-cell transcriptome analysis reveals tox as a promoting factor for t cell exhaustion and a predictor for anti-pd-1 responses in human cancer. *Genome Med*, 12(1):22, 2020. Kim, Kyungsoo Park, Seyeon Park, Seong Yong Kim, Gamin Park, Su Myeong Cho, Jae-Won Kim, Da Hee Park, Young Min Koh, Yoon Woo Kim, Hye Ryun Ha, Sang-Jun Lee, Insuk eng Research Support, Non-U.S. Gov't England 2020/03/01 Genome Med. 2020 Feb 28;12(1):22. doi: 10.1186/s13073-020-00722-9.
- [86] J. C. Beltra, S. Manne, M. S. Abdel-Hakeem, M. Kurachi, J. R. Giles, Z. Chen, V. Casella, S. F. Ngiow, O. Khan, Y. J. Huang, P. Yan, K. Nzingha, W. Xu, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, L. M. Schuchter, A. C. Huang, and E. J. Wherry. Developmental relationships of four exhausted cd8(+) t cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. *Immunity*, 52(5):825–841 e8,

2020. Beltra, Jean-Christophe Manne, Sasikanth Abdel-Hakeem, Mohamed S Kurachi, Makoto Giles, Josephine R Chen, Zeyu Casella, Valentina Ngiow, Shin Foong Khan, Omar Huang, Yinghui Jane Yan, Patrick Nzingha, Kito Xu, Wei Amaravadi, Ravi K Xu, Xiaowei Karakousis, Giorgos C Mitchell, Tara C Schuchter, Lynn M Huang, Alexander C Wherry, E John eng P01 CA210944/CA/NCI NIH HHS/ U19 AI117950/AI/NIAID NIH HHS/ F99 CA234842/CA/NCI NIH HHS/ K08 CA230157/CA/NCI NIH HHS/ U19 AI082630/AI/NIAID NIH HHS/ P50 CA174523/CA/NCI NIH HHS/ R01 AI105343/AI/NIAID NIH HHS/ T32 CA009140/CA/NCI NIH HHS/ P01 CA114046/CA/NCI NIH HHS/ P30 CA016520/CA/NCI NIH HHS/ K00 CA234842/CA/NCI NIH HHS/ T32 CA009615/CA/NCI NIH HHS/ P01 AI108545/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2020/05/13 Immunity. 2020 May 19;52(5):825-841.e8. doi: 10.1016/j.immuni.2020.04.014. Epub 2020 May 11.

[87] B. C. Miller, D. R. Sen, R. Al Abosy, K. Bi, Y. V. Virkud, M. W. LaFleur, K. B. Yates, A. Lako, K. Felt, G. S. Naik, M. Manos, E. Gjini, J. R. Kuchroo, J. J. Ishizuka, J. L. Collier, G. K. Griffin, S. Maleri, D. E. Comstock, S. A. Weiss, F. D. Brown, A. Panda, M. D. Zimmer, R. T. Manguso, F. S. Hodi, S. J. Rodig, A. H. Sharpe, and W. N. Haining. Subsets of exhausted cd8(+) t cells differentially mediate tumor control and respond to checkpoint blockade. *Nat Immunol*, 20(3):326–336, 2019. Miller, Brian C Sen, Debattama R Al Abosy, Rose Bi, Kevin Virkud, Yamini V LaFleur, Martin W Yates, Kathleen B Lako, Ana Felt, Kristen Naik, Girish S Manos, Michael Gjini, Evisa Kuchroo, Juhi R Ishizuka, Jeffrey J Collier, Jenna L Griffin, Gabriel K Maleri, Seth Comstock, Dawn E Weiss, Sarah A Brown, Flavian D Panda, Arpit Zimmer, Margaret D Manguso, Robert T Hodi, F Stephen Rodig, Scott J Sharpe, Arlene H Haining, W Nicholas eng KL2 TR002542/TR/NCATS NIH HHS/ P01 AI056299/AI/NIAID NIH HHS/ T32 HL007627/HL/NHLBI NIH HHS/ T32 GM007753/GM/NIGMS NIH HHS/ K23 AI130408/AI/NIAID NIH HHS/ T32 CA009172/CA/NCI NIH HHS/ U19 AI133524/AI/NIAID NIH HHS/ R01 AI115712/AI/NIAID NIH HHS/ T32 CA207021/CA/NCI NIH HHS/ Research Support, Non-U.S. Gov't 2019/02/20 *Nat Immunol*. 2019 Mar;20(3):326-336. doi: 10.1038/s41590-019-0312-6. Epub 2019 Feb 18.

[88] K. Bi, M. X. He, Z. Bakouny, A. Kanodia, S. Napolitano, J. Wu, G. Grimaldi, D. A. Braun, M. S. Cuoco, A. Mayorga, L. DelloStritto, G. Bouchard, J. Steinharter, A. K. Tewari, N. I. Vokes, E. Shannon, M. Sun, J. Park, S. L. Chang, B. A. McGregor, R. Haq, T. Denize, S. Signoretti, J. L. Guerriero, S. Vigneau, O. Rozenblatt-Rosen, A. Rotem, A. Regev, T. K. Choueiri, and E. M. Van Allen. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. *Cancer Cell*, 39(5):649–661 e5, 2021. Bi, Kevin He, Meng Xiao Bakouny, Ziad Kanodia, Abhay Napolitano, Sara Wu, Jingyi Grimaldi, Grace Braun, David A Cuoco, Michael S Mayorga, Angie DelloStritto, Laura Bouchard, Gabrielle Steinharter, John Tewari, Alok K Vokes, Natalie I Shannon, Erin Sun, Maxine Park, Jihye Chang, Steven L McGregor, Bradley A Haq, Rizwan Denize, Thomas Signoretti, Sabina Guerriero, Jennifer L Vigneau, Sebastien Rozenblatt-Rosen, Orit Rotem, Asaf Regev, Aviv Choueiri, Toni K Van Allen, Eliezer M eng U01 CA233100/CA/NCI NIH HHS/ P50 CA101942/CA/NCI NIH HHS/ T32 CA009172/CA/NCI NIH HHS/ R01 CA227388/CA/NCI NIH HHS/ T32

- GM008313/GM/NIGMS NIH HHS/ U2C CA233195/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2021/03/13 Cancer Cell. 2021 May 10;39(5):649-661.e5. doi: 10.1016/j.ccell.2021.02.015. Epub 2021 Mar 11.
- [89] B. Daniel, K. E. Yost, S. Hsiung, K. Sandor, Y. Xia, Y. Qi, K. J. Hiam-Galvez, M. Black, J. Raposo C, Q. Shi, S. L. Meier, J. A. Belk, J. R. Giles, E. J. Wherry, H. Y. Chang, T. Egawa, and A. T. Satpathy. Divergent clonal differentiation trajectories of t cell exhaustion. *Nat Immunol*, 23(11):1614–1627, 2022. Daniel, Bence Yost, Kathryn E Hsiung, Sunnie Sandor, Katalin Xia, Yu Qi, Yanyan Hiam-Galvez, Kamir J Black, Mollie J Raposo, Colin Shi, Quanming Meier, Stefanie L Belk, Julia A Giles, Josephine R Wherry, E John Chang, Howard Y Egawa, Takeshi Satpathy, Ansuman T eng K08 CA230188/CA/NCI NIH HHS/ S10 OD021763/OD/NIH HHS/ T32 CA009140/CA/NCI NIH HHS/ T32 AI007290/AI/NIAID NIH HHS/ HHMI/Howard Hughes Medical Institute/ RM1 HG007735/HG/NHGRI NIH HHS/ F99 CA253729/CA/NCI NIH HHS/ U01 CA260852/CA/NCI NIH HHS/ R01 AI130152/AI/NIAID NIH HHS/ S10 OD018220/OD/NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2022/10/28 *Nat Immunol*. 2022 Nov;23(11):1614-1627. doi: 10.1038/s41590-022-01337-5. Epub 2022 Oct 26.
- [90] A. Haslam, J. Gill, and V. Prasad. Estimation of the percentage of us patients with cancer who are eligible for immune checkpoint inhibitor drugs. *JAMA Netw Open*, 3(3):e200423, 2020. Haslam, Alyson Gill, Jennifer Prasad, Vinay eng 2020/03/10 *JAMA Netw Open*. 2020 Mar 2;3(3):e200423. doi: 10.1001/jamanetworkopen.2020.0423.
- [91] M. Philip and A. Schietinger. Heterogeneity and fate choice: T cell exhaustion in cancer and chronic infections. *Curr Opin Immunol*, 58:98–103, 2019. Philip, Mary Schietinger, Andrea eng DP2 CA225212/CA/NCI NIH HHS/ P30 CA008748/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review England 2019/06/11 *Curr Opin Immunol*. 2019 Jun;58:98-103. doi: 10.1016/j.coi.2019.04.014. Epub 2019 Jun 8.
- [92] S. Wang, Q. Zhang, H. Hui, K. Agrawal, M. A. Y. Karris, and T. M. Rana. An atlas of immune cell exhaustion in hiv-infected individuals revealed by single-cell transcriptomics. *Emerg Microbes Infect*, 9(1):2333–2347, 2020. Wang, Shaobo Zhang, Qiong Hui, Hui Agrawal, Kriti Karris, Maile Ann Young Rana, Tariq M eng P30 AI036214/AI/NIAID NIH HHS/ 2020/09/22 *Emerg Microbes Infect*. 2020 Dec;9(1):2333-2347. doi: 10.1080/22221751.2020.1826361.
- [93] J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, and J. Shendure. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496–502, 2019. Cao, Junyue Spielmann, Malte Qiu, Xiaojie Huang, Xingfan Ibrahim, Daniel M Hill, Andrew J Zhang, Fan Mundlos, Stefan Christiansen, Lena Steemers, Frank J Trapnell, Cole Shendure, Jay eng HHMI/Howard Hughes Medical Institute/ R01 HL118342/HL/NHLBI NIH HHS/ U54 HL145611/HL/NHLBI NIH HHS/ R01 HG006283/HG/NHGRI NIH HHS/ DP2 HD088158/HD/NICHD NIH HHS/ U54 DK107979/DK/NIDDK NIH HHS/ DP1 HG007811/HG/NHGRI NIH HHS/ Research Support,

N.I.H., Extramural Research Support, Non-U.S. Gov't England 2019/02/23 Nature. 2019 Feb;566(7745):496-502. doi: 10.1038/s41586-019-0969-x. Epub 2019 Feb 20.

- [94] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–50, 2005. Subramanian, Aravind Tamayo, Pablo Mootha, Vamsi K Mukherjee, Sayan Ebert, Benjamin L Gillette, Michael A Paulovich, Amanda Pomeroy, Scott L Golub, Todd R Lander, Eric S Mesirov, Jill P eng R01 CA109467/CA/NCI NIH HHS/ T32 CA009172/CA/NCI NIH HHS/ 2005/10/04 Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50. doi: 10.1073/pnas.0506580102. Epub 2005 Sep 30.
- [95] S. Lukhele, D. A. Rabbo, M. Guo, J. Shen, H. J. Elsaesser, R. Quevedo, M. Carew, R. Gadalla, L. M. Snell, L. Mahesh, M. T. Ciudad, B. E. Snow, A. You-Ten, J. Haight, A. Wakeham, P. S. Ohashi, T. W. Mak, W. Cui, T. L. McGaha, and D. G. Brooks. The transcription factor irf2 drives interferon-mediated cd8(+) t cell exhaustion to restrict anti-tumor immunity. *Immunity*, 55(12):2369–2385 e10, 2022. Lukhele, Sabelo Rabbo, Diala Abd Guo, Mengdi Shen, Jian Elsaesser, Heidi J Quevedo, Rene Carew, Madeleine Gadalla, Ramy Snell, Laura M Mahesh, Lawanya Ciudad, M Teresa Snow, Bryan E You-Ten, Annick Haight, Jillian Wakeham, Andrew Ohashi, Pamela S Mak, Tak W Cui, Weiguo McGaha, Tracy L Brooks, David G eng R01 CA255670/CA/NCI NIH HHS/ R01 AI085043/AI/NIAID NIH HHS/ R01 AI148403/AI/NIAID NIH HHS/ P20 GM121176/GM/NIGMS NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2022/11/13 Immunity. 2022 Dec 13;55(12):2369-2385.e10. doi: 10.1016/j.immuni.2022.10.020. Epub 2022 Nov 11.
- [96] Q. Shan, S. Zhu, X. Chen, J. Liu, S. Yuan, X. Li, W. Peng, and H. H. Xue. Tcf1-ctcf cooperativity shapes genomic architecture to promote cd8(+) t cell homeostasis. *Nat Immunol*, 23(8):1222–1235, 2022. Shan, Qiang Zhu, Shaoqi Chen, Xia Liu, Jia Yuan, Shuang Li, Xiang Peng, Weiqun Xue, Hai-Hui eng I01 BX005771/BX/BLRD VA/ R01 AI112579/AI/NIAID NIH HHS/ R01 AI121080/AI/NIAID NIH HHS/ R01 AI139874/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. 2022/07/27 Nat Immunol. 2022 Aug;23(8):1222-1235. doi: 10.1038/s41590-022-01263-6. Epub 2022 Jul 26.
- [97] S. Quon, B. Yu, B. E. Russ, K. Tsyganov, H. Nguyen, C. Toma, M. Heeg, J. D. Hocker, J. J. Milner, S. Crotty, M. E. Pipkin, S. J. Turner, and A. W. Goldrath. Dna architectural protein ctcf facilitates subset-specific chromatin interactions to limit the formation of memory cd8(+) t cells. *Immunity*, 56(5):959–978 e10, 2023. Quon, Sara Yu, Bingfei Russ, Brendan E Tsyganov, Kirill Nguyen, Hongtuyet Toma, Clara Heeg, Maximilian Hocker, James D Milner, J Justin Crotty, Shane Pipkin, Matthew E Turner, Stephen J Goldrath, Ananda W eng P01 AI132122/AI/NIAID NIH HHS/ R01 AI067545/AI/NIAID NIH HHS/ R37 AI067545/AI/NIAID NIH HHS/ U19 AI109976/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2023/04/12 Immunity. 2023 May 9;56(5):959-978.e10. doi: 10.1016/j.immuni.2023.03.017. Epub 2023 Apr 10.

- [98] X. Gao, K. Tewari, J. Svaren, and M. Suresh. Role of cell cycle regulator e2f1 in regulating cd8 t cell responses during acute and chronic viral infection. *Virology*, 324(2):567–76, 2004. Gao, Xiaoyan Tewari, Kavita Svaren, John Suresh, M eng AI 48785/AI/NIAID NIH HHS/ Comparative Study Research Support, U.S. Gov't, P.H.S. 2004/06/23 Virology. 2004 Jul 1;324(2):567-76. doi: 10.1016/j.virol.2004.04.012.
- [99] M. Yang, C. Lin, Y. Wang, K. Chen, Y. Han, H. Zhang, and W. Li. Cytokine storm promoting t cell exhaustion in severe covid-19 revealed by single cell sequencing data analysis. *Precis Clin Med*, 5(2):pbac014, 2022. Yang, Minglei Lin, Chenghao Wang, Yanni Chen, Kang Han, Yutong Zhang, Haiyue Li, Weizhong eng England 2022/06/14 Precs Clin Med. 2022 May 23;5(2):pbac014. doi: 10.1093/pccmedi/pbac014. eCollection 2022 Jun.
- [100] J. R. Giles, S. Manne, E. Freilich, D. A. Oldridge, A. E. Baxter, S. George, Z. Chen, H. Huang, L. Chilukuri, M. Carberry, L. Giles, N. P. Weng, R. M. Young, C. H. June, L. M. Schuchter, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, A. C. Huang, J. Shi, and E. J. Wherry. Human epigenetic and transcriptional t cell differentiation atlas for identifying functional t cell-specific enhancers. *Immunity*, 55(3):557–574 e7, 2022. Giles, Josephine R Manne, Sasikanth Freilich, Elizabeth Oldridge, Derek A Baxter, Amy E George, Sangeeth Chen, Zeyu Huang, Hua Chilukuri, Lakshmi Carberry, Mary Giles, Lydia Weng, Nan-Ping P Young, Regina M June, Carl H Schuchter, Lynn M Amaravadi, Ravi K Xu, Xiaowei Karakousis, Giorgos C Mitchell, Tara C Huang, Alexander C Shi, Junwei Wherry, E John eng P01 CA210944/CA/NCI NIH HHS/ U19 AI117950/AI/NIAID NIH HHS/ U19 AI082630/AI/NIAID NIH HHS/ P50 CA174523/CA/NCI NIH HHS/ R01 AI115712/AI/NIAID NIH HHS/ R38 HL143613/HL/NHLBI NIH HHS/ P01 AI108545/AI/NIAID NIH HHS/ T32 CA009140/CA/NCI NIH HHS/ P50 CA261608/CA/NCI NIH HHS/ R01 AI155577/AI/NIAID NIH HHS/ F99 CA234842/CA/NCI NIH HHS/ P30 AI045008/AI/NIAID NIH HHS/ K08 CA230157/CA/NCI NIH HHS/ P30 CA016520/CA/NCI NIH HHS/ K00 CA234842/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2022/03/10 Immunity. 2022 Mar 8;55(3):557-574.e7. doi: 10.1016/j.immuni.2022.02.004.
- [101] A. El Hage and O. Dormond. Combining mtor inhibitors and t cell-based immunotherapies in cancer treatment. *Cancers (Basel)*, 13(6), 2021. El Hage, Alexandre Dormond, Olivier eng Review Switzerland 2021/04/04 Cancers (Basel). 2021 Mar 17;13(6):1359. doi: 10.3390/cancers13061359.
- [102] Masashi Numata, Ramon Klein Geltink, and Gerard Grosveld. The ets transcription factor etv7 exhausts hematopoietic stem cells by enhancing the cell cycle entry and cell proliferation. *Blood*, 122(21):733–733, 2013.
- [103] D. Corridoni, A. Antanaviciute, T. Gupta, D. Fawcner-Corbett, A. Aulicino, M. Jagielowicz, K. Parikh, E. Repapi, S. Taylor, D. Ishikawa, R. Hatano, T. Yamada, W. Xin, H. Slawinski, R. Bowden, G. Napolitani, O. Brain, C. Morimoto, H. Koohy, and A. Simmons. Single-cell atlas of colonic cd8(+) t cells in ulcerative colitis. *Nat Med*, 26(9):1480–1490, 2020. Corridoni, Daniele Antanaviciute, Agne Gupta, Tarun Fawcner-Corbett, David Aulicino, Anna Jagielowicz, Marta Parikh, Kaushal Repapi, Emmanouela Taylor, Steve Ishikawa, Dai Hatano, Ryo Yamada, Taketo Xin, Wei Slawinski, Hubert Bowden, Rory Napolitani, Giorgio Brain, Oliver Morimoto, Chikao Koohy, Hashem Simmons, Alison eng MR/M00919X/1/MRC\_/Medical

Research Council/United Kingdom 219523/Z/19/Z/WT\_/Wellcome Trust/United Kingdom DH\_/Department of Health/United Kingdom MC\_PC\_MR/S025952/1/MRC\_/Medical Research Council/United Kingdom MR/S036377/1/MRC\_/Medical Research Council/United Kingdom MC\_UU\_00008/7/MRC\_/Medical Research Council/United Kingdom MC\_UU\_12010/7/MRC\_/Medical Research Council/United Kingdom Research Support, Non-U.S. Gov't 2020/08/05 Nat Med. 2020 Sep;26(9):1480-1490. doi: 10.1038/s41591-020-1003-4. Epub 2020 Aug 3.

- [104] L. M. McLane, S. F. Ngiow, Z. Chen, J. Attanasio, S. Manne, G. Ruthel, J. E. Wu, R. P. Staupe, W. Xu, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, L. M. Schuchter, A. C. Huang, B. D. Freedman, M. R. Betts, and E. J. Wherry. Role of nuclear localization in the regulation and function of t-bet and eomes in exhausted cd8 t cells. *Cell Rep*, 35(6):109120, 2021. McLane, Laura M Ngiow, Shin Foong Chen, Zeyu Attanasio, John Manne, Sasikanth Ruthel, Gordon Wu, Jennifer E Staupe, Ryan P Xu, Wei Amaravadi, Ravi K Xu, Xiaowei Karakousis, Giorgos C Mitchell, Tara C Schuchter, Lynn M Huang, Alexander C Freedman, Bruce D Betts, Michael R Wherry, E John eng R21 AI117718/AI/NIAID NIH HHS/ U19 AI117950/AI/NIAID NIH HHS/ P01 CA114046/CA/NCI NIH HHS/ U19 AI082630/AI/NIAID NIH HHS/ P50 CA174523/CA/NCI NIH HHS/ R01 AI115712/AI/NIAID NIH HHS/ S10 RR027128/RR/NCRR NIH HHS/ S10 OD021633/OD/NIH HHS/ P01 AI108545/AI/NIAID NIH HHS/ R01 AI105343/AI/NIAID NIH HHS/ P01 AI112521/AI/NIAID NIH HHS/ R56 AI125415/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2021/05/13 Cell Rep. 2021 May 11;35(6):109120. doi: 10.1016/j.celrep.2021.109120.
- [105] M. Buggert, J. Tauriainen, T. Yamamoto, J. Frederiksen, M. A. Ivarsson, J. Michaelsson, O. Lund, B. Hejdeman, M. Jansson, A. Sonnerborg, R. A. Koup, M. R. Betts, and A. C. Karlsson. T-bet and eomes are differentially linked to the exhausted phenotype of cd8+ t cells in hiv infection. *PLoS Pathog*, 10(7):e1004251, 2014. Buggert, Marcus Tauriainen, Johanna Yamamoto, Takuya Frederiksen, Juliet Ivarsson, Martin A Michaelsson, Jakob Lund, Ole Hejdeman, Bo Jansson, Marianne Sonnerborg, Anders Koup, Richard A Betts, Michael R Karlsson, Annika C eng P30 AI045008/AI/NIAID NIH HHS/ R01 AI076066/AI/NIAID NIH HHS/ Clinical Trial Research Support, Non-U.S. Gov't 2014/07/18 PLoS Pathog. 2014 Jul 17;10(7):e1004251. doi: 10.1371/journal.ppat.1004251. eCollection 2014 Jul.
- [106] J. Li, Y. He, J. Hao, L. Ni, and C. Dong. High levels of eomes promote exhaustion of anti-tumor cd8(+) t cells. *Front Immunol*, 9:2981, 2018. Li, Jing He, Yi Hao, Jing Ni, Ling Dong, Chen eng Research Support, Non-U.S. Gov't Switzerland 2019/01/09 Front Immunol. 2018 Dec 18;9:2981. doi: 10.3389/fimmu.2018.02981. eCollection 2018.
- [107] H. He, Y. Yi, X. Cai, J. Wang, X. Ni, Y. Fu, and S. Qiu. Down-regulation of eomes drives t-cell exhaustion via abolishing eomes-mediated repression of inhibitory receptors of t cells in liver cancer. *J Cell Mol Med*, 25(1):161–169, 2021. He, Hongwei Yi, Yong Cai, Xiaoyan Wang, Jiaying Ni, Xiaochun Fu, Yipeng Qiu, Shuangjian eng 2012ZX10002010-001/002/National Key Sci-Tech Special Project of China/ 81302102/National Natural Science Foundation of China/ 81602486/National Natural Science Foundation of China/ Research



Support, Non-U.S. Gov't England 2020/12/17 J Cell Mol Med. 2021 Jan;25(1):161-169. doi: 10.1111/jcmm.15898. Epub 2020 Dec 16.

- [108] Yi Zhong, Sarah K. Walker, Yuri Pritykin, Christina S. Leslie, Alexander Y. Rudensky, and Joris Van Der Veecken. Hierarchical regulation of the resting and activated t cell epigenome by major transcription factor families. *Nature Immunology*, 23(1):122–134, 2022.
- [109] R. Grenningloh, T. S. Tai, N. Frahm, T. C. Hongo, A. T. Chicoine, C. Brander, D. E. Kaufmann, and I. C. Ho. Ets-1 maintains il-7 receptor expression in peripheral t cells. *J Immunol*, 186(2):969–76, 2011. Grenningloh, Roland Tai, Tzong-Shyuan Frahm, Nicole Hongo, Tomoyuki C Chicoine, Adam T Brander, Christian Kaufmann, Daniel E Ho, I-Cheng eng HL092565/HL/NHLBI NIH HHS/ R03 AI081052/AI/NIAID NIH HHS/ R01 HL092565/HL/NHLBI NIH HHS/ R03 AI081052-02/AI/NIAID NIH HHS/ R01 HL092565-05/HL/NHLBI NIH HHS/ AI081052/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2010/12/15 J Immunol. 2011 Jan 15;186(2):969-76. doi: 10.4049/jimmunol.1002099. Epub 2010 Dec 10.
- [110] J. K. Gicobi, E. R. Dellacecca, and H. Dong. Resilient t-cell responses in patients with advanced cancers. *Int J Hematol*, 117(5):634–639, 2023. Gicobi, Joanina K Dellacecca, Emilia R Dong, Haidong eng Review Japan 2022/07/22 Int J Hematol. 2023 May;117(5):634-639. doi: 10.1007/s12185-022-03424-7. Epub 2022 Jul 21.
- [111] B. Gerby, C. S. Tremblay, M. Tremblay, S. Rojas-Sutterlin, S. Herblot, J. Hebert, G. Sauvageau, S. Lemieux, E. Lecuyer, D. F. Veiga, and T. Hoang. Sc1, lmo1 and notch1 reprogram thymocytes into self-renewing cells. *PLoS Genet*, 10(12):e1004768, 2014. Gerby, Bastien Tremblay, Cedric S Tremblay, Mathieu Rojas-Sutterlin, Shanti Herblot, Sabine Hebert, Josee Sauvageau, Guy Lemieux, Sebastien Lecuyer, Eric Veiga, Diogo F T Hoang, Trang eng Canadian Institutes of Health Research/Canada Research Support, Non-U.S. Gov't 2014/12/19 PLoS Genet. 2014 Dec 18;10(12):e1004768. doi: 10.1371/journal.pgen.1004768. eCollection 2014 Dec.
- [112] W. Wu, X. Liang, H. Li, X. Huang, C. Wan, Q. Xie, and Z. Liu. Landscape of t cells in nk-aml(m4/m5) revealed by single-cell sequencing. *J Leukoc Biol*, 112(4):745–758, 2022. Wu, Wenqi Liang, Xiaolin Li, Huiqun Huang, Xiaoke Wan, Chengyao Xie, Qiongni Liu, Zhenfang eng Research Support, Non-U.S. Gov't England 2022/03/09 J Leukoc Biol. 2022 Oct;112(4):745-758. doi: 10.1002/JLB.5A0721-396RR. Epub 2022 Mar 8.
- [113] B. Shin and E. V. Rothenberg. Multi-modular structure of the gene regulatory network for specification and commitment of murine t cells. *Front Immunol*, 14:1108368, 2023. Shin, Boyoung Rothenberg, Ellen V eng R01 AI135200/AI/NIAID NIH HHS/ R01 AI151704/AI/NIAID NIH HHS/ R01 HD100039/HD/NICHD NIH HHS/ Research Support, N.I.H., Extramural Review Switzerland 2023/02/24 Front Immunol. 2023 Jan 31;14:1108368. doi: 10.3389/fimmu.2023.1108368. eCollection 2023.
- [114] W. Zhou, F. Gao, M. Romero-Wolf, S. Jo, and E. V. Rothenberg. Single-cell deletion analyses show control of pro-t cell developmental speed and pathways by tcf7, spi1, gata3, bcl11a, erg, and bcl11b. *Sci Immunol*, 7(71):eabm1920, 2022. Zhou, Wen Gao, Fan Romero-Wolf, Maile Jo, Suin Rothenberg, Ellen V eng R01 AI135200/AI/NIAID NIH HHS/ R01 HD100039/HD/NICHD NIH HHS/ R01 HL119102/HL/NHLBI NIH

- HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2022/05/21 Sci Immunol. 2022 May 20;7(71):eabm1920. doi: 10.1126/sciimmunol.abm1920. Epub 2022 May 20.
- [115] S. Luc, J. Huang, J. L. McEldoon, E. Somuncular, D. Li, C. Rhodes, S. Mamoor, S. Hou, J. Xu, and S. H. Orkin. Bcl11a deficiency leads to hematopoietic stem cell defects with an aging-like phenotype. *Cell Rep*, 16(12):3181–3194, 2016. Luc, Sidinh Huang, Jialiang McEldoon, Jennifer L Somuncular, Ece Li, Dan Rhodes, Claire Mamoor, Shahan Hou, Serena Xu, Jian Orkin, Stuart H eng R03 DK101665/DK/NIDDK NIH HHS/ R01 HL032259/HL/NHLBI NIH HHS/ HHMI/Howard Hughes Medical Institute/ P30 DK049216/DK/NIDDK NIH HHS/ P01 HL032262/HL/NHLBI NIH HHS/ U54 DK110805/DK/NIDDK NIH HHS/ K01 DK093543/DK/NIDDK NIH HHS/ R01 DK111430/DK/NIDDK NIH HHS/ 2016/09/23 Cell Rep. 2016 Sep 20;16(12):3181-3194. doi: 10.1016/j.celrep.2016.08.064.
- [116] Ken-ichi Hirano, Hiroyuki Hosokawa, Maria Koizumi, Yusuke Endo, Takashi Yahata, Kiyoshi Ando, and Katsuto Hozumi. Lmo2 is essential to maintain the ability of progenitors to differentiate into t-cell lineage in mice. *eLife*, 10:e68227, 2021.
- [117] Y. W. Zhang and N. Cabezas-Wallscheid. Hscs: slow me down with plag1. *Blood*, 140(9):935–936, 2022. Zhang, Yu Wei Cabezas-Wallscheid, Nina eng Comment Editorial 2022/09/02 Blood. 2022 Sep 1;140(9):935-936. doi: 10.1182/blood.2022017069.
- [118] X. Zhao, Q. Shan, and H. H. Xue. Tcf1 in t cell immunity: a broadened frontier. *Nat Rev Immunol*, 22(3):147–157, 2022. Zhao, Xudong Shan, Qiang Xue, Hai-Hui eng Review England 2021/06/16 Nat Rev Immunol. 2022 Mar;22(3):147-157. doi: 10.1038/s41577-021-00563-6. Epub 2021 Jun 14.
- [119] J. G. Crompton, M. Sukumar, and N. P. Restifo. Uncoupling t-cell expansion from effector differentiation in cell-based immunotherapy. *Immunol Rev*, 257(1):264–276, 2014. Crompton, Joseph G Sukumar, Madhusudhanan Restifo, Nicholas P eng ZIA BC010763-08/Intramural NIH HHS/ Review England 2013/12/18 Immunol Rev. 2014 Jan;257(1):264-276. doi: 10.1111/imr.12135.
- [120] E. J. Wherry, S. J. Ha, S. M. Kaech, W. N. Haining, S. Sarkar, V. Kalia, S. Subramaniam, J. N. Blattman, D. L. Barber, and R. Ahmed. Molecular signature of cd8+ t cell exhaustion during chronic viral infection. *Immunity*, 27(4):670–84, 2007. Wherry, E John Ha, Sang-Jun Kaech, Susan M Haining, W Nicholas Sarkar, Surojit Kalia, Vandana Subramaniam, Shruti Blattman, Joseph N Barber, Daniel L Ahmed, Rafi eng AI071309/AI/NIAID NIH HHS/ AI30048/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2007/10/24 Immunity. 2007 Oct;27(4):670-84. doi: 10.1016/j.immuni.2007.09.006. Epub 2007 Oct 18.
- [121] J. Nah and R. H. Seong. Kruppel-like factor 4 regulates the cytolytic effector function of exhausted cd8 t cells. *Sci Adv*, 8(47):eadc9346, 2022. Nah, Jinwoo Seong, Rho H eng 2022/11/26 Sci Adv. 2022 Nov 25;8(47):eadc9346. doi: 10.1126/sciadv.adc9346. Epub 2022 Nov 25.
- [122] Hui Feng, Erik M. Lenarcic, Daisuke Yamane, Eliane Wauthier, Jinyao Mo, Haitao Guo, David R. McGivern, Olga González-López, Ichiro Misumi, Lola M. Reid, Jason K. Whitmire, Jenny P. Y. Ting, Joseph A. Duncan,

- Nathaniel J. Moorman, and Stanley M. Lemon. Nlr1 promotes immediate irf1-directed antiviral responses by limiting dsrna-activated translational inhibition mediated by pkr. *Nature Immunology*, 18(12):1299–1309, 2017.
- [123] Marta Iwanaszko and Marek Kimmel. Nf-b and irf pathways: cross-regulation on target genes promoter level. *BMC Genomics*, 16(1), 2015.
- [124] Chen Zhu, Kaori Sakuishi, Sheng Xiao, Zhiyi Sun, Sarah Zaghouni, Guangxiang Gu, Chao Wang, Dewar J. Tan, Chuan Wu, Manu Rangachari, Thomas Pertel, Hyun-Tak Jin, Rafi Ahmed, Ana C. Anderson, and Vijay K. Kuchroo. An il-27/nfil3 signalling axis drives tim-3 and il-10 expression and t-cell dysfunction. *Nature Communications*, 6(1):6072, 2015.
- [125] Amber M. Smith, Joseph E. Qualls, Kevin O’Brien, Liza Balouzian, Peter F. Johnson, Stacey Schultz-Cherry, Stephen T. Smale, and Peter J. Murray. A distal enhancer in il12b is the target of transcriptional repression by the stat3 pathway and requires the basic leucine zipper (b-zip) protein nfil3. *Journal of Biological Chemistry*, 286(26):23582–23590, 2011.
- [126] Shawn D. Blackburn and E. John Wherry. Il-10, t cell exhaustion and viral persistence. *Trends in Microbiology*, 15(4):143–146, 2007.
- [127] Deepali V. Sawant, Hiroshi Yano, Maria Chikina, Qianxia Zhang, Mengting Liao, Chang Liu, Derrick J. Callahan, Zhe Sun, Tao Sun, Tracy Tabib, Arjun Pennathur, David B. Corry, James D. Luketich, Robert Lafyatis, Wei Chen, Amanda C. Poholek, Tullia C. Bruno, Creg J. Workman, and Dario A. A. Vignali. Adaptive plasticity of il-10+ and il-35+ treg cells cooperatively promotes tumor t cell exhaustion. *Nature Immunology*, 20(6):724–735, 2019.
- [128] Anna M. Wolf, Dominik Wolf, Holger Rumpold, Barbara Enrich, and Herbert Tilg. Adiponectin induces the anti-inflammatory cytokines il-10 and il-1ra in human leukocytes. *Biochemical and Biophysical Research Communications*, 323(2):630–635, 2004.
- [129] F. Petitprez, M. Meylan, A. de Reynies, C. Sautes-Fridman, and W. H. Fridman. The tumor microenvironment in the response to immune checkpoint blockade therapies. *Front Immunol*, 11:784, 2020. Petitprez, Florent Meylan, Maxime de Reynies, Aurelien Sautes-Fridman, Catherine Fridman, Wolf H eng Research Support, Non-U.S. Gov’t Review Switzerland 2020/05/28 Front Immunol. 2020 May 7;11:784. doi: 10.3389/fimmu.2020.00784. eCollection 2020.
- [130] F. A. Wolf, P. Angerer, and F. J. Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biol*, 19(1):15, 2018. Wolf, F Alexander Angerer, Philipp Theis, Fabian J eng Helmholtz Postdoc Grant/Helmholtz-Gemeinschaft/International Research Support, Non-U.S. Gov’t England 2018/02/08 Genome Biol. 2018 Feb 6;19(1):15. doi: 10.1186/s13059-017-1382-0.
- [131] Consortium UniProt. Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Res*, 47(D1):D506–D515, 2019. eng R01 GM080646/GM/NIGMS NIH HHS/ U41 HG002273/HG/NHGRI NIH HHS/ UL1 TR001409/TR/NCATS NIH HHS/ U01 GM120953/GM/NIGMS NIH HHS/ P20 GM103446/GM/NIGMS NIH

HHS/ U41 HG007822/HG/NHGRI NIH HHS/ BB/M011674/1/BB\_/Biotechnology and Biological Sciences Research Council/United Kingdom RG/13/5/30112/BHF\_/British Heart Foundation/United Kingdom Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2018/11/06 Nucleic Acids Res. 2019 Jan 8;47(D1):D506-D515. doi: 10.1093/nar/gky1049.

- [132] Adam Gayoso and Jonathan Shor. Doubletdetection, 2022.
- [133] L. Haghverdi, A. T. L. Lun, M. D. Morgan, and J. C. Marioni. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*, 36(5):421–427, 2018. Haghverdi, Laleh Lun, Aaron T L Morgan, Michael D Marioni, John C eng Wellcome Trust/United Kingdom Research Support, Non-U.S. Gov't 2018/04/03 Nat Biotechnol. 2018 Jun;36(5):421-427. doi: 10.1038/nbt.4091. Epub 2018 Apr 2.
- [134] V. A. Traag, L. Waltman, and N. J. van Eck. From louvain to leiden: guaranteeing well-connected communities. *Sci Rep*, 9(1):5233, 2019. Traag, V A Waltman, L van Eck, N J eng England 2019/03/28 Sci Rep. 2019 Mar 26;9(1):5233. doi: 10.1038/s41598-019-41695-z.
- [135] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [136] E. Becht, L. McInnes, J. Healy, C. A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell. Dimensionality reduction for visualizing single-cell data using umap. *Nat Biotechnol*, 2018. Becht, Etienne McInnes, Leland Healy, John Dutertre, Charles-Antoine Kwok, Immanuel W H Ng, Lai Guan Ginhoux, Florent Newell, Evan W eng 2018/12/12 Nat Biotechnol. 2018 Dec 3. doi: 10.1038/nbt.4314.
- [137] D. van Dijk, R. Sharma, J. Nainys, K. Yim, P. Kathail, A. J. Carr, C. Burdziak, K. R. Moon, C. L. Chaffer, D. Pattabiraman, B. Bierie, L. Mazutis, G. Wolf, S. Krishnaswamy, and D. Pe'er. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3):716–729 e27, 2018. van Dijk, David Sharma, Roshan Nainys, Juozas Yim, Kristina Kathail, Pooja Carr, Ambrose J Burdziak, Cassandra Moon, Kevin R Chaffer, Christine L Pattabiraman, Diwakar Bierie, Brian Mazutis, Linas Wolf, Guy Krishnaswamy, Smita Pe'er, Dana eng R01 CA164729/CA/NCI NIH HHS/ P30 DK045735/DK/NIDDK NIH HHS/ T32 HD007149/HD/NICHD NIH HHS/ P30 CA008748/CA/NCI NIH HHS/ DP1 HD084071/HD/NICHD NIH HHS/ U19 AI089992/AI/NIAID NIH HHS/ U54 CA209975/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2018/07/03 Cell. 2018 Jul 26;174(3):716-729.e27. doi: 10.1016/j.cell.2018.05.061. Epub 2018 Jun 28.
- [138] Matthieu Viry. jenkspy.
- [139] T. Moerman, S. Aibar Santos, C. Bravo Gonzalez-Blas, J. Simm, Y. Moreau, J. Aerts, and S. Aerts. Grnboost2 and arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12):2159–2161, 2019. Moerman, Thomas Aibar Santos, Sara Bravo Gonzalez-Blas, Carmen Simm, Jaak Moreau, Yves Aerts, Jan Aerts, Stein eng Research Support, Non-U.S. Gov't England 2018/11/18 Bioinformatics. 2019 Jun 1;35(12):2159-2161. doi: 10.1093/bioinformatics/bty916.

- [140] B. Van de Sande, C. Flerin, K. Davie, M. De Waegeneer, G. Hulselmans, S. Aibar, R. Seurinck, W. Saelens, R. Cannoodt, Q. Rouchon, T. Verbeiren, D. De Maeyer, J. Reumers, Y. Saeys, and S. Aerts. A scalable scenic workflow for single-cell gene regulatory network analysis. *Nat Protoc*, 15(7):2247–2276, 2020. Van de Sande, Bram Flerin, Christopher Davie, Kristofer De Waegeneer, Maxime Hulselmans, Gert Aibar, Sara Seurinck, Ruth Saelens, Wouter Cannoodt, Robrecht Rouchon, Quentin Verbeiren, Toni De Maeyer, Dries Reumers, Joke Saeys, Yvan Aerts, Stein eng Research Support, Non-U.S. Gov't England 2020/06/21 Nat Protoc. 2020 Jul;15(7):2247-2276. doi: 10.1038/s41596-020-0336-2. Epub 2020 Jun 19.
- [141] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P. L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader. Integration of biological networks and gene expression data using cytoscape. *Nat Protoc*, 2(10):2366–82, 2007. Cline, Melissa S Smoot, Michael Cerami, Ethan Kuchinsky, Allan Landys, Nerius Workman, Chris Christmas, Rowan Avila-Campilo, Iliana Creech, Michael Gross, Benjamin Hanspers, Kristina Isserlin, Ruth Kelley, Ryan Killcoyne, Sarah Lotia, Samad Maere, Steven Morris, John Ono, Keiichiro Pavlovic, Vuk Pico, Alexander R Vailaya, Aditya Wang, Peng-Liang Adler, Annette Conklin, Bruce R Hood, Leroy Kuiper, Martin Sander, Chris Schmulevich, Ilya Schwikowski, Benno Warner, Guy J Ideker, Trey Bader, Gary D eng R01 GM070743-03/GM/NIGMS NIH HHS/ R01 GM070743-04/GM/NIGMS NIH HHS/ R01 GM070743-03S1/GM/NIGMS NIH HHS/ R01 GM070743-05/GM/NIGMS NIH HHS/ R01 GM070743-01/GM/NIGMS NIH HHS/ R01 GM070743/GM/NIGMS NIH HHS/ GM070743-01/GM/NIGMS NIH HHS/ R01 GM070743-02/GM/NIGMS NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't England 2007/10/20 Nat Protoc. 2007;2(10):2366-82. doi: 10.1038/nprot.2007.324.
- [142] A. M. Wepler, L. Da Meda, I. Pires da Silva, W. Xu, G. Grignani, A. M. Menzies, M. S. Carlino, G. V. Long, S. N. Lo, I. Nordman, C. B. Steer, M. Lyle, C. Trojaniello, P. A. Ascierto, C. Lebbe, and S. Sandhu. Durability of response to immune checkpoint inhibitors in metastatic merkel cell carcinoma after treatment cessation. *Eur J Cancer*, 183:109–118, 2023. Wepler, Alison M Da Meda, Laetitia Pires da Silva, Ines Xu, Wen Grignani, Giovanni Menzies, Alexander M Carlino, Matteo S Long, Georgina V Lo, Serigne N Nordman, Ina Steer, Christopher B Lyle, Megan Trojaniello, Claudia Ascierto, Paolo A Lebbe, Celeste Sandhu, Shahneen eng England 2023/02/27 Eur J Cancer. 2023 Apr;183:109-118. doi: 10.1016/j.ejca.2023.01.016. Epub 2023 Jan 30.
- [143] W. Ma, R. Xue, Z. Zhu, H. Farrukh, W. Song, T. Li, L. Zheng, and C. X. Pan. Increasing cure rates of solid tumors by immune checkpoint inhibitors. *Exp Hematol Oncol*, 12(1):10, 2023. Ma, Weijie Xue, Ruobing Zhu, Zheng Farrukh, Hizra Song, Wenru Li, Tianhong Zheng, Lei Pan, Chong-Xian eng P01 CA247886/CA/NCI NIH HHS/ R01 CA197296/CA/NCI NIH HHS/ Review England 2023/01/17 Exp Hematol Oncol. 2023 Jan 16;12(1):10. doi: 10.1186/s40164-023-00372-8.

- [144] D. Sun, X. Guan, A. E. Moran, L. Y. Wu, D. Z. Qian, P. Schedin, M. S. Dai, A. V. Danilov, J. J. Alumkal, A. C. Adey, P. T. Spellman, and Z. Xia. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol*, 40(4):527–538, 2022. Sun, Duanchen Guan, Xiangnan Moran, Amy E Wu, Ling-Yun Qian, David Z Schedin, Pepper Dai, Mu-Shui Danilov, Alexey V Alumkal, Joshi J Adey, Andrew C Spellman, Paul T Xia, Zheng eng R01 CA251245/CA/NCI NIH HHS/ P50 CA097186/CA/NCI NIH HHS/ R01 CA244576/CA/NCI NIH HHS/ P50 CA186786/CA/NCI NIH HHS/ R35 GM124704/GM/NIGMS NIH HHS/ R01 CA207377/CA/NCI NIH HHS/ R21 HL145426/HL/NHLBI NIH HHS/ K01 LM012877/LM/NLM NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2021/11/13 *Nat Biotechnol*. 2022 Apr;40(4):527-538. doi: 10.1038/s41587-021-01091-3. Epub 2021 Nov 11.
- [145] A. Ravi, M. D. Hellmann, M. B. Arniella, M. Holton, S. S. Freeman, V. Naranbhai, C. Stewart, I. Leshchiner, J. Kim, Y. Akiyama, A. T. Griffin, N. I. Vokes, M. Sakhi, V. Kamesan, H. Rizvi, B. Ricciuti, P. M. Forde, V. Anagnostou, J. W. Riess, D. L. Gibbons, N. A. Pennell, V. Velcheti, S. R. Digumarthy, M. Mino-Kenudson, A. Califano, J. V. Heymach, R. S. Herbst, J. R. Brahmer, K. A. Schalper, V. E. Velculescu, B. S. Henick, N. Rizvi, P. A. Janne, M. M. Awad, A. Chow, B. D. Greenbaum, M. Luksza, A. T. Shaw, J. Wolchok, N. Hacohen, G. Getz, and J. F. Gainor. Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer. *Nat Genet*, 55(5):807–819, 2023. Ravi, Arvind Hellmann, Matthew D Arniella, Monica B Holton, Mark Freeman, Samuel S Naranbhai, Vivek Stewart, Chip Leshchiner, Ignaty Kim, Jaegil Akiyama, Yo Griffin, Aaron T Vokes, Natalie I Sakhi, Mustafa Kamesan, Vashine Rizvi, Hira Ricciuti, Biagio Forde, Patrick M Anagnostou, Valsamo Riess, Jonathan W Gibbons, Don L Pennell, Nathan A Velcheti, Vamsidhar Digumarthy, Subba R Mino-Kenudson, Mari Califano, Andrea Heymach, John V Herbst, Roy S Brahmer, Julie R Schalper, Kurt A Velculescu, Victor E Henick, Brian S Rizvi, Naiyer Janne, Pasi A Awad, Mark M Chow, Andrew Greenbaum, Benjamin D Luksza, Marta Shaw, Alice T Wolchok, Jedd Hacohen, Nir Getz, Gad Gainor, Justin F eng P30 CA093373/CA/NCI NIH HHS/ U01 CA272610/CA/NCI NIH HHS/ UL1 TR001863/TR/NCATS NIH HHS/ F30 CA257765/CA/NCI NIH HHS/ S10 OD021764/OD/NIH HHS/ T32 CA009172/CA/NCI NIH HHS/ P30 CA008748/CA/NCI NIH HHS/ R01 CA208756/CA/NCI NIH HHS/ R01 CA240317/CA/NCI NIH HHS/ R35 CA197745/CA/NCI NIH HHS/ K12 CA138464/CA/NCI NIH HHS/ U01 CA217858/CA/NCI NIH HHS/ T32 HG002295/HG/NHGRI NIH HHS/ S10 OD012351/OD/NIH HHS/ T32 GM007367/GM/NIGMS NIH HHS/ K08 CA248723/CA/NCI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2023/04/07 *Nat Genet*. 2023 May;55(5):807-819. doi: 10.1038/s41588-023-01355-5. Epub 2023 Apr 6.
- [146] Yuying Liu, Nannan Zhou, Li Zhou, Jing Wang, Yabo Zhou, Tianzhen Zhang, Yi Fang, Jinwei Deng, Yunfeng Gao, Xiaoyu Liang, Jiadi Lv, Zhenfeng Wang, Jing Xie, Yuanbo Xue, Huafeng Zhang, Jingwei Ma, Ke Tang, Yiliang Fang, Feiran Cheng, Chengjuan Zhang, Bing Dong, Yuzhou Zhao, Peng Yuan, Quanli Gao, Haizeng Zhang, F. Xiao-Feng Qin, and Bo Huang. Il-2 regulates tumor-reactive cd8+ t cell exhaustion by activating the aryl hydrocarbon receptor. *Nature Immunology*, 22(3):358–369, 2021.

- [147] J. W. Cho, M. H. Hong, S. J. Ha, Y. J. Kim, B. C. Cho, I. Lee, and H. R. Kim. Genome-wide identification of differentially methylated promoters and enhancers associated with response to anti-pd-1 therapy in non-small cell lung cancer. *Exp Mol Med*, 52(9):1550–1563, 2020. Cho, Jae-Won Hong, Min Hee Ha, Sang-Jun Kim, Young-Joon Cho, Byoung Chul Lee, Insuk Kim, Hye Ryun eng 2018M3C9A5064709/National Research Foundation of Korea (NRF)/ 2018R1A5A2025079/National Research Foundation of Korea (NRF)/ 2016M3C9A4922809/National Research Foundation of Korea (NRF)/ 2017M3A9E9072669/National Research Foundation of Korea (NRF)/ 2017M3A9E8029717/National Research Foundation of Korea (NRF)/ 6-2018-0072/Yonsei University | Yonsei University College of Dentistry (YUCD)/ Research Support, Non-U.S. Gov't 2020/09/04 *Exp Mol Med*. 2020 Sep;52(9):1550-1563. doi: 10.1038/s12276-020-00493-8. Epub 2020 Sep 2.
- [148] S. Litiere, S. Collette, E. G. de Vries, L. Seymour, and J. Bogaerts. Recist - learning from the past to build the future. *Nat Rev Clin Oncol*, 14(3):187–192, 2017. Litiere, Saskia Collette, Sandra de Vries, Elisabeth G E Seymour, Lesley Bogaerts, Jan eng Research Support, Non-U.S. Gov't Review England 2016/12/21 *Nat Rev Clin Oncol*. 2017 Mar;14(3):187-192. doi: 10.1038/nrclinonc.2016.195. Epub 2016 Dec 20.
- [149] N. Riaz, J. J. Havel, V. Makarov, A. Desrichard, W. J. Urba, J. S. Sims, F. S. Hodi, S. Martin-Algarra, R. Mandal, W. H. Sharfman, S. Bhatia, W. J. Hwu, T. F. Gajewski, Jr. Slingluff, C. L., D. Chowell, S. M. Kendall, H. Chang, R. Shah, F. Kuo, L. G. T. Morris, J. W. Sidhom, J. P. Schneck, C. E. Horak, N. Weinhold, and T. A. Chan. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell*, 171(4):934–949 e16, 2017. Riaz, Nadeem Havel, Jonathan J Makarov, Vladimir Desrichard, Alexis Urba, Walter J Sims, Jennifer S Hodi, F Stephen Martin-Algarra, Salvador Mandal, Rajarsi Sharfman, William H Bhatia, Shailender Hwu, Wen-Jen Gajewski, Thomas F Slingluff, Craig L Jr Chowell, Diego Kendall, Sviatoslav M Chang, Han Shah, Rachna Kuo, Fengshen Morris, Luc G T Sidhom, John-William Schneck, Jonathan P Horak, Christine E Weinhold, Nils Chan, Timothy A eng T32 CA160001/CA/NCI NIH HHS/ R01 CA205426/CA/NCI NIH HHS/ R01 CA177828/CA/NCI NIH HHS/ K08 DE024774/DE/NIDCR NIH HHS/ P30 CA008748/CA/NCI NIH HHS/ 2017/10/17 *Cell*. 2017 Nov 2;171(4):934-949.e16. doi: 10.1016/j.cell.2017.09.028. Epub 2017 Oct 12.
- [150] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. R. Loh, and S. Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods*, 16(12):1289–1296, 2019. Korsunsky, Ilya Millard, Nghia Fan, Jean Slowikowski, Kamil Zhang, Fan Wei, Kevin Baglaenko, Yuriy Brenner, Michael Loh, Po-Ru Raychaudhuri, Soumya eng U19 AI111224/AI/NIAID NIH HHS/ T32 HG002295/HG/NHGRI NIH HHS/ T32 AR007530/AR/NIAMS NIH HHS/ K00 CA222750/CA/NCI NIH HHS/ R01 AR063759/AR/NIAMS NIH HHS/ U01 HG009088/HG/NHGRI NIH HHS/ U01 HG009379/HG/NHGRI NIH HHS/ UH2 AR067677/AR/NIAMS NIH HHS/ Research Support, N.I.H., Extramural 2019/11/20 *Nat Methods*. 2019 Dec;16(12):1289-1296. doi: 10.1038/s41592-019-0619-0. Epub 2019 Nov 18.

- [151] K. H. Prazanowska and S. B. Lim. An integrated single-cell transcriptomic dataset for non-small cell lung cancer. *Sci Data*, 10(1):167, 2023. Prazanowska, Karolina Hanna Lim, Su Bin eng 2020R1A6A1A03043539/National Research Foundation of Korea (NRF)/ 2020M3A9D8037604/National Research Foundation of Korea (NRF)/ 2022R1C1C1004756/National Research Foundation of Korea (NRF)/ HR22C1734/Ministry of Health and Welfare (Ministry of Health, Welfare and Family Affairs)/ Dataset England 2023/03/28 Sci Data. 2023 Mar 27;10(1):167. doi: 10.1038/s41597-023-02074-6.
- [152] Y. Chen, A. T. Lun, and G. K. Smyth. From reads to genes to pathways: differential expression analysis of rna-seq experiments using rsubread and the edger quasi-likelihood pipeline. *F1000Res*, 5:1438, 2016. Chen, Yunshun Lun, Aaron T L Smyth, Gordon K eng England 2016/08/11 F1000Res. 2016 Jun 20;5:1438. doi: 10.12688/f1000research.8987.2. eCollection 2016.
- [153] D. J. McCarthy, Y. Chen, and G. K. Smyth. Differential expression analysis of multifactor rna-seq experiments with respect to biological variation. *Nucleic Acids Res*, 40(10):4288–97, 2012. McCarthy, Davis J Chen, Yunshun Smyth, Gordon K eng Research Support, Non-U.S. Gov't England 2012/01/31 Nucleic Acids Res. 2012 May;40(10):4288-97. doi: 10.1093/nar/gks042. Epub 2012 Jan 28.
- [154] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–40, 2010. Robinson, Mark D McCarthy, Davis J Smyth, Gordon K eng Research Support, Non-U.S. Gov't England 2009/11/17 Bioinformatics. 2010 Jan 1;26(1):139-40. doi: 10.1093/bioinformatics/btp616. Epub 2009 Nov 11.
- [155] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res*, 43(7):e47, 2015. Ritchie, Matthew E Phipson, Belinda Wu, Di Hu, Yifang Law, Charity W Shi, Wei Smyth, Gordon K eng Research Support, Non-U.S. Gov't England 2015/01/22 Nucleic Acids Res. 2015 Apr 20;43(7):e47. doi: 10.1093/nar/gkv007. Epub 2015 Jan 20.
- [156] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol*, 33(5):495–502, 2015. Satija, Rahul Farrell, Jeffrey A Gennert, David Schier, Alexander F Regev, Aviv eng F32 HD075541/HD/NICHD NIH HHS/ U01 MH105960/MH/NIMH NIH HHS/ P50 HG006193/HG/NHGRI NIH HHS/ R01 GM056211/GM/NIGMS NIH HHS/ HHMI/Howard Hughes Medical Institute/ 1P50HG006193/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2015/04/14 Nat Biotechnol. 2015 May;33(5):495-502. doi: 10.1038/nbt.3192. Epub 2015 Apr 13.
- [157] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*, 36(5):411–420, 2018. Butler, Andrew Hoffman, Paul Smibert, Peter Papalexi, Efthymia Satija, Rahul eng DP2 HG009623/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. 2018/04/03 Nat Biotechnol. 2018 Jun;36(5):411-420. doi: 10.1038/nbt.4096. Epub 2018 Apr 2.



- [158] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, 3rd Mauck, W. M., Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902 e21, 2019. Stuart, Tim Butler, Andrew Hoffman, Paul Hafemeister, Christoph Papalexi, Efthymia Mauck, William M 3rd Hao, Yuhan Stoeckius, Marlon Smibert, Peter Satija, Rahul eng F32 HD075541/HD/NICHD NIH HHS/ R01 MH071679/MH/NIMH NIH HHS/ R01 HD096770/HD/NICHD NIH HHS/ DP2 HG009623/HG/NHGRI NIH HHS/ OT2 OD026673/OD/NIH HHS/ R21 HG009748/HG/NHGRI NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. 2019/06/11 Cell. 2019 Jun 13;177(7):1888-1902.e21. doi: 10.1016/j.cell.2019.05.031. Epub 2019 Jun 6.
- [159] Y. Hao, S. Hao, E. Andersen-Nissen, 3rd Mauck, W. M., S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, and R. Satija. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587 e29, 2021. Hao, Yuhan Hao, Stephanie Andersen-Nissen, Erica Mauck, William M 3rd Zheng, Shiwei Butler, Andrew Lee, Maddie J Wilk, Aaron J Darby, Charlotte Zager, Michael Hoffman, Paul Stoeckius, Marlon Papalexi, Efthymia Mimitou, Eleni P Jain, Jaison Srivastava, Avi Stuart, Tim Fleming, Lamar M Yeung, Bertrand Rogers, Angela J McElrath, Juliana M Blish, Catherine A Gottardo, Raphael Smibert, Peter Satija, Rahul eng U19 AI089986/AI/NIAID NIH HHS/ RM1 HG011014/HG/NHGRI NIH HHS/ DP2 HG009623/HG/NHGRI NIH HHS/ UM1 AI068614/AI/NIAID NIH HHS/ UM1 AI068618/AI/NIAID NIH HHS/ OT2 OD026673/OD/NIH HHS/ R21 HG009748/HG/NHGRI NIH HHS/ UM1 AI068635/AI/NIAID NIH HHS/ U19 AI128914/AI/NIAID NIH HHS/ U01 AI068618/AI/NIAID NIH HHS/ Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2021/06/02 Cell. 2021 Jun 24;184(13):3573-3587.e29. doi: 10.1016/j.cell.2021.04.048. Epub 2021 May 31.
- [160] Y. Hao, T. Stuart, M. H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, A. Srivastava, G. Molla, S. Madad, C. Fernandez-Granda, and R. Satija. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*, 42(2):293–304, 2024. Hao, Yuhan Stuart, Tim Kowalski, Madeline H Choudhary, Saket Hoffman, Paul Hartman, Austin Srivastava, Avi Molla, Gesmira Madad, Shaista Fernandez-Granda, Carlos Satija, Rahul eng OT2 OD033760/OD/NIH HHS/ K99 CA267677/CA/NCI NIH HHS/ K99 HG011489/HG/NHGRI NIH HHS/ RM1 HG011014/HG/NHGRI NIH HHS/ R01 HD096770/HD/NICHD NIH HHS/ DP2 HG009623/HG/NHGRI NIH HHS/ R35 NS097404/NS/NINDS NIH HHS/ OT2 OD026673/OD/NIH HHS/ 2023/05/26 Nat Biotechnol. 2024 Feb;42(2):293-304. doi: 10.1038/s41587-023-01767-y. Epub 2023 May 25.

[161] B. A. Cordier. Ohsu latex dissertation template, 2021. Available on Overleaf.

The following figures were created using *BioRender*: Figures 2, 3, 4, 12.

The following Overleaf template was used: Oregon Health & Science University (OHSU) dissertation template [161].

This work was supported by the following NIH NLM training grant: T15-LM007088