



Longitudinal and Multi-Omics Analysis of Metastatic Melanoma Reveals Dynamic Tumor Evolution and Multi-Level Heterogeneity

Citation

Bao, Yourong. 2026. Longitudinal and Multi-Omics Analysis of Metastatic Melanoma Reveals Dynamic Tumor Evolution and Multi-Level Heterogeneity. Masters Thesis, Harvard Medical School.

Link

<https://dash.harvard.edu/handle/1/42738360>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

**Longitudinal and Multi-Omics Analysis of Metastatic Melanoma Reveals Dynamic Tumor
Evolution and Multi-Level Heterogeneity**

Yourong Bao

A Thesis Submitted to the Faculty of

The Harvard Medical School

in Partial Fulfillment of the Requirements

for the Degree of Master of Medical Sciences in Biomedical Informatics

Harvard University

Boston, Massachusetts.

May, 2026

Longitudinal and Multi-Omics Analysis of Metastatic Melanoma Reveals Dynamic Tumor Evolution and Multi-Level Heterogeneity

Abstract

Melanoma is the most lethal cutaneous cancer. Advanced-stage (Stage III-IV) melanoma has a 10-year survival rate of less than 30%. Despite recent advances in targeted and immunotherapies, clinical outcomes of melanoma patients remain highly variable, largely due to its highly heterogeneous nature. This heterogeneity is observed in 3 levels: intratumoral, intertumoral, and inter-patient, complicating clinicians' efforts to identify consistent biomarkers and therapeutic targets. Therefore, traditional approaches that focus on individual genes or pathways are insufficient to capture the complexity of metastatic melanoma progression and treatment response. To address this limitation, integrative multi-omics strategies are required to systematically characterize tumor evolution and tumor-immune interactions across different biological scales.

In this study, we analyzed two complementary melanoma patient cohorts. A longitudinal in-transit metastasis (ITM) cohort with whole-exome sequencing (WES) and bulk RNA-seq data was used to reconstruct phylogenetic trees and investigate tumor evolutionary dynamics over time. A second cohort from the Human Tumor Atlas Network (HTAN), comprising WES, single-nucleus RNA sequencing (snRNA-seq), spatial transcriptomics (MERFISH), and spatial proteomics data, enables the characterization of tumor cell states and immune interactions across defined clinical response groups. We integrated genomic, transcriptomic, and spatial-omics analyses to identify recurrent patterns of tumor progression, including clinically relevant clonal expansion, biological or functional pathways, mutational signatures, and melanoma state transitions.

Our results reveal that melanoma progression is characterized by dynamic clonal evolution that is associated with therapeutic interventions. During melanoma progression, we observed a decreased genomic heterogeneity but increased phenotypic and mutational process diversities over time. Non-progressor tumors exhibited enhanced antigen presentation and interferon signaling, highlighting the importance of tumor-immune interactions in treatment outcomes. We also identified biologically interpretable tumor programs that capture coherent tumor-intrinsic states beyond single-gene signals, which can be validated across modalities and applied

clinically. Together, this study demonstrates that melanoma progression is driven by coordinated genomic evolution, phenotypic plasticity, and tumor-immune interactions, which underscores the value of multi-modal integration for understanding complex cancers.

Table of Contents

Abstract	ii
Table of Contents	iv
List of Figures	vi
Acknowledgement	vii
CHAPTER 1	1
CHAPTER 2	6
2.1 Study Cohort and Data Collection	6
2.1.1 The ITM Patient Cohort	6
2.1.2 The HTAN Patient Cohort	7
2.2 Whole Exome Sequencing Analysis and Phylogenetic Reconstruction	7
2.2.1 WES Data Processing and Quality Control	7
2.2.2 Variant Calling and Annotation	8
2.2.3 Tumor Purity and Ploidy Estimation and Cross-sample Merging	9
2.3 Single-Patient Phylogenetic Tree Reconstruction	9
2.3.1 Clonal Decomposition	9
2.3.2 Phylogenetic Tree Reconstruction, Refinement, and Validation	10
2.3.3 Single Base Substitutions Mutational Signature Analysis	11
2.4 Transcriptomic Analysis	11
2.4.1 RNA-seq Quality Control and Preprocessing	11
2.4.2 Bulk RNA-seq Data Cell Type Deconvolution	12
2.4.3 Gene Set Enrichment Analysis	12
2.4.4 Differential Expression Analysis	12
2.5 Tumor Program Identification	13
2.5.1 NMF-Based Tumor Programs	13
2.5.2 Signature-Based Tumor Program Discovery	13
2.5.3 Cluster Annotation and Program Consolidation	14
2.5.4 Tumor Program Marker Gene Identification and Validation	14
CHAPTER 3	15
3.1 Evolutionary dynamics of three tumor lineages in a melanoma progressor patient reveals temporal relationship between tumor clones' survival and treatment	15
3.2 Distinct genomic features and mutational processes characterize brain metastasis-associated lineage in a melanoma progressor	19

3.3 Longitudinal tumor analysis reveals decreased genomic heterogeneity and increased mutational process heterogeneity over time	24
CHAPTER 4	28
4.1 Melanoma cell state transitions from invasive to proliferative over time within the tumor evolutionary lineages	28
4.2 Transcriptional plasticity and increased phenotypic heterogeneity facilitate prolonged tumor survival	31
4.3 Immune infiltration levels of macrophages and dendritic cells weakly associated with specific melanoma cell states	36
CHAPTER 5	39
5.1 HTAN patient cohort enables multi-modal analysis of tumor-immune interactions across clinical response groups	39
5.2 Antigen presentation and interferon-related pathways are enriched in non-progressors	40
5.3 NMF-defined tumor programs successfully identify melanocytic and proliferative tumor programs but remain ambiguous otherwise	43
5.4 Defining novel biologically relevant tumor programs based on recurrent pathway and signature activity	45
5.5 Preliminary spatial mapping of novel tumor programs reveals context-dependent tumor-immune associations	50
CHAPTER 6	53
6.1 Summary of key findings	53
6.2 Melanoma progression and intratumoral heterogeneity in the scope of tumor evolution	54
6.3 Intertumoral heterogeneity across modalities reveals tumor evolution dynamics and environmental-driven plasticity	55
6.4 Inter-patient tumor evolution heterogeneity in terms of phylogenetic tree structural variations	56
6.5 Genomic features of melanoma brain metastasis and limitations of SBS mutational signature interpretation	58
6.6 Additional limitations of the study and future steps	59
References	61
Supplementary Material	69
Supplementary Figures	69
Supplementary Tables	83

List of Figures

Figure 1 A-B. Reconstructed phylogenetic tree for patient 118. *Page 17.*

Figure 1 C-D. Reconstructed phylogenetic tree for patient 135. *Page 21.*

Figure 2. Mutational signature compositions and comut plot. *Page 26.*

Figure 3. Global and longitudinal pattern of melanoma signature and tumor state transitions. *Page 30.*

Figure 4. Enrichment of selected melanoma signatures across clinical and biological sample groups. *Page 32-33.*

Figure 5. Spearman correlation between immune cell infiltration level and melanoma signature enrichment scores.
Page 37.

Figure 6. Enrichment of melanoma signatures and human hallmark pathways across clinical response groups. *Page 41.*

Figure 7. Characterization of NMF-defined tumor programs. *Page 44.*

Figure 8. Identification and annotation of novel tumor programs. *Page 48-50.*

Acknowledgement

I would like to express my deepest gratitude to my principal investigator, Dr. David Liu, for overseeing this project and for his invaluable guidance throughout the process. His high-level insight, patience, and continuous support were instrumental in shaping this work. I am especially grateful for his thoughtful feedback and revision suggestions, which greatly strengthened both the scientific direction and presentation of this thesis.

I am sincerely thankful to my postdoctoral mentor, Dr. Giuseppe Tarantino, for his constant support and mentorship. His patience, availability, and willingness to provide guidance on a day-to-day basis were crucial to my growth. I deeply appreciate his help in both the technical and scientific aspects of this project.

I would also like to thank Dr. Tuulia Vallius for her guidance when training me for spatial-omics analysis.

I feel incredibly fortunate to have been part of the Liu lab as I was surrounded by a group of amazing colleagues and researchers. I would like to thank all members of the lab for their daily support and encouragement. The engaging discussions, technical assistance, and shared curiosity created an inspiring environment that made this experience both productive and enjoyable. These everyday interactions have played a significant role in my growth.

I would like to acknowledge the important contributions of previous lab members, including Priyanka Solanky, Xinyu Cui, and Yiwen He, whose foundational work and carefully curated datasets made this project possible. I am grateful for the opportunity to build upon their efforts.

I am deeply grateful to our clinical collaborators and, most importantly, to the patients who contributed to the datasets used in this study. Without their participation and generosity, this work would not have been possible.

Finally, I would like to thank my family and friends for their unwavering support and encouragement. Their understanding and emotional support have been essential in helping me navigate and complete my master's journey.

This thesis includes figures and tables from a work that has been previously submitted for publication:

- Tarantino, G., Zaremba, A., Vallius, T., Woodnorth, M., Hua, Y., Pelletier, R., Lopez Leon, M., Shi, Y., Maliga, Z., Makhzami, S., Aprati, T. J., Karlaš, B., Glutsch, V., Schilling, B., Hassel, J., Berking, C., Utikal, J., Meier, F., Meiss, F., ... Schadendorf, D. (2025). *Multiomic Profiling of a Unique In-Transit Melanoma Cohort Identifies Melanoma Differentiation as Predictor of Tumor Progression and Therapy Response*. <https://doi.org/10.1101/2025.10.01.679282>

CHAPTER 1

Introduction

AI Usage Statement: Portions of this chapter were edited with the assistance of ChatGPT for grammar and clarity.

All scientific content, analysis, and interpretations are the author's own.

Melanoma has the highest death rate for cutaneous cancers due to its aggressive and heterogeneous nature¹, which means it can progress quickly and metastasize across multiple organ systems². Stage III melanoma patients can develop in-transit metastases, where tumors spread intralymphatically and are typically found between the primary tumor site and regional lymph nodes³⁻⁴. Stage IV melanoma exhibits distant metastases with tumors already spread to distant organs such as the lungs, liver, and brain⁴⁻⁵. Current treatments for advanced melanoma include immune single agents and combinations of checkpoint inhibitors (ICIs) such as anti-CTLA-4 (ipilimumab)⁶, anti-PD-1 (nivolumab)⁶⁻⁸, and anti-LAG3⁸⁻⁹, BRAF and MEK inhibitors targeting BRAF-mutant melanoma¹⁰, and combination therapies¹¹. However, the effectiveness of these treatments varies widely among patients due to differences in their response to therapy, tumor metastatic rates, and overall tumor aggressiveness. While some tumors remain localized for extended periods, others rapidly spread to distant sites⁵. Likewise, some patients show long-term remission with immunotherapy or targeted therapy, while others develop resistance, leading to disease progression despite initial responses¹²⁻¹⁴.

The variability observed clinically reflects the interpatient heterogeneity of melanoma, while tumors also exhibit intertumoral and intratumoral heterogeneities, both of which play critical roles in tumor evolution. Tumor evolution is a sequential process. Acquired mutations due to genetic instability in tumor cells accumulate in the clonal expansions of the primary tumor, shaping them into new tumor lineages over time. Multiple divergent lineages can co-exist and co-evolve simultaneously but may respond differently to innate immune responses and clinical interventions due to intertumoral heterogeneity^{2,15}. Within each tumor clone, either primary or metastatic, defined by a driver clonal mutation like BRAF, there can be subpopulations of cells harboring additional mutations not shared by all the tumor cells. These subclonal mutations contribute to intratumoral heterogeneity and influence

tumor progression^{2,16}. As tumor clones metastasize to different sites, intrinsic heterogeneities drive subclones to further divide, with some forming new downstream lineages while others are gradually eliminated by selective pressure¹⁷.

Tumor evolution is also a dynamic, Darwinian-like process^{2,17}. Tumor lineages and their subclones undergo stepwise selection by the innate and adaptive immune responses¹⁸⁻¹⁹, where only those that develop immune-resistant and immune-privileged characteristics survive²⁰. Studies have shown that the interactions between immune cells and tumor cells shape the characteristics of tumor clones at specific metastatic sites¹⁸⁻²¹. Intratumoral heterogeneity fuels this Darwinian-like selection, enabling tumor adaptation through various immune escape mechanisms^{15,21}. In many Stage III and Stage IV cancers, including melanoma, tumors evade immune responses²²⁻²³ by exploiting immune checkpoint mechanisms to suppress T-cell activity²⁴. The most common pathways involve tumor cells leveraging immune checkpoint proteins on T cells, such as Cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and Programmed Cell Death Protein 1 (PD-1)²⁵⁻²⁷. When these checkpoint proteins bind to their specific ligands, they act as “brakes” on T cell function, preventing an immune attack on tumor cells²⁶⁻²⁷. This immune suppression allows tumors to persist, proliferate, and metastasize to nearby tissues or distant organs²².

Due to these immune-resistant properties, immune checkpoint blockade (ICB) has emerged as a promising therapeutic strategy for advanced melanoma²⁸. ICB therapy applies anti-CTLA-4, anti-PD-1, or anti-LAG-3 checkpoint inhibition^{8,29-31}, preventing them from blocking immune checkpoint proteins, thereby restoring T-cell activity and enhancing anti-tumor immunity²⁸. Nevertheless, despite initial positive responses to ICB, most stage III-IV melanoma patients eventually develop resistance, leading to tumor recurrence and disease progression^{7,12-13}. Studies have identified several mechanisms associated with ICB resistance³², including PTEN loss³³⁻³⁴ and impaired interferon- γ signaling³⁵⁻³⁶. However, researchers emphasize the need for longitudinal samples to gain a more comprehensive understanding of tumor evolutionary dynamics under prolonged long-term immune pressure^{3,19-20}.

Longitudinal analyses of patient data and tumor samples have provided deeper insights into tumor evolution and changes in tumoral heterogeneity over time. Constructing phylogenetic maps of multiple tumor clones within a single patient can reveal the co-evolution of distinct tumor lineages, illustrating the interplay between cancer and immune pathways¹⁸. Longitudinal studies incorporating multimodal data can also establish connections between clinical interventions, phenotypic responses, and molecular mechanisms of immune resistance. Compared to cross-sectional studies using tumor samples from different patients, longitudinal tumor samples from the same

patient serve as internal controls, reducing the confounding effects of intratumoral heterogeneity²¹. Previous longitudinal studies used sequential tumor biopsies to link treatment responses with tumor heterogeneity³⁷⁻³⁸. However, many are limited by incomplete sampling across treatment stages due to collection constraints and relying on a single data type.

In 2021, Liu et al. conducted a longitudinal study to thoroughly examine the mechanisms of melanoma resistance to ICB³⁹. This study analyzed an extensive collection of 37 tumor samples from a single patient over 9 years, covering all key progression and treatment time points. This phylogenetic study reconstructed the tumor's evolutionary trajectory, identifying site-specific alterations and tumor-immune interactions driving resistance. Specifically, it revealed the association between the de-differentiation of neural-crest (NGFR^{hi}) tumor cells and immunotherapy resistance with different spatial patterns at distinct tumor sites. Still, these findings remain limited to a single patient with an ancestral tumor lineage characterized by chromosome 15q loss who received sequential ICB treatments, so validation in a broader melanoma cohort is necessary.

Additionally, many longitudinal analyses assume monophyletic tumor samples and fail to investigate the intratumoral heterogeneity arising from potential subclonal tumors. Recent studies suggest that metastatic tumors can originate from diverse seeding patterns, including consecutive and polyclonal seeding⁴⁰. Consecutive seeding describes a stepwise process where a metastatic site is seeded multiple times by different tumor cells over time⁴⁰. Polyclonal seeding occurs when a metastatic site arises from multiple distinct tumors subclones or a cluster of tumor cells that co-migrate simultaneously⁴¹⁻⁴². These new models, in contrast to the classic monoclonal seeding, where all secondary tumors originate directly from the primary tumor⁴³, explain the rise of intratumoral heterogeneity that is often associated with immunotherapy resistance in metastatic tumors.

While genomic analyses provide a framework for reconstructing tumor evolutionary trajectories, they do not fully capture the functional diversity of tumor cells. In melanoma, phenotypic plasticity, or where tumor cells transition between distinct transcriptional states without substantial genetic change, has been recognized to be associated with disease progression and therapeutic resistance⁴⁴⁻⁴⁷. For example, multiple studies have suggested that the AXL-high mesenchymal-like melanoma cell state is highly invasive and is associated with immunotherapy resistance⁴⁶⁻⁴⁹. To better characterize these dynamic cell states, inter-tumoral heterogeneity can be examined using multiple methods and categorized into tumor programs through shared gene signatures.

Tumor programs are recurring tumor-intrinsic signatures shared across patients and even cohorts, commonly defined by transcriptional states derived from gene expression data. Classical methods of defining tumor programs often include unsupervised machine learning approaches such as non-negative matrix factorization (NMF), which deconvolutes complex tumor gene expression data into interpretable metaprograms. However, these data-driven approaches may not always yield biologically interpretable programs, nor could they be easily validated across datasets.

An alternative method is to define tumor programs based on recurrent pathway activity and curated gene signatures, like the previously defined melanoma signatures⁴⁵⁻⁶¹ and human hallmark gene sets⁶². Such a more supervised approach enables a more interpretable characterization of functional states grounded in prior biological knowledge and avoids random noise or dominating expression of non-coding genes intrinsic to each gene expression data set. While it is lacking in the current melanoma research, novel approaches that integrate genomic, transcriptomic, spatial-omics, and pathway signatures to define biologically robust and cross-validated tumor programs are an urgent next step to address intra- and inter-tumoral heterogeneity by capturing patterns of gene activity associated with phenotypes such as proliferation, invasion, immune-interacting, or stress-responding. This will strengthen our understanding of how tumor functional states evolve over time and under therapeutic pressure.

To address the gaps summarized above, this thesis work integrates two complementary melanoma patient cohorts spanning multiple data modalities, including genomics, transcriptomics, spatial transcriptomics, and spatial proteomics, to systematically characterize tumor evolution, heterogeneity, and tumor-immune interactions. We define three primary aims. First, we reconstruct phylogenetic maps from sequential tumor biopsies of patients with in-transit metastasis to trace the evolutionary trajectories of melanoma and identify key genomic changes associated with clonal selection and metastatic progression. Second, we characterize tumor cell states and transcriptional programs underlying melanoma progression, focusing on how cell state transitions and phenotypic heterogeneity evolve over time within individual patients. Lastly, we perform integrative analyses across multi-patient cohorts and spatially resolved datasets to investigate tumor-immune interactions and identify biologically interpretable tumor programs associated with clinical response. Overall, this study provides a multi-scale view of melanoma progression, linking genomic evolution with phenotypic plasticity and immune context. These findings have important clinical implications, including improving the understanding of treatment response variability, identifying

tumor-intrinsic programs associated with resistance, and informing the development of more effective therapeutic strategies for melanoma patients.

CHAPTER 2

Materials and Methods

AI Usage Statement: Portions of this chapter were edited with the assistance of ChatGPT for grammar and clarity.

All scientific content, analysis, and interpretations are the author's own.

2.1 Study Cohort and Data Collection

This study incorporates two independent melanoma patient cohorts. The first cohort focuses on melanoma patients who developed in-transit metastasis (ITM) at any stage of the disease and is referred to as the ITM patient cohort. The second cohort was generated as part of the Human Tumor Atlas Network (HTAN)⁶³⁻⁶⁴ and is referred to as the HTAN patient cohort. This cohort includes multi-omics data from melanoma patients with diverse demographic backgrounds, disease subtypes, tumor progression states, treatment regimens, and clinical responses.

2.1.1 The ITM Patient Cohort

The ITM patient cohort consists of 41 patients with unresectable melanoma, categorized into progressors (n = 22) and non-progressors (n = 19) based on their clinical outcomes (*Supplementary Table 1*). Patient metadata includes sex, age at diagnosis, primary tumor location, histological subtype, Breslow thickness, ulceration status, and treatment history. All data were de-identified prior to analysis.

Among these, five patients have longitudinal tumor samples with matched whole-exome sequencing (WES) and bulk RNA sequencing (RNA-seq) data. Tissue cyclic immunofluorescence (t-CyCIF) data are available upon request for validation purposes. Two patients (Patients 136 and 222) were previously analyzed by my post-doc mentor, Giuseppe Tarantino and were included in a submitted manuscript⁶⁵. This work focuses on two of the remaining three patients with longitudinal data (Patients 118 and 135). Both selected patients are male progressors with primary tumors located in the lower extremities. Tumor biopsies were collected across 14 and 21 time points, respectively, spanning multiple stages of disease progression and treatment. DNA and RNA extraction, library

preparation, and WES were performed as previously described^{39, 66-67} by a collaborating laboratory at the University Hospital Essen, Germany.

2.1.2 The HTAN Patient Cohort

The HTAN melanoma cohort is a large-scale, multi-omics dataset designed to characterize melanoma at single-cell resolution. Available data modalities include single-nucleus RNA sequencing (snRNA-seq), WES, spatial transcriptomics (MERFISH), and spatial proteomics (mIF, CODEX, and Lunaphore). This cohort enables comprehensive investigation of the tumor microenvironment (TME) and tumor-immune microenvironment (TIME), with the goal of understanding mechanisms underlying immune response and treatment resistance.

The cohort comprises 67 tumor samples from melanoma patients with diverse tumor subtypes, treatment regimens, and clinical outcomes (*Supplementary Table 2*). The majority of patients received either a single anti-PD1 therapy (64%) or a combined ICB (27%). Most samples were collected in a metastatic treatment setting (73%), compared to adjuvant (24%) or neoadjuvant (3%) settings, reflecting a cohort enriched for advanced-stage, surgically unresectable melanoma.

Clinical responses were categorized into four groups: durable clinical benefit (DCB), intermediate non-progressor (IMnotP), intermediate progressor (IMP), and rapid progressor (RP), based on each patient's progression-free survival (PFS) time. DCB patients are defined by PFS > 1095 days (3 years), while RP patients are defined by PFS < 180 days (6 months). Intermediate patients fall between these thresholds and are further classified as IMP or IMnotP based on the presence or absence of disease progression following treatment. In total, the cohort includes 10 DCB, 12 IMnotP, 23 IMP, and 22 RP tumor samples.

2.2 Whole Exome Sequencing Analysis and Phylogenetic Reconstruction

2.2.1 WES Data Processing and Quality Control

WES data were generated at the Broad Institute and obtained as tumor-normal matched BAM files aligned to the hg19 reference genome, accessed through dbGaP under controlled access. Data preprocessing was performed using an adapted version of the Getz Lab CGA WES Characterization pipeline

(https://github.com/broadinstitute/CGA_Production_Analysis_Pipeline) developed at the Broad Institute to ensure

standardized quality control and downstream variant analysis. All preprocessing and some downstream analyses were conducted on the Broad Institute's Terra cloud platform using high-memory compute nodes (≥ 64 GB RAM) with multi-core processing (≥ 8 CPUs per job).

Quality control procedures were applied to ensure sample integrity and sequencing reliability. Tumor-normal sample pairing and potential lane swaps were assessed using CopyNumberQC and Picard Cross Check Lane Fingerprints⁶⁸, with a likelihood threshold >15 indicating correct sample matching. Cross-sample contamination was evaluated using ContEst⁶⁹, with samples required to have an estimated contamination fraction <0.01 . Tumor-in-Normal (TiN) contamination was assessed and corrected using DeTiN⁷⁰, with TiN <0.3 indicating effective somatic single nucleotide variants (sSNVs) rescuing. Other sequencing quality metrics, including coverage and uniformity, were evaluated using Picard Multiple Metrics⁷¹⁻⁷². Samples were required to meet thresholds for median target coverage and the proportion of targeted bases sequenced at a depth of at least 100X. All samples included in downstream mutation detection and phylogenetic analysis passed these quality control criteria.

2.2.2 Variant Calling and Annotation

The Getz Lab CGA WES Characterization pipeline was also utilized to call somatic mutations and copy number alterations (CNAs) from tumor-normal matched WES data. Single-nucleotide variants (SNVs) and small insertions and deletions (indels) were detected using MuTect^{68, 73} and Strelka⁷⁴, respectively, with variants called relative to matched germline samples obtained at treatment initiation. CNAs were inferred using CapSeg (<https://github.com/aaronmck/CapSeg>), which segments the genome based on read depth and allelic imbalance. Both total and allelic copy number segments, including focal amplifications and deletions, were derived using the AllelicCapSeg framework^{68, 75}

All detected variants were functionally annotated using Oncotator⁷⁶, a functional annotator available on GATK at the Broad Institute^{68, 77}, incorporating gene-level and variant-level annotations. Additional annotations were integrated from external resources, including OncoKB Annotator v3.4⁷⁸ for clinically relevant and oncogenic variants, AlphaMissense⁷⁹ for pathogenicity predictions of missense mutations, and cBioPortal⁸⁰⁻⁸² for looking up the exact location of the variant in the genome as well as cross-referencing recurrent alterations observed in the TCGA 2015 Melanomas database⁸³.

2.2.3 Tumor Purity and Ploidy Estimation and Cross-sample Merging

Tumor purity and ploidy were estimated using ABSOLUTE⁸⁴ based on integrated SNV, indel, and CNA data. Tumor purity, defined as the proportion of cancerous cells within a sample, was used to adjust variant allele frequencies (VAFs) for downstream analyses. Cancer cell fraction (CCF) values were inferred for each mutation following purity and ploidy correction and were subsequently used for clonal decomposition. The median allelic copy number across the genome was calculated from purity- and ploidy-corrected segments and used to quantify genomic instability via the Genome Instability Index⁸⁵.

To enable cross-sample comparison within individual patients, mutations and CVAs were harmonized across longitudinal samples using the mergedMAF function (vanallenlab/MergeMAFs/1). This step ensured consistent mutation representation and facilitated downstream phylogenetic reconstruction. ABSOLUTE-derived solutions were manually reviewed post-merging for consistency, and finalized purity, ploidy, and CCF estimates were extracted for subsequent analyses.

2.3 Single-Patient Phylogenetic Tree Reconstruction

2.3.1 Clonal Decomposition

Clonal decomposition was performed to infer tumor subpopulations and reconstruct evolutionary trajectories. Somatic mutations, including SNVs, indels, and CNAs, were clustered using PyClone-VI⁸⁶, a Bayesian clustering framework that estimates CCF for each mutation based on tumor purity and allelic copy number information. Default parameters were applied as previously described in Liu et al.³⁹. Compared to the original PyClone implementation⁸⁹, PyClone-VI utilizes variational inference in place of the Markov Chain Monte Carlo (MCMC) sampling for improved computational efficiency while maintaining comparable accuracy.

The resulting mutational clusters were interpreted as distinct tumor clones or subclones. Clusters were subsequently filtered based on mutation count and biological relevance. Small clusters ($n < 5$ mutations) were retained if they contained known driver mutations or oncogenic alterations, as annotated by Oncotator or OncoKB, or if they were associated with copy number events contributing to lineage definition. Clusters were further categorized based on their CCF distribution across samples, including ancestral clusters (clonal across all samples), lineage-specific clusters (clonal in a subset of samples), and private clusters (restricted to a single sample). Clusters

consistent with low-confidence signals, such as those with near-zero CCF across all samples or interpreted as passenger subclones without any biological relevance, were excluded.

To infer relationships between tumor samples, hierarchical clustering was performed on mutational cluster CCF profiles using Euclidean distance. This clustering method is under the assumption that tumors with shared clusters of alterations arise from shared ancestry. The resulting heatmaps were used to identify shared and sample-specific clusters, enabling grouping of samples into putative evolutionary lineages. Clusters present across all samples were considered ancestral, while the acquisition of additional clusters in later samples was used to infer tumor evolution over time.

2.3.2 Phylogenetic Tree Reconstruction, Refinement, and Validation

Phylogenetic trees were reconstructed using a stepwise approach integrating mutational clustering and copy number information. Initial tree structures were defined based on PyClone-derived mutational clusters only, which serves as the backbone for lineage relationships.

Copy number alterations were subsequently incorporated to refine phylogenetic relationships. CNA profiles were examined across samples to identify patterns such as loss of heterozygosity (LOH) that were consistent with or supportive of PyClone-inferred lineages. In cases where mutations within a cluster exhibited reduced CCF in specific samples, corresponding CNA segments were examined to assess whether allelic imbalance or deletion events explained the observed patterns. CNA information was used to support and annotate lineage relationships but did not override the primary structure defined by mutational clustering.

Phylogenetic reconstruction was further validated using PhylogicNDT⁹⁰, which applies a multidimensional Dirichlet process to model subclonal architecture and identify additional private or subclonal events. Concordance between PyClone-based and PhylogicNDT-based results was used to assess robustness of inferred evolutionary relationships.

Finally, clinical metadata, including biopsy timing, anatomical location, treatment history, and clinical response following each treatment like progressive disease (PD) and partial response (PR), were overlaid onto the phylogenetic trees to provide clinical context for tumor evolution.

2.3.3 Single Base Substitutions Mutational Signature Analysis

Single base substitution (SBS) mutational signature analysis was performed using Mutation Annotation Format (MAF) files generated prior to cross-sample merging to preserve sample-specific mutation profiles and avoid artificial inflation of shared variants across longitudinal samples.

Mutational signatures were extracted using the deconstructSigs R package⁹¹ which estimates the weight of predefined reference signatures in each sample. Reference signatures were obtained from the COSMIC SBS mutational signatures v2 database⁹², comprising 30 curated SBS signatures either named or unknown. To account for differences between whole-exome and whole-genome contexts, normalization for trinucleotide frequencies was performed using the “exome2genome” option within deconstructSigs, enabling appropriate comparison of the WES patient data to the genome-based COSMIC reference signatures.

Signature analysis was conducted at both the sample level and the mutational cluster level. At the sample level, signature weights reflect the overall mutational processes active within each tumor sample. At the cluster level, signatures were computed using mutations assigned to each PyClone-derived cluster, providing insight into subclonal mutational processes. Results from clusters with fewer than 50 mutations were retained but interpreted with caution due to increased susceptibility to sampling bias.

2.4 Transcriptomic Analysis

2.4.1 RNA-seq Quality Control and Preprocessing

Bulk RNA-seq data from the ITM cohort were subjected to quality control using the MultiQC pipeline⁹³, which aggregates sequencing quality metrics across samples including sequence count, sequence duplication level, STAR alignment score, FastQC quality score, per base sequence content, overrepresented sequence percentage, and adaptor content. Samples meeting standard quality thresholds were retained for downstream analysis (*Supplementary Figure 1*).

The HTAN snRNA-seq data preprocessing and quality control were performed previously by collaborating lab members and were not repeated in this study. Cell type annotation was also performed previously by Yiwen He, Xinyu Cui, and Priyanka Solanky using canonical marker genes and refined with expert input from domain specialists. Annotated immune cell subtypes were used for downstream analyses in this study.

2.4.2 Bulk RNA-seq Data Cell Type Deconvolution

Due to cost and clinical restrictions, high-resolution single-cell RNA-seq (scRNA-seq) data are commonly unavailable for cancer patients, so the deconvolution of bulk RNA-seq data is crucial to evaluate the transcriptomic aspect of metastasis progression and intratumoral heterogeneity. InstaPrism⁹⁴, was used to estimate tumor cell fractions within each sample. It is adapted from the traditional deconvolution method, BayesPrism⁹⁵, which is a Bayesian probabilistic model priming on scRNA-seq data across various tumor types. InstaPrism replaces the time-consuming Gibbs sampling step in BayesPrism with a rapid fixed-point algorithm to increase debulking efficiency. Immune cell composition and infiltration levels were quantified using the online tools of CIBERSORT⁹⁶, which infers the relative abundance of immune cell populations based on reference gene expression profiles.

2.4.3 Gene Set Enrichment Analysis

Gene set enrichment analysis was performed using single-sample Gene Set Enrichment Analysis (ssGSEA)⁹⁷ to quantify pathway and signature activity at the sample and cell level. Gene sets used in this work include Human Hallmark pathways from the Molecular Signatures Database (MSigDB)⁶² and melanoma signatures. We pulled a total of 48 melanoma signature gene sets from 17 published literatures⁴⁵⁻⁶¹ that previously identified melanoma tumor-intrinsic pathways or cell states cross-referenced in large patient datasets (*Supplementary Table 3*). Custom gene sets, including those representing tumor programs and antigen presentation pathways, for example, pre-proliferative tumor program marker genes and MHC-related genes, were also evaluated. Enrichment scores were computed in transcriptomics modalities, across bulk RNA-seq, snRNA-seq, and MERFISH data. Pairwise comparisons of enrichment scores of selected gene sets between sample groups were performed using the Wilcoxon rank-sum test, and nominal p-values were used to assess statistical significance with a threshold of 0.05.

2.4.4 Differential Expression Analysis

Differential expression analysis was performed on snRNA-seq data using a pseudo-bulk approach, in which gene expression counts were aggregated at the sample level. Differential expression testing was conducted using DESeq2⁹⁸. Genes or gene sets are considered differentially expressed if they meet the thresholds of p-value < 0.05

and absolute log2 fold change ($|\log_2\text{fc}| > 1$). Differential expression analysis was applied to identify changes in gene expression, as well as differences in pathway and signature activity across conditions, including the previously defined human hallmark pathways and melanoma signatures.

2.5 Tumor Program Identification

2.5.1 NMF-Based Tumor Programs

Previously defined tumor transcriptional programs were derived from snRNA-seq data using non-negative matrix factorization (NMF) by a prior member of the lab (Priyanka Solanky). In this approach, the gene expression matrix V was decomposed into two non-negative matrices W and H , such that $V \approx WH$. The matrix W represents gene weights for each program (number of programs predefined in function parameter, $k = 6$), while H captures the activity of each program across individual cells, enabling the assignment and quantification of tumor programs at single-cell resolution.

2.5.2 Signature-Based Tumor Program Discovery

Tumor programs were independently identified using a signature-based clustering approach. Single-cell enrichment scores were computed for both human hallmark pathways and melanoma signatures from Section 2.4.3 and used as input features for clustering. Clustering was performed using the Leiden algorithm with multiple resolutions evaluated to identify biologically meaningful groupings while avoiding clusters dominated by single samples. Final resolutions of 0.03 for hallmark-based features and 0.06 for melanoma signature-based features were selected, each yielding eight clusters.

Clusters were filtered and refined by examining the expression of canonical melanoma marker genes (*MLANA*, *PMEL*, *TYR*, *MITF*, *SOX10*, *DCT*) and tumor-associated pathways to exclude non-tumor or artifactual clusters. Differentially expressed genes (DEGs) for each cluster were identified using DESeq2, with thresholds of adjusted p-value < 0.05 and absolute log2 fold change > 0.25 . To assess similarity between clusters, the Jaccard similarity index was computed based on DEG overlap. Clusters with high similarity were merged to define broader transcriptional programs.

2.5.3 Cluster Annotation and Program Consolidation

Resulting clusters were manually annotated based on enrichment patterns of human hallmark pathways and melanoma signatures, as well as the expression of key marker genes associated with known melanoma cell states. These include melanocytic markers (*MITF*, *PMEL*, *MLANA*, *SOX10*), neural crest-like markers (*NGFR*), and dedifferentiated state markers (*AXL*, *SOX9*, *EGFR*, *ZEB1*). Sequencing depth was also evaluated to ensure that identified clusters were not driven by technical artifacts.

To define robust tumor programs, results from hallmark-based and melanoma signature-based clustering were compared. Overlap of the top 30 DEGs per cluster, as well as shared pathway enrichment and cell state characteristics, were used to identify consistent programs across both approaches. Clusters with concordant features were consolidated into unified tumor programs.

2.5.4 Tumor Program Marker Gene Identification and Validation

For each consolidated tumor program, the top 100 DEGs were identified, and mutually exclusive gene sets were defined as candidate marker genes. These gene sets were validated by mapping back to both human hallmark-based and melanoma signature-based cluster assignments to ensure consistency of program definition. To enable spatial validation, marker genes were further restricted to those included in the MERFISH gene panel. These filtered markers will be used to identify and validate tumor programs in spatial transcriptomic data in future steps.

CHAPTER 3

Phylogenetic reconstruction of single-patient tumor samples reveals branched evolution and dynamic genomic changes during melanoma progression

AI Usage Statement: Portions of this chapter were edited with the assistance of ChatGPT for grammar and clarity.

All scientific content, analysis, and interpretations are the author's own.

3.1 Evolutionary dynamics of three tumor lineages in a melanoma progressor patient reveals temporal relationship between tumor clones' survival and treatment

The availability of longitudinal samples spanning multiple treatment regimens and disease stages in the ITM patient cohort provides a unique opportunity to study tumor evolution across both local and metastatic progression. We therefore conducted a longitudinal analysis of clinical metadata and whole-exome sequencing (WES) data from patient 118 ($n = 14$) and patient 135 ($n = 21$) to reconstruct phylogenetic trees and characterize their tumor evolutionary trajectories.

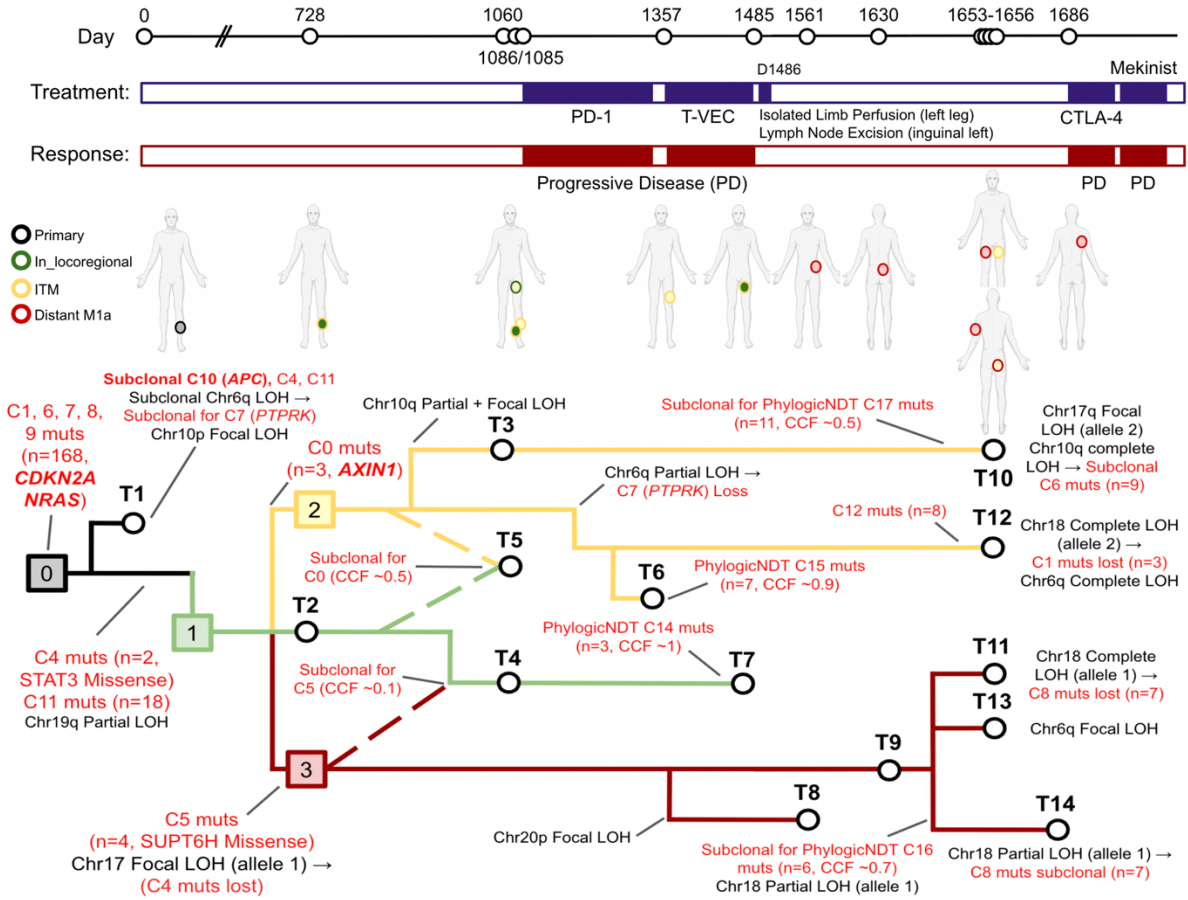
To infer evolutionary relationships across samples, we leveraged mutational clusters defined using PyClone-VI and validated with PhylogicNDT, which estimated cancer cell fractions (CCFs) to identify clonal and subclonal populations (*Supplementary Figure 2A–B, E–F*). Hierarchical clustering of these mutational clusters based on CCF profiles was then used to group samples into putative evolutionary lineages (*Supplementary Figure 2C, G*), following the assumption that tumors with shared clusters of alterations arise from shared ancestry. For example, in patient 118, PyClone Cluster 9 is simply an ancestral mutational cluster because these mutations are shared (CCF ~ 1) in all samples (*Supplementary Figure 2C*). Copy number variation (CNV) profiling were also incorporated to refine lineage relationships (*Supplementary Figure 2D–H*), identifying copy number alteration (CNA) events that may contributed to the loss of certain mutational clusters in some samples due to deletions or loss of heterozygosity (LOH), like PyClone Clusters 1, 6, and 8 that are also ancestral clusters shared among all samples (*Supplementary*

Figure 2C-D). In addition, CNVs also help infer direct ancestry. In patient 118, for example, the primary tumor T1 is not a direct ancestor of all following tumors samples because it contains private clonal events and LOH events not found in later samples (*Figure 1A, Supplementary Figure 2D*). More details on inferring phylogenetic relationships see *Methods 2.3*.

We first reconstructed the phylogenetic map of melanoma progression for patient 118, with clinical metadata overlaid (*Figure 1A–B*). Patient 118 is a slow progressor from the ITM melanoma cohort who developed stage IV disease over 4 years. This patient harbors an oncogenic mutation in *NRAS* and received anti-PD-1 systemic therapy 3 years after the initial diagnosis. The treatment lasted for over 2 months but was terminated due to progressive disease. Following progression to stage IV melanoma, the patient received anti-CTLA-4 blockade and Mekinist treatment, but both resulted in progressive disease, ultimately leading to patient death.

The phylogenetic tree captures 3 major lineages from an ancestral tumor clone (black lines, denoted as L0), which contains 168 mutations, including oncogenic driver mutations in genes *CDKN2A* and *NRAS* (*Figure 1A*). The primary tumor belonging to L0 was sampled from the lower left leg at diagnosis and is named T1. However, T1 is not the direct ancestor of all downstream tumors sampled because it is uniquely subclonal for the PyClone Cluster 10 mutations (n=88), including an oncogenic mutation in *APC*, and has a focal loss of heterozygosity (LOH) on Chr10p not found in later samples. Nevertheless, T1 already contains subclonal tumor cells that represent the later clones. It is subclonal for PyClone Cluster 4 (n=2) and Cluster 11 (n=18) mutations, which became clonal in Lineage 1 tumors and are shared by all downstream samples, indicating that the seeds of later dominant lineages were already present early at diagnosis. Similarly, a subpopulation of T1 cells have LOH at Chr6q, which leads to loss of PyClone Cluster 7 mutation *PTPRK*, which corresponds to a later CNA event in T6 and T12 in Lineage 2.

A



B

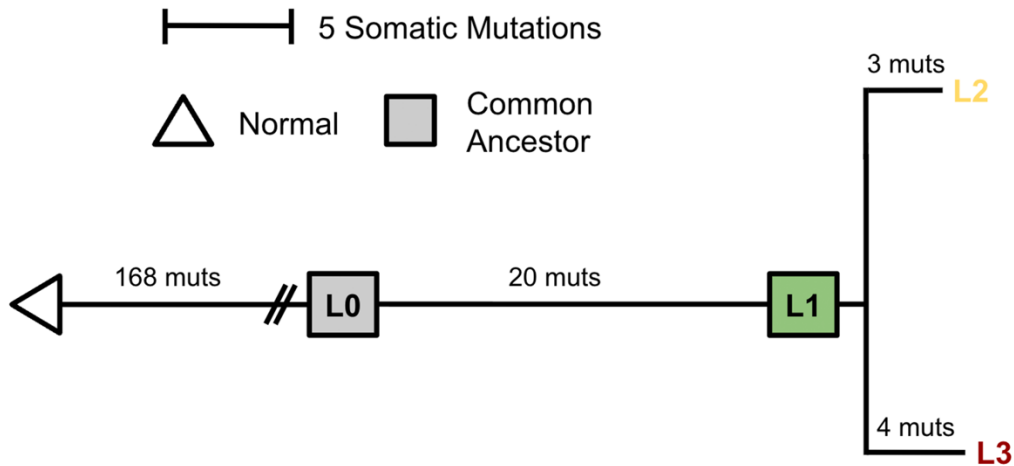


Figure 1A-B. Reconstructed phylogenetic tree for patient 118. (A) Reconstructed phylogenetic tree with clinical metadata (clinical timeline, treatment history, patient response, and tumor biopsy locations) overlaid. Line colors represent tumor lineages. Mutational and copy number alteration (CNA) events are labeled in red and black text, respectively. **(B)** Lineage phylogenetic tree for patient 118 with distance corresponding to the number of mutations between lineages. Scale provided.

Lineage 1 (green lines, denoted as L1) is defined by 2 mutational clusters, PyClone Cluster 4 and Cluster 11, becoming clonal (*Figure 1A*). This lineage can be described as ITM-driven (all 3 samples are ITM melanoma), early tumor clones that gave rise to the later metastasis but became extinct due to local treatment. Lineage 2 (yellow lines, denoted as L2) branched off from L1 by gaining additional PyClone Cluster 0 mutations (n=3), including a mutation in the oncogene *AXIN1*. It contains 4 samples (T3, T6, T10, and T12), varied in time, location, treatment group, and contains both ITM and distant metastatic samples. This lineage is the most mixed in terms of tumor type and metastatic location but is persistent over all clinical intervention. Interestingly, T5, the only locoregional lymph node tumor biopsy from patient 118, is subclonal for the L2-defining PyClone Cluster 0 with a CCF of 0.5, marking its shared membership in both L1 and L2.

Lineage 3 (red lines, denoted as L3), similar to L2, branched off from L1 early in time. It gained PyClone Cluster 5 mutations (n=4) while lost PyClone Cluster 4 mutations due to a focal LOH event on chr17q allele 1. It is a late, all-distant-metastasis tumor lineage that was able to spread globally, potentially linked to the tumor's ability to evade immune surveillance and resist systematic treatments. Worth noticing that T4 and T7 from L1 are both subclonal for PyClone Cluster 5 mutations with a very low CCF around 0.1, suggesting either a shared subclonal mutational profile between independent branches or an alternative metastatic pattern where L3 samples were partially seeded by late L1 tumors.

What really distinguishes L3 from L2 is its late appearance after major clinical events and the all-distant-metastasis component of this tumor clone. Despite its molecular divergence early from L1, L3 first emerged phenotypically post-treatment at day 1561. More importantly, all L3 tumors were biopsied following a series of treatments. Between 1086 and 1357 days post-diagnosis, patient 118 received anti-PD-1 blockade treatment but had progressive disease with T6 emerged at the left leg site. Following the immunotherapy, patient 118 received a T-VEC therapy between 1357 and 1485 days post-diagnosis which also led to progressive disease with T7 appeared at the inguinal left side. In response to T7 metastasis, local treatments were administered on day 1486, including isolated limb perfusion (ILP) at the primary tumor site (lower left leg) and lymph node excision at inguinal left to remove T7. Despite patient 118 continuing to show progressive disease clinically, the L1 tumor line disappeared completely after the excision of inguinal T7 tumor. Coinciding with L1 extinction, the first L3 tumor and also the first distant metastatic tumor appeared on day 1561 at the left flank, and this distant metastatic lineage continued to expand and spread to multiple body parts including upper left arm and right scapula. The emergence of L3 distant

metastases marked progressive disease after T-VEC and local treatments, as well as patient 118's progression to Stage IV (Distant M1a) melanoma.

The series of clinical interventions in patient 118 may have acted as selective pressures shaping tumor lineage dynamics. Notably, the emergence of L3 coincided with the disappearance of L1 following surgical excision and prior systemic therapies, suggesting a shift in dominant tumor populations. In contrast, L2 tumors persisted and reappeared at later timepoints, indicating sustained survival across treatment conditions. Between 1653 and 1686 days post-diagnosis, both L2 and L3 tumors were observed across multiple anatomical sites, including both in-transit and distant metastases (*Figure 1A, T10–T14*). T10, an L2 tumor, represents the only post-treatment ITM sample and reappeared at the site of prior lymph node excision (inguinal left). This recurrence at a previously treated location may suggest site-specific factors or intrinsic properties of L2 tumors that support continued survival following local intervention.

Although the mutational clusters defining L2 and L3 are relatively small ($n = 3$ and $n = 4$, respectively) (*Figure 1B*), these lineages, particularly at later timepoints, exhibit increased CNA events and the presence of private subclonal populations compared to early L1 tumors (*Figure 1A*). This pattern suggests an accumulation of genomic alterations over time that may be associated with tumor progression. From an evolutionary perspective, the three tumor lineages display distinct temporal and genomic characteristics: L1 represents an early expanding lineage that is no longer observed following treatment, L2 is a mixed-type clone persisted across multiple timepoints and clinical contexts, and L3 emerges later and is primarily associated with distant metastatic disease that is resistant to immunotherapy. Together, these observations indicate that melanoma progression in patient 118 involves early diversification followed by the emergence and persistence of distinct lineages with differing clinical and genomic features.

3.2 Distinct genomic features and mutational processes characterize brain metastasis-associated lineage in a melanoma progressor

Like patient 118, patient 135 is also a slow progressor from the ITM melanoma cohort who ended up with stage IV disease that led to his decease 10 years after the initial diagnosis. This patient harbors the canonical melanoma driver mutation *BRAF* V600E and thus was a candidate of receiving BRAF inhibitor (BRAFi) targeted

therapy. Patient 135 received an ILP 2226 days post-diagnosis resulting in partial response, followed by a 2-cycle anti-CTLA4 treatment through Ipilimumab 7 years post-diagnosis but terminated before showing any clinical benefit due to adverse events. After progression to Stage IV distant metastasis, patient 135 stayed on a combined BRAFi+MEKi targeted therapy with continued partial response until he deceased.

Using the same method, we reconstructed a phylogenetic tree for patient 135 (*Figure 1C-D*). Patient 135's ancestral tumor clone contains 130 mutations, including the melanoma driver mutation, *BRAF* V600E. Unlike patient 118, we observed during phylogenetic reconstruction that patient 135 tumor samples harbor a substantially higher total number of somatic mutations ($n = 6861$ vs $n = 2632$), despite comparable mutational burden at the individual sample level. Patient 135 also exhibits increased CNA events, including more frequent amplifications and deletions (*Supplementary Figure 2H*), as well as a whole genome doubling (WGD) event marking the T14–T16–T17 sub-branch under Lineage 2 (green lines, denoted as L2) (*Figure 1C*).

We found L2 in patient 118 relatively more comparable with L2 in patient 135. Both represent a persistent lineage spanning both early and late timepoints and are predominantly composed of ITM tumors, although patient 135 ITM samples are spatially constrained to the lower right leg. However, compared to patient 118, L2 in patient 135 exhibits greater subclonal complexity. Patient 135 has more private mutational clusters, such as PyClone Cluster 3 ($n = 66$) and Cluster 5 ($n = 78$) in T15 and T12, respectively, along with multiple subclonal sub-branches defined by clusters such as PyClone Cluster 8 ($n = 62$) and Cluster 13 ($n = 99$) (*Supplementary Figure 2E, G*). Notably, these private and subclonal clusters often contain many mutations, resulting in greater mutational distances between sub-branches than between major lineages. However, because subclonal mutations provide weaker phylogenetic signals than shared clonal mutations, these clusters define sub-groups rather than independent tumor lineages. Together, these features highlight increased genomic complexity and distinguish the phylogenetic landscape of patient 135 from that of patient 118.

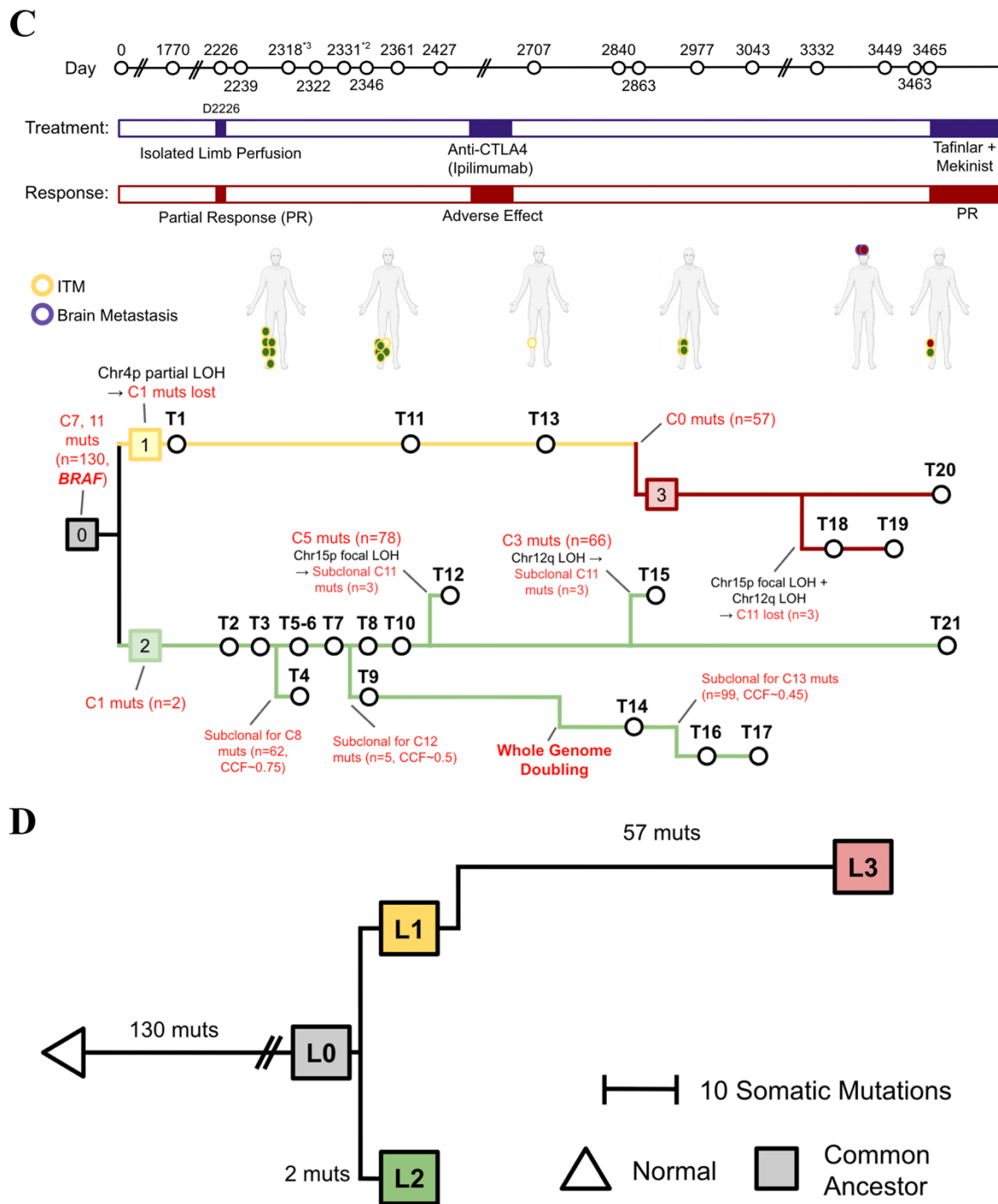


Figure 1C-B. Reconstructed phylogenetic tree for patient 135. (C) Reconstructed phylogenetic tree with clinical metadata (clinical timeline, treatment history, patient response, and tumor biopsy locations) overlaid. Line colors represent tumor lineages. Mutational and copy number alteration (CNA) events are labeled in red and black text, respectively. (D) Lineage phylogenetic tree for patient 135 with distance corresponding to the number of mutations between lineages. Scale provided.

In contrast to patient 118, where tumor lineages demonstrate clear evolutionary and phenotypic distinctions, lineage structure in patient 135 is less well aligned with metastatic phenotype. The majority of tumor biopsies in patient 135 are ITM lesions, and each lineage contains a mixture of dermal, cutaneous, subcutaneous, and amelanotic metastases, making it difficult to assign specific metastatic behaviors to individual lineages. This difference is also reflected at the clinical level. Patient 135 is characterized by predominantly ITM disease with only two biopsies of brain metastases, whereas patient 118 exhibits widespread distant metastases across multiple anatomical sites without brain involvement. This phenotypic difference is consistent with the phylogenetic structure of patient 135, where Lineage 1 (yellow lines, denoted as L1) appears relatively linear without any sub-branching until Lineage 3 (red lines, denoted as L3) branched off with additional PyClone Cluster 0 mutations ($n=57$), while L2 contains numerous sub-branches defined by either private or subclonal genomic events (*Figure 1C*).

Looking at patient 135's phylogenetic tree alone also reveals an interesting tumor evolutionary history. The separation between the two major lineages, L1 and L2, is defined by a partial LOH event on chr4p rather than by mutational clusters. This LOH event corresponds to the loss of PyClone Cluster 1 mutations ($n = 2$) in L1. However, due to the small size and lack of known oncogenic drivers in this cluster, the lineage distinction is more appropriately defined by the CNV event. Moreover, based on the available mutation and CNA data, the temporal order of these events cannot be resolved: whether the patient first gained Cluster 1 mutations in L2 but lost them due to LOH in L1, or the CNA event occurred in L1 prior to the gaining of Cluster 1 mutations in L2. Nevertheless, this chronological uncertainty does not affect the downstream comparisons.

Additional genomic observations further highlight the intertwined relationships between patient 135's tumor lineages. For example, T12 and T15 in L2 exhibit CNA events, focal LOH in chr15p and LOH of chr12q, respectively, that are also present in L3 brain metastasis samples (T18 and T19), although these occur on separate branches without clear ancestry (*Figure 1C*). The two CNA events in T12 and T15 separately caused the ancestral Cluster 11 to become subclonal in each sample, and the combined LOH of both chr15p focal location and chr12q result in complete loss of Cluster 11 in T18 and T19. This pattern may reflect parallel or convergent genomic evolution, but it may also contribute to the theory of alternative metastatic seeding, where a subpopulation of cells in L2 tumors partially give rise to the later L3 metastases. This theory involves additional modeling and exploration and can become an interesting next step, which will be elaborated later in the discussion.

Supporting the overall genomic complexity of L2, a WGD event occurred in the sub-branch under T9 after the anti-CTLA4 immunotherapy, affecting T14, T16, and T17, with additional subclonal mutations (PyClone Cluster 13) observed in T16 and T17 (*Figure 1C*). Around the same time, T15 emerged as a distinct sub-branch with a private mutational cluster PyClone Cluster 3 and the LOH event on chr12q leading to its subclonality for ancestral Cluster 11 mutations.

While L3 is the major lineage of interest by containing the only two distant brain metastases, its emergence is not directly associated with a specific clinical intervention, as its earliest samples appeared after day 3332 post-diagnosis (*Figure 3C*), temporally distant from any known treatment. Although L3 may have arisen earlier molecularly with the gaining of Cluster 1 mutations from L1, it is only observed phenotypically about two years after the anti-CTLA4 therapy, suggesting it may not be directly driven by immediate treatment pressure. Unlike patient 118, where L1 became extinct following the clinical treatment, no L1 samples are observed beyond T13 in patient 135, but this lineage may have contributed to or transitioned into the later-emerging L3 lineage rather than disappearing entirely.

L3 is characterized by substantial genomic divergence from its ancestor, L1, with a mutational distance of 57 mutations (*Figure 1D*), considerably greater than that observed between other lineages in both patients. This suggests that L3, and potentially brain metastases, are associated with increased genomic divergence from the ancestral lineage, rather than incremental accumulation of small mutational or CNA events.

The distinct genomic features of the L3 lineage are further supported by mutational signature analysis (*Figure 2A*). Single base substitution (SBS) mutational signatures represent characteristic patterns of nucleotide substitutions arising from specific endogenous and exogenous DNA damage and repair mechanisms, providing insight into the underlying mutagenic processes. Notably, a signature corresponding to polymerase η activity is observed exclusively in T18–T20, corresponding precisely to the L3 lineage (marked by the red frame in *Figure 2A*). This signature reflects a distinct mutational process that may be associated with context-specific DNA damage or microenvironmental influences. The restriction of this signature to L3 samples suggests that tumors associated with brain metastasis in patient 135 are characterized by unique mutational processes that differentiate them from other tumor samples.

3.3 Longitudinal tumor analysis reveals decreased genomic heterogeneity and increased mutational process heterogeneity over time

To assess the longitudinal heterogeneity of melanoma samples across disease progression, we first established a standardized metric for characterizing genomic profiles. In WES data with a large number of SNPs, SBS mutational signatures provide a systematic approach to infer underlying DNA damage processes and identify persistent patterns across samples and patients.

As described above, the three L3 tumor samples in patient 135 uniquely exhibit a somatic hypermutation signature. In addition to that, these samples share a common baseline mutational profile with many other samples, characterized by dominant clock-like (spontaneous deamination of 5-methylcytosine) and UV light exposure signatures (*Figure 2A*). A similar pattern is observed in patient 118, where most samples display a combination of clock-like and UV-associated signatures, with additional minor signatures appearing sporadically (*Figure 2B*). The prevalence of these two dominant signatures is biologically expected. Clock-like signatures reflect endogenous mutational processes associated with aging and are typically observed across many tissue types, independent of tumor-specific processes. In contrast, UV light exposure signatures are a hallmark of cutaneous melanoma and arise from ultraviolet-induced DNA damage, a well-established environmental risk factor for melanoma development. Together, these signatures form a stable mutational background across melanoma samples.

Notably, not all samples exhibit the clock-like signature. T13 in patient 135 and T14 in patient 118 both lack this baseline signature. T13, the final sample of Lineage 1 (L1) in patient 135 and collected shortly after the anti-CTLA4 therapy, instead shows a pronounced increase in the UV-associated signature, along with the presence of defective DNA mismatch repair and a chemotherapy-associated signature. The observed spike in UV-associated signature in T13 may reflect clonal selection or expansion of tumor cells with pre-existing UV-induced mutations, rather than new UV exposure, especially given the lack of temporal proximity to UV-related environmental changes. Alternatively, shifts in mutational signature proportions may result from changes in the relative abundance of subclones with distinct mutational profiles following treatment.

While T13 is not temporally associated with ILP treatment and earlier samples do not show this signature, the chemotherapy-associated signature may reflect limitations of the SBS mutational signature assignment and confusion with an alternative DNA damage process with similar substitution patterns. Similar inconsistencies are observed with the temozolomide (alkylating agent) signature, which appears sporadically in patient 118, including in

samples collected prior to treatment (*Figure 2B*), suggesting potential false-positive assignments in signature deconvolution.

Despite these limitations, several minor mutational signatures provide meaningful insights when interpreted in the phylogenetic context. Particularly, defective DNA repair-associated signatures appear in T12 of patient 118, which represents the only ITM tumor following the extinction of the L1 lineage. This signature is also observed in T11 and T13 of patient 135, both late-stage ITM tumors in L1 that precede either lineage disappearance or transition toward the more aggressive L3 lineage containing brain metastases. This pattern suggests that impaired DNA repair mechanisms may emerge during late stages of tumor evolution, potentially contributing to increased genomic instability and facilitating lineage transition or adaptation under selective pressure.

Although SBS mutational signature analysis has limitations, including incomplete annotation of signatures and occasional false-positive assignments, the overall mutational signature landscape is consistent across both patients. Early tumor samples are relatively homogeneous, dominated by the two canonical melanoma-associated signatures (clock-like and UV exposure), especially high in the clock-like signature. As tumors progress, metastasize, or diverge into distinct lineages, additional mutational signatures emerge at lower proportions, including those associated with defective DNA mismatch repair, homologous recombination deficiency, and APOBEC activity. These additional signatures alter the relative composition of the mutational profile, often reducing or eliminating the contribution of clock-like signatures, while the UV-associated signature remains consistently present across samples. This trend reflects an increase in the mutational process heterogeneity over time, where the underlying DNA damage and repair mechanisms become more complex, on top of a stable UV-driven mutational background.

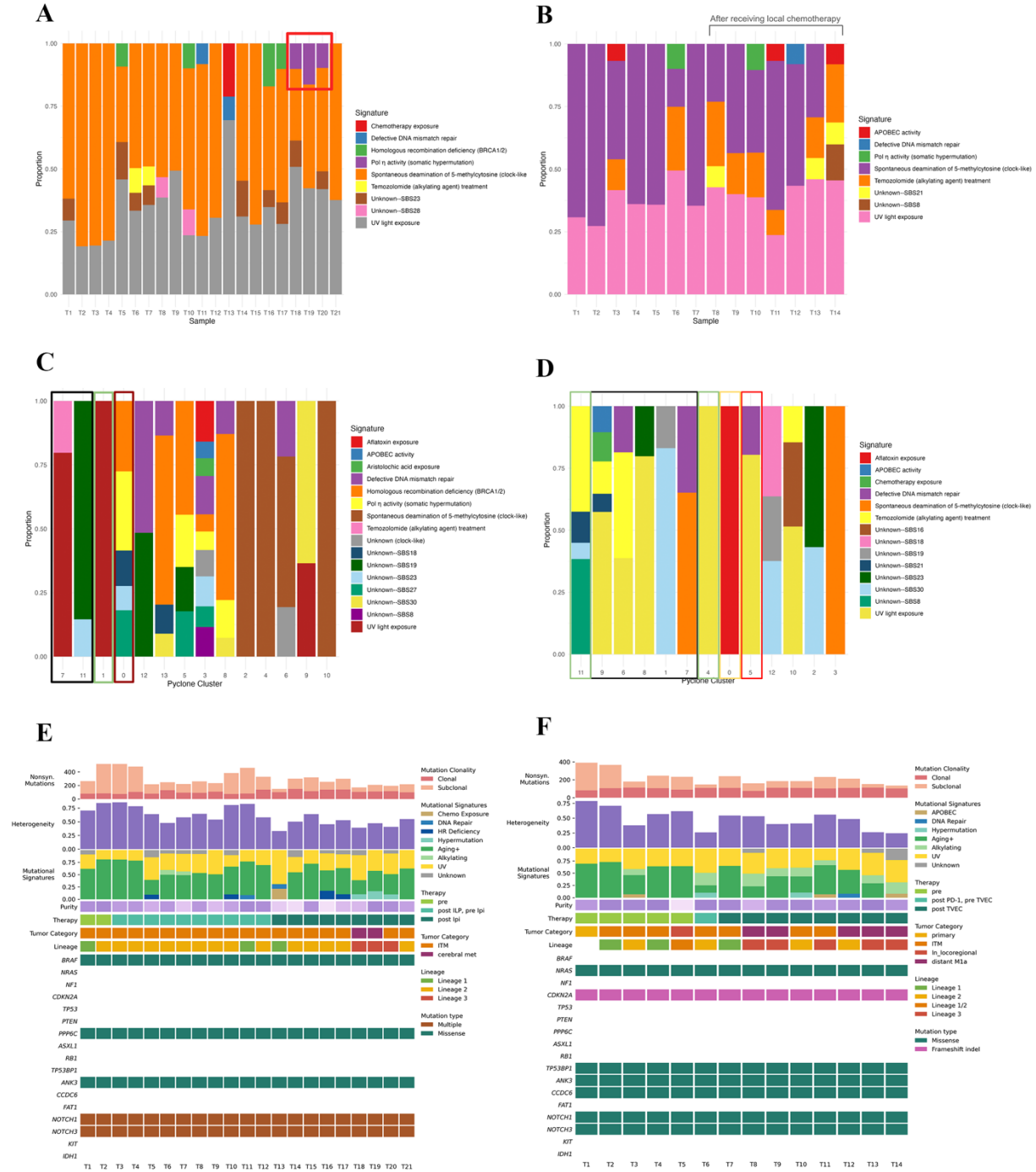


Figure 2. Mutational signature compositions and comutation plot. Stacked bar plots showing Single base substitutions (SBS) mutational signature compositions in (A) patient 135 and (B) patient 118 samples. SBS Mutational signature decomposition by PyClone mutational clusters in (C) patient 135 and (D) patient 118. Comutation plot summarizing genomic information for (E) patient 135 and (F) patient 118, including clonal/subclonal non-synonymous mutation proportions, genomic heterogeneity, mutational signatures, post-merging purity score, treatment status, tumor category, assigned lineage, and mutational profile of selected melanoma driver genes.

To further characterize tumor-clone-specific mutational processes, we calculated mutational signature proportions at the level of PyClone mutational clusters (*Figure 2C–D*). As expected, PyClone clusters previously defined as non-clonal due to low CCF and removed before hierarchical clustering (e.g., Cluster 2, 4, 6, 10 in patient 135 and Cluster 2, 3 in patient 118) are dominated by clock-like signatures, consistent with non-tumoral background mutational processes. In patient 135, UV-associated signatures dominate the ancestral clusters (*Figure 2C*, marked by black frame), whereas clock-like signatures are absent in either Cluster 7 or Cluster 11. In contrast, ancestral clusters in patient 118 contain both UV and clock-like signatures, although these are partitioned into distinct clusters (*Figure 2D*, marked by black frame).

Finally, the comutation (comut) plots provide an integrated summary of genomic features across samples, including genomic heterogeneity (defined as the proportion of non-synonymous clonal mutations relative to total non-synonymous mutations), mutational signature composition, tumor purity, treatment stage, tumor progression category, lineage assignment, and key melanoma driver mutations (*Figure 2E–F*). Across both patients, a consistent pattern of genomic heterogeneity over time is observed: early lesions, particularly primary tumors, exhibit higher subclonal diversity, whereas later lesions appear more homogeneous. This pattern is consistent with a model in which a subset of clones presents in early, heterogeneous tumors undergo selective expansion, ultimately giving rise to more homogeneous descendant tumor populations.

CHAPTER 4

Melanoma cells exhibit structured cell state transitions and increasing phenotypic heterogeneity during tumor progression and over time

AI Usage Statement: Portions of this chapter were edited with the assistance of ChatGPT for grammar and clarity.

All scientific content, analysis, and interpretations are the author's own.

4.1 Melanoma cell state transitions from invasive to proliferative over time within the tumor evolutionary lineages

Melanoma tumor cells can be broadly categorized into canonical cell states, including proliferative, invasive, and intermediate states, with substantial variability across samples. The proliferative (melanocytic-like) state is characterized by high expression of *MITF* and low expression of *AXL*, reflecting a more differentiated and growth-oriented phenotype. In contrast, the invasive (mesenchymal-like) state exhibits high *AXL* and low *MITF* expression and has been associated with increased migratory capacity and resistance to therapy. Rather than representing discrete categories, these states exist along a continuum, with additional intermediate phenotypes such as transitory and neural crest-like (NC-like) states. Although definitions of these intermediate states vary across studies, they can be broadly grouped as intermediate or mixed cell states.

The gene expression programs underlying these cell states are captured by curated melanoma signature gene sets. Leveraging 28 melanoma signatures derived from eight published studies (*Supplementary Table 3*; see *Methods 2.4.3*), we performed single-sample gene set enrichment analysis (ssGSEA) on bulk RNA-seq data from patient 118 to quantify the enrichment of these signatures across tumor samples, clinical contexts, biological subgroups. Visualization of melanoma signature enrichment scores across samples (*Figure 3A*) reveals that tumor samples can be broadly assigned to dominant cell states, including proliferative, invasive, intermediate (transitory), or mixed states. When these inferred cell states are overlaid onto the phylogenetic tree, a structured pattern of cell state transitions emerges along evolutionary lineages.

Specifically, while the primary tumor (T1) is characterized by a proliferative and high pigmentation state, samples corresponding to the early branching points in each lineage, such as T2 for L1, T3 for L2, T8 for L3, and the locoregional lymph node tumor T5 belonging to both L1 and L2, all consistently exhibit elevated invasive signatures (*Supplementary Table 3, Figure 3B*, marked by red frame). Intermediate timepoint samples within each lineage, like T6 in L2 and T9 in L3, are enriched for a transitory or mixed state (*Supplementary Table 3, Figure 3B*, marked by orange frame), whereas later-stage lesions (T10–T14) are predominantly proliferative or proliferative with mixed intermediate features (marked by green/brown frames). Notably, instead of a linear transition from proliferative to invasive state over time, the observed pattern demonstrates that invasive signatures are most prominent at key branching points in the phylogeny, suggesting a potential association between invasive cell states and the emergence of new tumor lineages.

The observations above suggest a possible pathway where invasive-like transcriptional programs facilitate tumor dissemination and adaptation during lineage diversification, although direct causality cannot be established from these data. The proliferative melanoma state in later tumor samples in each branch suggests a potential shift from invasive-associated programs during early lineage expansion toward proliferative programs that may support tumor growth and clonal expansion at later stages. Additionally, the enrichment of transitory signatures in intermediate samples within each lineage may reflect ongoing transitions between invasive and proliferative states, consistent with a dynamic and reversible cell state landscape during tumor progression. Lastly, instead of simply proliferative, later tumor samples tend to exhibit a mixed state with proliferative and transitory signatures, explainable by the remainder transcriptomic programs carried over from previous intermediate state but may also have implication in the growing phenotypic heterogeneity during tumor progression. This is why more quantitative characterization of these melanoma signatures is necessary.

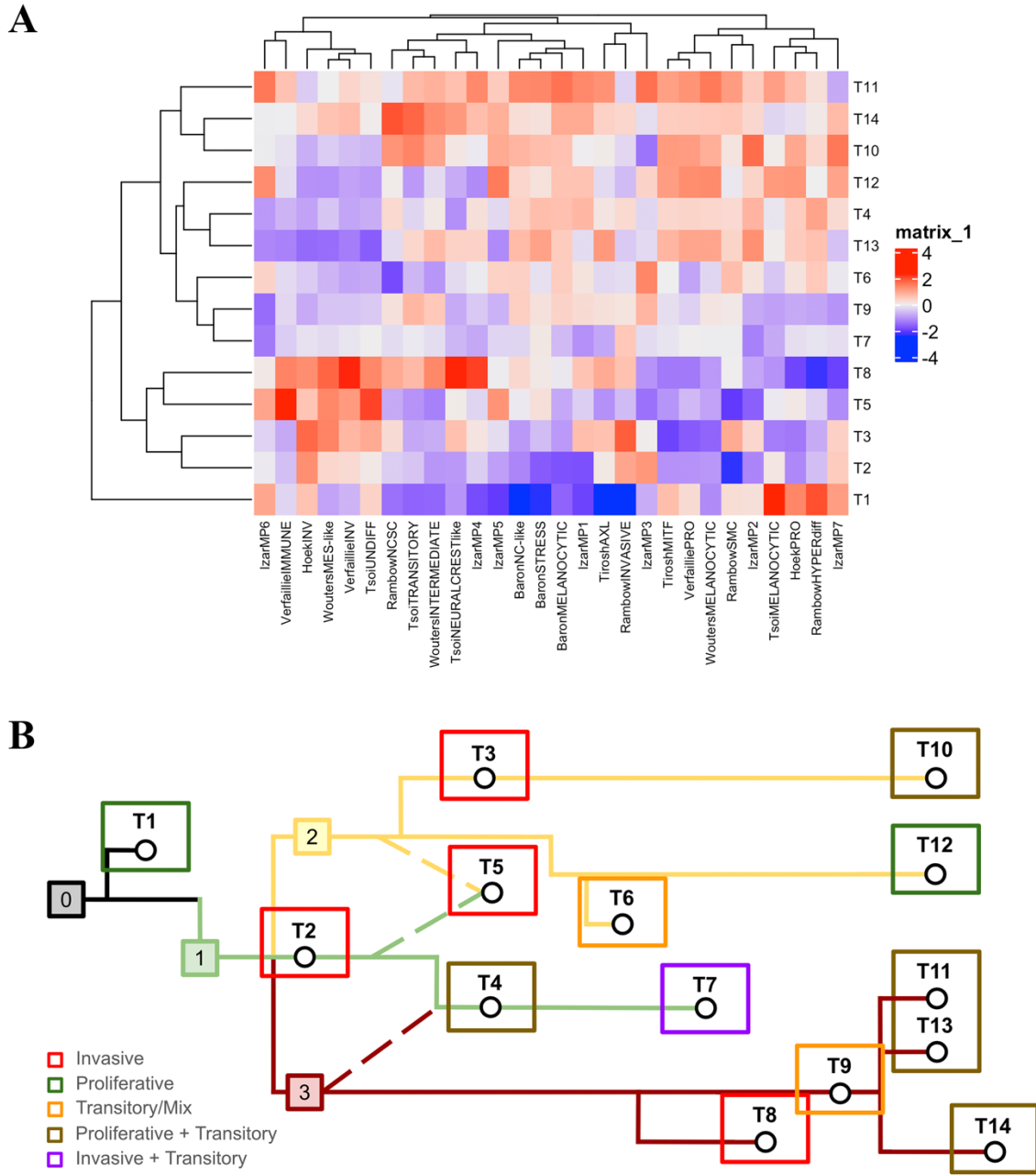


Figure 3. Global and longitudinal pattern of melanoma signature and tumor state transitions. (A) Heatmap with row and column hierarchical clustering to show the Single Sample Gene Set Enrichment Analysis (ssGSEA) enrichment scores of selected melanoma signatures by patient 118 tumor samples. (B) Roughly annotated sample-level melanoma cell states overlaid on patient 118 phylogenetic map. The colors of rectangular frames indicate cell state categories.

4.2 Transcriptional plasticity and increased phenotypic heterogeneity facilitate prolonged tumor survival

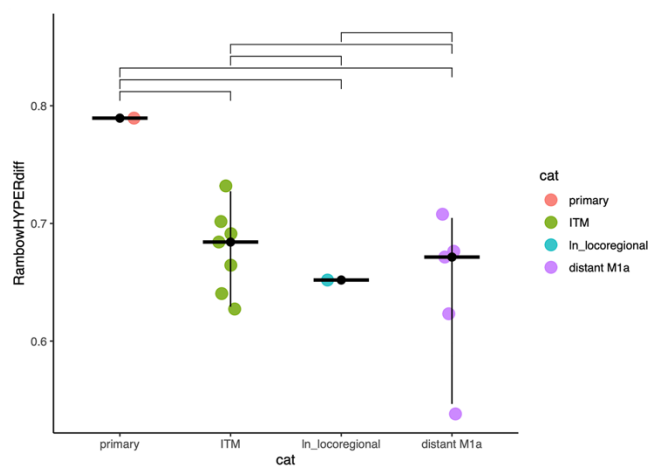
To quantitatively assess differences in melanoma cell states across biological and clinical contexts, we performed pairwise Wilcoxon rank-sum tests on melanoma signature enrichment scores across lineage, treatment, tumor progression category, and timepoint groupings. Note that due to the limited sample size within each subgroup, Benjamini-Hochberg (BH) adjusted p-values were overly conservative and resulted in predominantly non-significant findings. Therefore, nominal p-values are reported below to assess statistical significance.

The hyper-differentiation signature associated with proliferative tumor cells exhibiting elevated pigmentation programs (*Ki67+*, *PCNA+*) has been linked to non-progressor clinical responses and shown to be suppressed in late-stage metastatic disease. Consistent with this prior observation, the primary tumor sample (T1) exhibited elevated pigmentation signature enrichment relative to downstream samples, although this difference did not reach statistical significance (*Figure 3A*, *Figure 4A*).

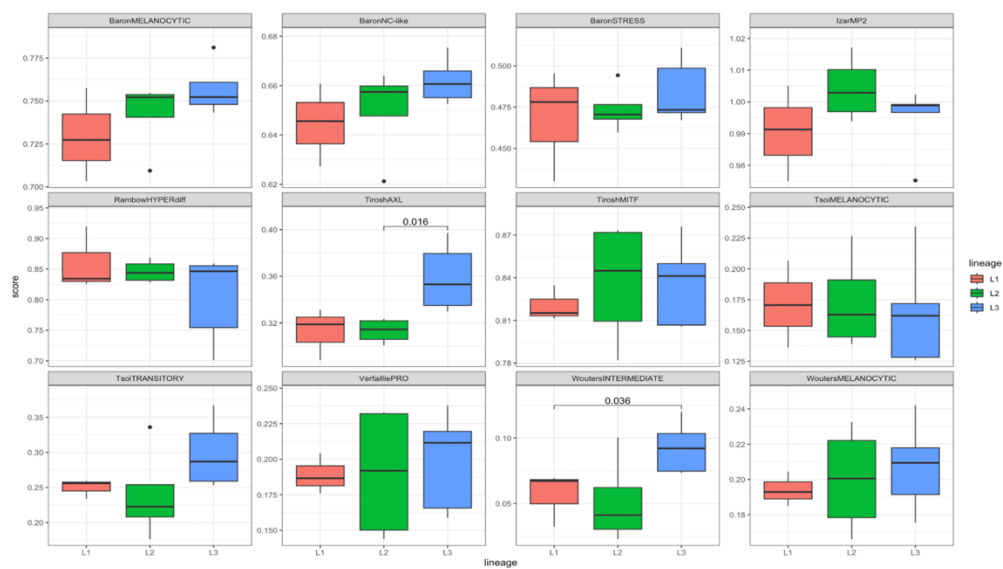
Due to low read depth in T1 based on MultiQC quality control metrics (*Supplementary Figure 1*), as well as its distinct transcriptomic profile observed in PCA analysis (*Supplementary Figure 3B-C*), this sample was excluded from downstream statistical comparisons. Given that T1 represents the sole sample in the ancestral lineage (L0), as well as the only primary tumor and earliest timepoint, its removal does not confound subgroup comparisons.

Two rounds of Wilcoxon comparisons were performed. The first included all bulk RNA-seq signals and all 28 melanoma signature enrichment scores were calculated. In the second round, to better isolate tumor-intrinsic transcriptional programs, bulk RNA-seq data were deconvoluted using InstaPrism to estimate tumor cell proportions and assign tumor cell identities (*Supplementary Figure 4A*). Subsequent analyses were restricted in tumor cells only and included only melanoma signatures that were significantly enriched in at least one prior comparison.

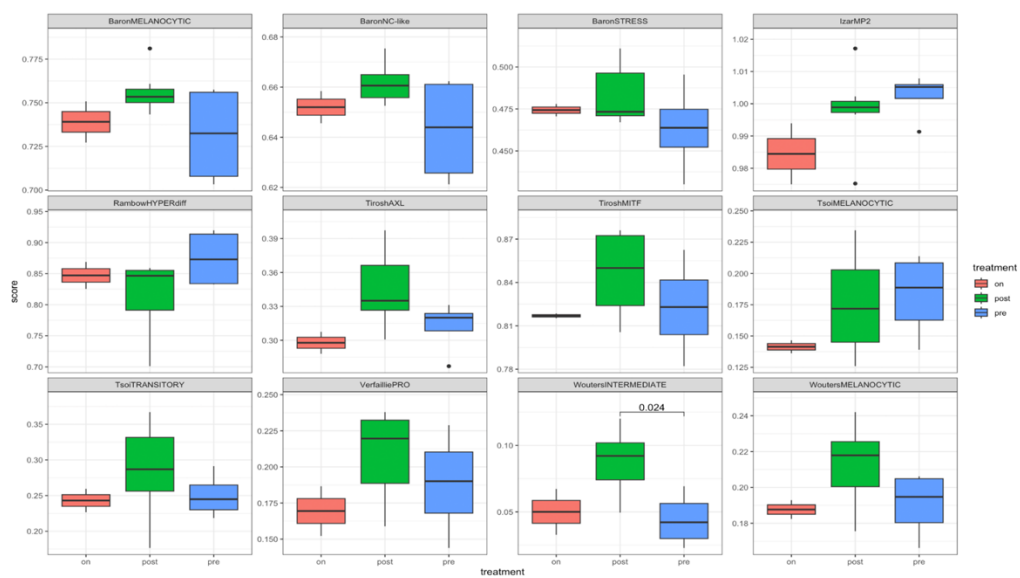
A



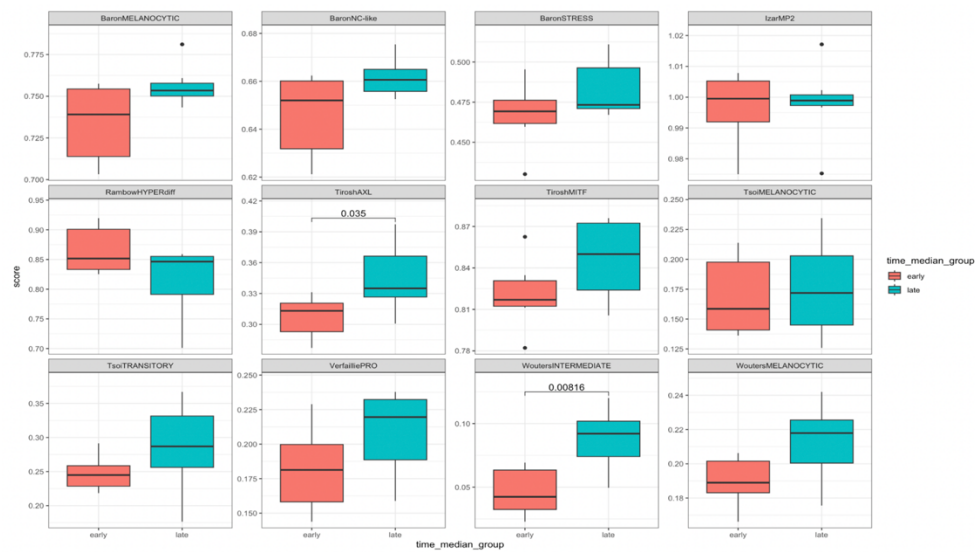
B



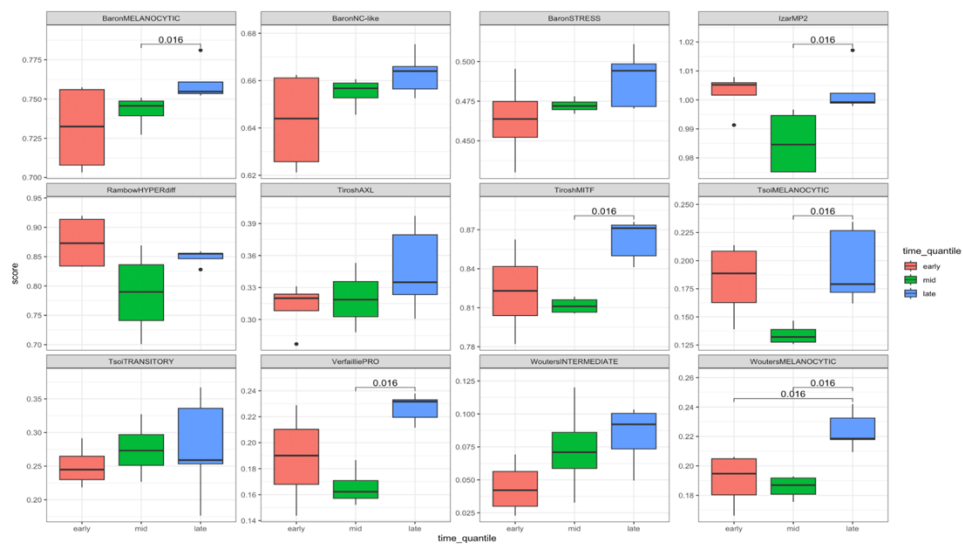
C



D



E



F

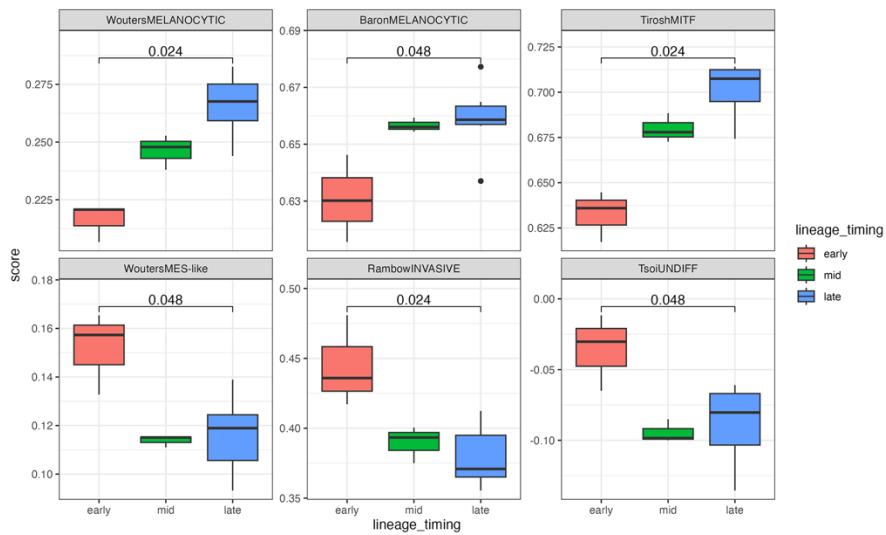


Figure 4. Enrichment of selected melanoma signatures across clinical and biological sample groups. (A)

Enrichment of the hyper-differentiation/pigmentation melanoma signature in the primary tumor. Comparison of the ssGSEA enrichment scores of selected melanoma signatures between **(B)** tumor lineages, **(C)** treatment groups, **(D)** time points separated by median, **(E)** time points separated by tertile, and **(F)** timing within tumor lineages. For panel F, comparison groups early (T2, T3, T8), mid (T4, T6, T9), and late (T7, T10-14) are manually defined based on the sampling time of each tumor and their relative order within their assigned lineage. Significant differences from the pairwise Wilcoxon rank sum test are indicated with the nominal p values (< 0.05). ****Panel A of this figure was created by me and adapted by Giuseppe Tarantino for better visualization.**

Across lineage comparisons, an invasive-associated signature (TiroshAXL) was significantly enriched in L3 relative to L2 (*Figure 4B*, $p = 0.016$). While this initially suggests a potential association between *AXL* expression and distant metastasis because L2 tumors span both early and late time points, this relationship is not statistically significant when comparing metastasis categories directly. Notably, T12, an L2 sample classified as a distant metastasis, exhibits a strong proliferative profile, which may obscure this association. These observations suggest that *AXL* enrichment may be more strongly associated with lineage-specific transcriptional programs rather than metastatic site alone. Enrichment of the invasive signature in L3 initially appear contradicting as the global pattern previously shows a more proliferative profile in these samples (T10-14). However, these samples, besides T12, are not solely enriched for proliferative melanoma signatures, but also have a high transitory and intermediate signature, suggesting that they are mixing both proliferative and invasive signatures in their transcriptomic profiles.

To further support this claim, we observed that an intermediate-state signature (WoutersIntermediate) is significantly enriched in L3 compared to L1 (*Figure 4B*, $p = 0.036$), and is also elevated in post-treatment samples, including all L3 samples and select L2 samples (T10 and T12) (*Figure 4C*, $p = 0.024$). Furthermore, both the invasive and intermediate signatures are significantly higher in later samples compared to earlier samples when stratified by median timepoint (*Figure 4D*, $p = 0.024$). However, when samples are stratified into tertiles, these differences are no longer statistically significant (*Figure 4E*), suggesting a gradual rather than discrete shift in these signatures over time.

In contrast, proliferative signatures do not show significant differences under median-based stratification. Nevertheless, when stratified by tertile timepoints, all previously enriched proliferative signatures are significantly enriched in late-stage samples compared to mid-stage samples (*Supplementary Table 3*), with one (WoutersMelanocytic) also elevated relative to early samples (*Figure 4E*). This pattern suggests that proliferative activity is highest in late-stage tumors, while mid-stage samples, particularly those near major phylogenetic branching points like T8, exhibit lower proliferative activity and higher invasive or intermediate characteristics.

Contrary to the initial expectation that invasive, mesenchymal-like signatures would dominate late-stage tumors due to their association with therapy resistance, late-stage samples instead display strong proliferative and melanocytic features. When considered alongside earlier observations that samples at lineage branching points are enriched for invasive signatures, this suggests a dynamic shift in tumor cell states over time but a constant transition pattern within tumor lineages. In fact, when we characterize the proliferative and invasive signatures between early, middle, and late samples within different tumor clones, we observed significantly higher proliferative signatures in later samples (T7, T10-14) and significantly higher invasive signatures in earlier samples (T2, T3, T8) across all tumor lineages (*Figure 4F*).

One explanation of the observed pattern is that invasive transcriptional programs are preferentially enriched during periods of lineage expansion or under environmental stress like clinical treatments, while proliferative programs become dominant during later stages of tumor growth and clonal expansion under more stable conditions. The persistent presence of intermediate signatures in later samples further supports the idea that tumor cell populations retain transcriptional plasticity, reflecting either ongoing state transitions or increased phenotypic heterogeneity within expanding tumor clones. While it is commonly believed that invasive cell states are associated with immunotherapy resistance, our findings show a high proliferative cell state among later, knowingly immune-evasive samples. This is surprising but does not directly contradict the prior knowledge, as immune interaction plays a pivotal role in shaping tumor states. If the environmental pressure from immune surveillance is low in later samples, even the melanocytic-like melanoma can survive and proliferate. Despite preliminary inspection using CIBERSORT does not show largely different immune infiltration levels between early and later samples or among late samples (*Supplementary Figure 4B-C*), further characterization of the immune infiltration level of specific immune subtypes is necessary.

Overall, the results above provide quantitative evidence that phenotypic heterogeneity, as reflected by diversity in melanoma cell states, increases over time during tumor progression. This trend contrasts with the decreasing genomic heterogeneity observed in Aim 1, suggesting that transcriptional plasticity may compensate for reduced genetic diversity in later-stage tumors. Overall, late-stage tumors demonstrate high *MITF*, proliferation, melanocytic, and intermediate (mixed) signatures compared to both the early and middle samples, while there is a trend of invasive shifting toward proliferative cell state within tumor lineages. *AXL* gene expression level may be a

key to both the clinical and phylogenetic transition point, marking the emergence of distant metastasis, gaining of therapy resistance, and seeding of more fit lineage.

4.3 Immune infiltration levels of macrophages and dendritic cells weakly associated with specific melanoma cell states

To investigate the tumor microenvironment and potential interaction between immune and tumor cells, we estimated immune cell infiltration from bulk RNA-seq data using CIBERSORT, a deconvolution algorithm that infers the relative proportions of immune cell types from gene expression profiles. This approach enables characterization of immune composition across tumor samples without requiring single-cell resolution.

Using CIBERSORT, we obtained both the relative immune cell type proportions (*Supplementary Figure 4B*) and the overall immune infiltration levels per sample (*Supplementary Figure 4C*). Total immune infiltration was lowest in the two early samples, T1 and T2, suggesting limited immune engagement at initial stages of tumor development. However, beyond these early timepoints, overall immune infiltration became more variable and did not show a clear trend with respect to lineage or time, indicating that bulk immune abundance alone may not capture structured biological differences across tumor progression.

To explore more specific patterns, we performed Wilcoxon rank-sum tests comparing individual immune cell types across predefined groupings. Only a small number of immune populations showed enrichment, and significant differences are only between tumor lineages (*Supplementary Figure 5A*). Specifically, activated natural killer (NK) cells are more abundant in L1 tumors compared to L2, while plasma cells are enriched in L2 tumors relative to both L1 and L3. These observations may reflect differences in immune contexture across lineages. Activated NK cells are part of the innate immune system and are involved in early anti-tumor responses through direct cytotoxicity. Their enrichment in L1, an early and ultimately extinct lineage, may suggest stronger initial immune surveillance contributing to its eventual elimination. In contrast, plasma cells, a type of terminally differentiated B cells responsible for antibody production, are more prominent in the longitudinally persisting lineage L2. This may indicate a shift toward a more adaptive, humoral immune response that is less effective at tumor clearance, potentially allowing sustained tumor survival.

Beyond lineage-specific differences, we examined associations between immune infiltration and tumor-intrinsic transcriptional programs. Although most immune cell types did not show strong group-wise differences, macrophage M1 cells displayed notable trends. M1 macrophages, which are classically pro-inflammatory and associated with anti-tumor activity, were more abundant in distant metastases and L3 samples, although this did not reach statistical significance (*Supplementary Figure 5A-B*). Importantly, macrophage M1 infiltration is positively correlated with the intermediate signature, WoutersIntermediate (*Figure 5A*, $\rho = 0.54$, $p = 0.0451$), suggesting a potential link between inflammatory microenvironments and tumor cells in intermediate or transitioning states.

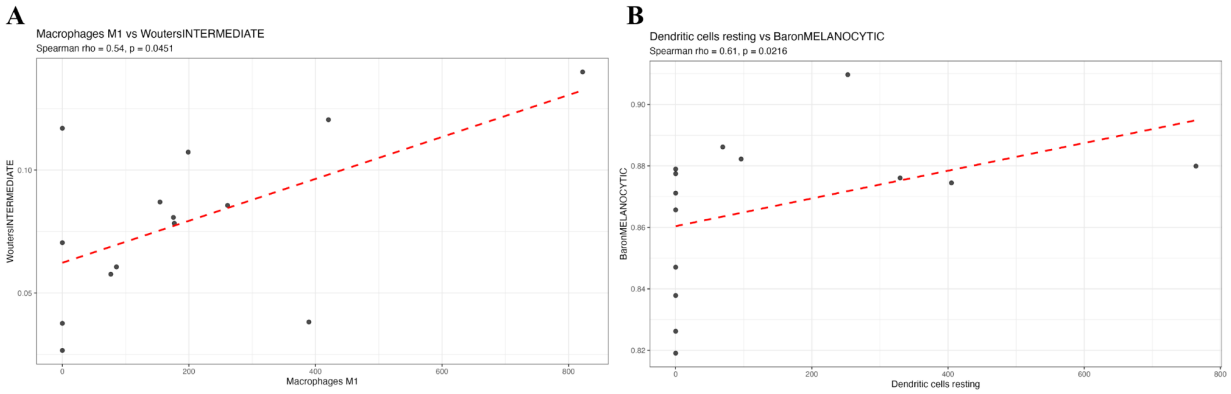


Figure 5. Spearman correlation between immune cell infiltration level and melanoma signature enrichment scores. Significant spearman correlation between (A) M1 macrophages and intermediate melanoma signature, and (B) resting dendritic cells and melanocytic melanoma signature. Correlation coefficient (rho) and p-value (p) are included.

In addition, resting dendritic cells showed a positive correlation with a proliferative signature, BaronMELANOCYTIC (*Figure 5B*, $\rho = 0.61$, $p = 0.0216$). While their overall abundance is higher in distant or L3 samples, this pattern was not consistently observed across all late timepoints, suggesting that their association may be more related to metastatic site or lineage-specific context rather than temporal progression alone. Given the role of dendritic cells in antigen presentation and T cell priming, this relationship may reflect interactions between differentiated tumor states and immune recognition processes.

Taken together, these results suggest that while global immune infiltration patterns are not strongly structured across tumor progression, specific immune populations like macrophages may play an important role in shaping tumor cell states and metastatic behavior. However, bulk transcriptomic data limits our ability to resolve direct interactions between tumor and immune cells or to determine their spatial organization within the tumor microenvironment. These findings therefore motivate further investigation using spatially resolved approaches,

where tumor–immune interactions can be examined in situ. In the next aim, we extend this analysis to spatial transcriptomic data (MERFISH) and a larger multi-patient HTAN cohort with defined clinical response groups to better understand how tumor programs and immune context jointly influence disease progression and therapeutic outcomes.

CHAPTER 5

Integrative multi-modal analysis identifies tumor programs and immune-associated pathways linked to clinical response in melanoma

AI Usage Statement: Portions of this chapter were edited with the assistance of ChatGPT for grammar and clarity.

All scientific content, analysis, and interpretations are the author's own.

5.1 HTAN patient cohort enables multi-modal analysis of tumor-immune interactions across clinical response groups

Building on the observed associations between melanoma cell states and immune infiltration in patient 118, we next sought to determine whether similar tumor-immune relationships are consistently observed across a larger patient cohort and at higher cellular resolution. While bulk RNA-seq analysis provides an overview of immune composition and its correlation with tumor-intrinsic programs, it does not capture the spatial organization of specific cell types within the tumor microenvironment or directly link these features to clinical response.

To address these limitations, we extended our analysis to the HTAN melanoma cohort, which includes snRNA-seq, WES, spatial transcriptomics (MERFISH), and spatial proteomics (CODEX, Lunaphore) data across multiple patients with well-defined clinical response categories. Subsequent analyses focused majorly on comparing tumor-intrinsic pathways and tumor-immune interactions between the progressor, including rapid progressors (RP) and intermediate progressors (IMP), and non-progressor, including durable clinical benefits (DCB) and intermediate non-progressors (IMnotP), response groups. At this stage, our analysis primarily focuses on snRNA-seq and MERFISH modalities, while spatial proteomics data will be incorporated in future work to further support these findings. This multi-modal dataset enables a more detailed characterization of immune-related pathways and tumor-immune interactions at both cellular and spatial levels.

For the snRNA-seq data, we analyzed tumor and immune cells separately. Data preprocessing, including quality control, batch correction, and initial annotation, was performed by previous members of the lab (Yiwen He,

Priyanka Solanky, and Xinyu Cui) (*Supplementary Figure 6A-B*). Immune cell type and subtype annotations were defined using established marker genes in collaboration with clinical and immunology experts. Major immune cell populations identified in the dataset include B cells, T cells, macrophages, natural killer (NK) cells, plasma cells (PC), and dendritic cells (DC) (*Supplementary Figure 6C*). Additional functionally relevant immune subtypes were also identified, including proliferating, memory, and exhausted CD4⁺ T cells, regulatory T cells (Tregs), tumor-associated macrophages (TAM), conventional and plasmacytoid dendritic cells (cDC and pDC), and more (*Supplementary Figure 6D*). The expression of canonical marker genes supports the accuracy of these annotations (*Supplementary Figure 6E*).

Leveraging the curated snRNA-seq data from the HTAN cohort, we first investigated the enrichment of intrinsic transcriptomic programs like melanoma signatures and human hallmark pathways across tumor and immune cell populations to identify features associated with clinical response. We then extended this analysis by defining tumor programs based on coordinated pathway activity, comparing previously established NMF-derived programs with newly defined, biologically interpretable programs, and evaluating their associations with clinical outcomes and tumor-immune interaction patterns.

5.2 Antigen presentation and interferon-related pathways are enriched in non-progressors

To investigate transcriptomic programs associated with clinical response, we performed pseudo-bulk aggregation of tumor snRNA-seq data followed by differential expression analysis of curated melanoma signatures and human hallmark gene sets. This approach enables robust comparison of pathway-level activity across patient groups while reducing single-cell noise.

Comparing rapid progressors (RP) with non-progressors (DCB + IMnotP), we observed significant enrichment of interferon-related pathways, including interferon- α and interferon- γ responses in human hallmark gene sets, in non-progressors (*Figure 6A*). This finding is further supported by melanoma-specific signatures, which are also elevated in non-progressors (*Figure 6B, Supplementary Table 3*). Consistent results were obtained when comparing all progressors (RP + IMP) against non-progressors, indicating that these immune-related pathways are a robust feature of favorable clinical response (*Figure 6C-D*).

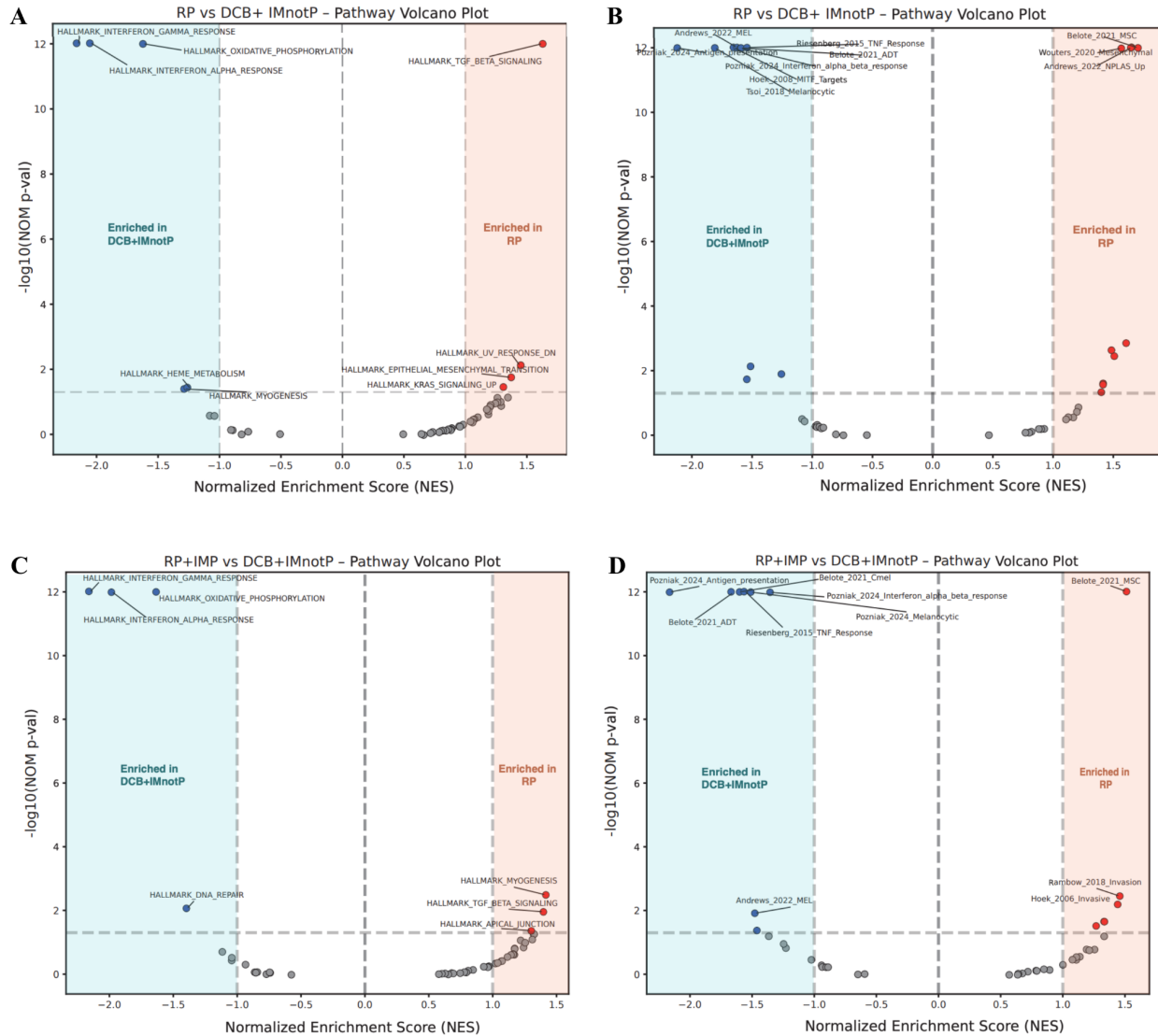


Figure 6. Enrichment of melanoma signatures and human hallmark pathways across clinical response groups. Volcano plot showing significantly enriched (A, C) melanoma signatures and (B, D) human hallmark pathways between different clinical response groups. Blue points indicate signatures significantly enriched in non-progressors. Red points indicate signatures significantly enriched in rapid progressor (A-B) or all progressor (C-D). *RP* = rapid progressor, *DCB* = durable clinical benefit, *IMnotP* = intermediate non-progressor, *IMP* = intermediate progressor. ****This figure was created by me and adapted by Giuseppe Tarantino for better visualization.**

Biologically, interferon signaling and antigen presentation are central to anti-tumor immunity. Interferon- γ , primarily produced by activated T cells and NK cells, enhances tumor cell visibility to the immune system by upregulating antigen presentation machinery, including major histocompatibility complex (MHC) molecules. Interferon- α and β are type I interferons also promoting immune activation, dendritic cell maturation, and T cell

responses. Enrichment of these pathways in non-progressors suggests a tumor microenvironment that is more immunologically active and capable of executing effective anti-tumor responses.

In parallel, melanoma signatures reveal that progressors are enriched for mesenchymal-like programs (Belote_2021_MSC, Wouters_2020_Mesenchymal), while non-progressors exhibit higher melanocytic signatures (Tsoi_2018_Melanocytic, Andrews_2022_MEL) (*Figure 6B, D*). This is consistent with prior findings that mesenchymal-like states are associated with immune evasion and resistance to immunotherapy, whereas melanocytic states tend to be associated with non-progressive diseases.

Given the enrichment of antigen presentation and interferon-related pathways in non-progressors, we next examined the expression of MHC class I and II genes, which are the core molecular machinery responsible for presenting antigens to CD8⁺ cytotoxic T cells and CD4⁺ helper T cells, respectively. From above, we hypothesized that MHC genes will also be upregulated in non-progressors. However, differential expression analysis reveals no significant difference in the MHC gene expression level across response groups in the snRNA-seq data, including analyses restricted to tumor cells, immune cells, macrophages, and CD163⁺ TAMs subsets (*Supplementary Table 4A-B*).

This lack of consistent single-gene signals suggests that pathway-level enrichment observed in transcriptomic analyses may not be driven by uniform upregulation of canonical MHC genes but could instead reflect more complex or context-dependent regulatory mechanisms. For example, the coordinated regulation of multiple pathway components in addition to the MHC genes or the activation of upstream signaling like the interferon pathways without strong transcriptomic shifts in individual genes.

To further investigate this discrepancy, we extended the analysis to the MERFISH spatial transcriptomics dataset, which provides cellular and spatial resolution of gene expression within the tumor microenvironment. Because antigen presentation is a multi-cellular process involving both tumor and immune cells, we examined MHC gene expression across multiple compartments, including all cells, tumor cells, immune cells, and macrophage plus monocyte populations (*Supplementary Figure 7A-D*).

HLA-DPBI, an MHC-II gene associated with CD4⁺ T cells, is expressed significantly higher in non-progressors in immune cells (including macrophage and monocytes specifically) but not in tumor cells (*Supplementary Figure 7C-D*). On the contrary, *HLA-DRBI*, another MHC-II gene associated with CD4⁺ T cells, is expressed higher in progressors than non-progressors in tumor cells only (*Supplementary Figure 7B*). These findings

highlight that different MHC-II genes may play distinct roles depending on cell type and the spatial context. Given the variability across individual genes and spatial environment, single-gene signals are limited in capturing the underlying biology of highly heterogeneous cancer, so we need to identify more coherent global tumor states, known as the tumor programs.

5.3 NMF-defined tumor programs successfully identify melanocytic and proliferative tumor programs but remain ambiguous otherwise

Tumors, particularly melanoma, are highly heterogeneous and consist of multiple dynamic cell states. Tumor programs are defined as coordinated gene expression patterns that reflect functional tumor states, including the underlying biological processes and cellular phenotypes. Rather than relying on individual gene expression, tumor programs capture recurrent global transcriptional patterns, making them a robust framework for studying intra-tumoral heterogeneity. These programs often correspond to key biological processes such as proliferation, invasion, stress response, and metabolic adaptation. As such, interpretable tumor programs are essential for integrating multi-modal datasets and enabling comparisons across studies.

To characterize transcriptional heterogeneity in melanoma, we first examined tumor programs derived using a data-driven approach based on non-negative matrix factorization (NMF) applied to single-nucleus RNA-seq data (see *Methods 2.5.1*). NMF decomposes complex transcriptomic data into a set of metaprograms that represent shared patterns of gene expression across samples, allowing for the identification of recurring tumor states without prior assumptions.

Using this approach, six NMF-based tumor programs (TP1-TP6), previously identified by members of the lab (Priyanka Solanky and Giuseppe Tarantino), were evaluated and annotated (*Figure 7A*) based on enrichment of curated melanoma signatures and human hallmark pathways (*Figure 7B-C*). These programs correspond to known melanoma cell states to varying degrees. TP1 represents a melanocytic state associated with active cell growth and cycling, while TP2 captures a proliferative program. TP3 reflects a neural-crest-like (NC-like) state with features of epithelial-to-mesenchymal transition (EMT). In contrast, several programs were less clearly defined. TP4 exhibits a mixture of melanocytic and NC-like features, while TP6 reflects a hybrid state combining NC- and mesenchymal-

Despite providing a useful framework for summarizing global transcriptional patterns, NMF-defined tumor programs solely rely on gene expression data, making them susceptible to noise and technical variation. In addition, the resulting programs can be disproportionately influenced by highly expressed housekeeping genes that are not directly related to tumor biology. To address these limitations, we next sought to define tumor programs using a complementary, biologically informed approach based on coordinated pathway and gene signature activity, enabling more interpretable and robust characterization of tumor cell states.

5.4 Defining novel biologically relevant tumor programs based on recurrent pathway and signature activity

Building on the limitations of NMF-defined tumor programs, which rely on gene-level expression and may lack biological interpretability, we next defined tumor programs using a biologically informed approach based on coordinated pathway and gene signature activity. Because tumor states are driven by integrated biological processes rather than individual genes, this approach provides a more robust and interpretable framework for characterizing tumor heterogeneity.

To achieve this, we used single-cell enrichment scores derived from both human hallmark gene sets⁶² and curated melanoma signature gene sets⁴⁵⁻⁶¹ (*Supplementary Table 3*; see *Methods 2.5.2*). These enrichment scores, rather than raw gene expression values, were used as input features for clustering. Human hallmark capture broadly conserved biological processes such as proliferation, metabolism, and immune signaling, while melanoma signatures reflect disease-specific cell states, including melanocytic, mesenchymal-like, neural crest-like, and other programs.

We performed Leiden clustering independently on hallmark-based and melanoma signature-based enrichment matrices, generating two parallel clustering representations of tumor cell states (*Supplementary Figure 8A-B*). To ensure robustness and biological relevance, clusters were filtered to remove non-tumor or artifactual populations based on canonical melanoma marker expression (*Supplementary Figure 8C-D*). Highly similar clusters were then merged using Jaccard similarity based on overlap of differentially expressed genes (*Supplementary Figure 8E*), resulting in a refined set of clusters for each approach. Notably, both hallmark-based and melanoma signature-based clustering converged on six biologically distinct clusters, enabling direct comparison and integration across methods.

We first annotated clusters derived from human hallmark pathways enrichment, incorporating melanoma signature enrichment and marker gene expression for interpretation. Clusters 0 and 4 represent basic tumor populations confirmed by the expression of canonical melanoma markers (*Supplementary Figure 8C*). Cluster 4 shows minimal activation of stress, immune, or proliferative pathways (*Figure 8A-B*) while retaining the core melanoma programs like *MLANA* and *PMEL*, suggesting a relatively stable tumor state. We define it as the **Baseline** tumor program. In contrast, Cluster 0 exhibits moderate melanocytic and proliferative signatures with relatively low MYC target and mTORC1 activity, consistent with an early-stage proliferative state, defined as the **Pre-Proliferative** tumor program.

Clusters 1 and 2 both display strong melanocytic features (*Figure 8B*) but differ in their associated stress and metabolic activity. Cluster 1 has the highest MYC target activity among all clusters, along with enriched mTORC1 signaling, unfolded protein response, glycolysis, and hypoxia pathways (*Figure 8A*), suggesting rapid cell growth under metabolic stress. This cluster is further supported by enrichment of hyper-differentiated melanoma signatures and is annotated as the **Proliferative/Hyper-differentiated** tumor program. In comparison, Cluster 2 exhibits high MYC activity without strong stress-related signals, indicating active proliferation in a less stressed environment. This cluster is annotated as the **Proliferative** tumor program.

Cluster 3 is characterized by high MYC activity combined with strong enrichment of epithelial-to-mesenchymal transition (EMT) pathways (*Figure 8A*), along with elevated NK-like signatures mixed with moderate mesenchymal-associated melanoma states (*Figure 8B*). This profile indicates a state that is actively transitioning toward invasiveness, annotated as the **EMT/Invasive** tumor program.

Cluster 5 exhibits the lowest MYC target activity, suggesting reduced proliferation. Nevertheless, it also shows strong enrichment of interferon- α and interferon- γ response pathways and moderate degree of antigen presentation signals (*Figure 8A-B*). This pattern is consistent with a tumor state under sustained immune interaction or immune-mediated growth suppression, defined as the **Immune-Interacting** tumor program.

Together, these analyses defined six biologically interpretable tumor programs: Baseline, Pre-Proliferative, Proliferative, Proliferative/Hyper-differentiated, EMT/Invasive, and Immune-Interacting (*Figure 8C*). These programs are supported by consistent patterns of pathway enrichment, melanoma signature activity, and expression of canonical marker genes, particularly the distinct markers found in melanocytic, NC-like, and mesenchymal-like cell states (*Supplementary Figure 8G-H*). Notably, the Proliferative cluster (Cluster 2) also expresses mesenchymal-

associated genes such as *SOX9*, *EGFR*, and *ZEB1*, suggesting partial overlap between proliferative and invasive states and highlighting potential transitional behavior.

Using the same clustering framework, we then analyzed tumor cell states using melanoma signature-based enrichment scores as input features (see *Methods 2.5.2*). This approach yielded a similar set of tumor programs as those derived from hallmark pathway enrichment, including Baseline, Pre-Proliferative, Proliferative/Hyper-differentiated, EMT/Invasive, and Immune-Interacting states (*Figure 8D-F*).

However, instead of the Proliferative cluster identified through hallmark-based clustering, the melanoma signature-based approach revealed a Transient/Intermediate cluster. This cluster is characterized by concurrent enrichment of melanocytic, proliferative, NC-like, and mesenchymal-like signatures, indicating a mixed transcriptional profile spanning multiple canonical melanoma states. When considered alongside the partial overlap and mixed features observed in the hallmark-derived Proliferative cluster, these results suggest the presence of a transitional tumor population undergoing dynamic shifts between phenotypic states.

To systematically integrate tumor programs identified from both approaches, we compared cluster similarity based on the top 30 differentially expressed genes (DEGs) per cluster using an UpSet plot (*Figure 8G*, see *Methods 2.5.3*). This analysis reveals a perfect one-to-one correspondence between Baseline, Pre-Proliferative, Proliferative/Hyper-differentiated, EMT/Invasive, and Immune-Interacting clusters derived from hallmark and melanoma signature-based clustering. Notably, the melanoma-defined Intermediate cluster corresponds most closely to the hallmark-defined Proliferative cluster, sharing 15 out of the top 30 DEGs. This partial overlap supports the interpretation that this population represents a hybrid or transitional state rather than a purely proliferative phenotype. Based on this integrative analysis, we redefined this group as the Intermediate tumor program.

To enable downstream application and cross-modality validation, we next defined marker gene sets for each tumor program by selecting mutually exclusive genes from the top 100 DEGs per cluster (*Supplementary Table 5A*; see *Methods 2.5.4*). These marker sets show strong concordance with tumor program assignments in the snRNA-seq dataset (*Supplementary Figure 9A-B*).

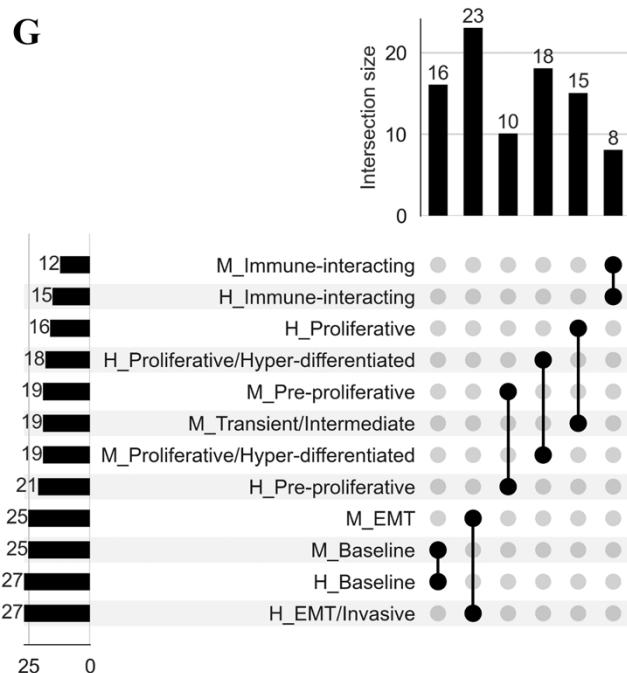


Figure 8. Identification and annotation of novel tumor programs. Leveraging (A, D) human hallmark and (B, E) melanoma signature expression information in each cluster, tumor programs are defined and annotated from the leiden clusters generated based on cell-level (C) human hallmark and (F) melanoma signature enrichment scores. (G) UpSet plot showing the overlapping of top 30 differentially expressed genes (DEGs) between biologically annotated leiden clusters. Suffixes of “H_” and “M_” indicate the clusters generated based on human hallmark and melanoma signature scores, respectively.

To assess generalizability, we further restricted these marker sets to genes present in our MERFISH panel (*Supplementary Table 5B*). Despite reduced gene coverage, these curated markers retain the ability to distinguish multiple tumor programs in snRNA-seq data, including EMT/Invasive, Intermediate, Immune-Interacting, and Proliferative/Hyper-differentiated states (*Supplementary Figure 9C-D*), supporting the transferability of pathway-defined tumor programs across modalities.

5.5 Preliminary spatial mapping of novel tumor programs reveals context-dependent tumor-immune associations

To assess whether transcriptionally defined tumor programs can be recapitulated in spatial tissue architecture, we performed a preliminary analysis using MERFISH data and marker genes derived from pathway-defined tumor programs (*Supplementary Table 5B*), as described in *Result 5.4*. Tumor cells were assigned to one of the six newly defined programs based on these markers, and spatial neighborhoods were defined by non-tumor cells

located within a 40-unit radius of each tumor cell. We then quantified the composition of immune and stromal cells in the local microenvironment of each tumor program.

Across all samples, tumor cells showed consistent spatial proximity to exhausted CD8⁺ T cells, monocytes/macrophages, and stromal cells (*Supplementary Figure 10A*). However, no substantial differences in neighborhood composition were observed between tumor programs at the cohort level, suggesting that global tumor-immune associations are not strongly program-specific. This lack of separation is likely influenced by inter-sample variability, as more distinct patterns emerge when observing individual slides or stratifying by clinical and anatomical context.

When stratified by clinical response, we found that tumors from intermediate response groups, both intermediate progressor (IMP) and intermediate non-progressors (IMnotP), exhibited increased proximity to exhausted CD8⁺ T cells and macrophages, whereas tumors from rapid progressor (RP) patients showed minimal spatial association with immune cell types and were instead predominantly adjacent to stromal cells (*Supplementary Figure 10B*). This suggests that tumors from intermediate response groups may exhibit an active but potentially suppressed immune engagement, while rapid progressors demonstrate an immune-excluded microenvironment that may contribute to poor therapeutic response. Still, these patterns were consistent across tumor programs, implying that clinical context may have a stronger role than tumor-intrinsic state in shaping global immune association.

Site-specific analysis further revealed distinct spatial organizations of tumor-immune interactions. Lymph node and soft tissue tumors showed increased proximity to exhausted CD8⁺ T cells and macrophages, with Immune-Interacting tumor programs in lymph nodes displaying particularly strong association with exhausted CD8⁺ T cells compared to all other tumor programs (*Supplementary Figure 10C*). In contrast, skin and lung tumors exhibited reduced immune association and increased proximity to stromal cells, while jejunum and nasal cavity tumors showed moderate association with immune populations like exhausted CD8⁺ T cells and macrophages. Dendritic cells have slightly higher proximity to tumor cells in skin and jejunum samples. These findings highlight the influence of anatomical context on tumor-microenvironment interactions.

To further examine spatial organization at higher resolution, we analyzed representative MERFISH slides from lymph node (LN) biopsies, as this site shows the clearest tumor program-specific distinctions. While inter-tumoral heterogeneity remains substantial across slides, spatial structures of tumor programs were observed in individual samples. For example, one LN biopsy exhibits organized clusters of tumor cells in which Pre-

Proliferative and Proliferative/Hyper-differentiated cells are mostly located at the core, surrounded by EMT/Invasive and Immune-Interacting tumor cells, with Intermediate tumor cells occupying transitional positions between these layers (*Supplementary Figure 11A*). Macrophages are frequently observed near tumor cells, whereas T cells are generally distant from large tumor clusters (*Supplementary Figure 11B*). Notably, EMT/Invasive tumor cells showed stronger spatial association with macrophages compared to other tumor states (*Supplementary Figure 11C*). Another LN biopsy highlights large clusters of Immune-Interacting tumor cells occupying the borderline between tumor and immune cells, with a particular proximity to macrophages (*Supplementary Figure 11D-F*).

While further spatial exploration is ongoing, these preliminary analyses suggest that while tumor programs may not exhibit strong global differences in immune association, spatial organization and tumor-immune interactions are highly context-dependent, varying by clinical responses and anatomical sites. These findings support the feasibility of mapping transcriptionally defined tumor programs onto spatial data and motivate future work to systematically characterize tumor spatial dynamics and functional roles under the clinical context.

CHAPTER 6

Discussion

AI Usage Statement: Portions of this chapter were edited with the assistance of ChatGPT for grammar and clarity.

All scientific content, analysis, and interpretations are the author's own.

6.1 Summary of key findings

In this thesis work, we demonstrated through the multi-omics analysis of longitudinal tumor samples and large cohort patient data that melanoma metastasis is a dynamic process shaped by the interplay between tumor evolution, phenotypic plasticity, and the tumor-immune microenvironment.

Longitudinal phylogenetic tree reconstruction of two melanoma patients revealed that distinct tumor lineages undergo extinction, persistence, or expansion in response to therapeutic pressure, suggesting tumor evolution may be characterized by dynamic clonal selection under clinical intervention.

At the phenotypic level, melanoma cell states were found to transition in a non-linear manner over the course of disease progression. However, when viewed in the context of branched tumor clones, these transitions follow a structured trajectory from invasive to intermediate to proliferative states, corresponding to initial dissemination, adaptation, and subsequent clonal expansion. Consistent with this model, we observed that while genomic heterogeneity, measured by subclonal mutation diversity, decreases over time, tumors exhibit increasing heterogeneity at the level of mutational processes and cell states, indicating a shift from genetic diversification to functional diversification during progression and over time.

Extending these findings to a larger cohort, transcriptomic analyses revealed that non-progressor patients are enriched for antigen presentation and interferon-related pathways, although these differences are not consistently captured at the level of individual genes such as MHC components in the pseudo-bulked transcriptomic modality, highlighting the importance of pathway-level interpretation and spatial-context dependency.

Finally, by integrating human hallmark pathways and melanoma-specific gene signatures, we defined biologically interpretable tumor programs that represent recurrent tumor-intrinsic states shared across patients,

providing a unifying framework to study tumor behavior across modalities. Preliminary validation of these tumor programs in MERFISH data demonstrate context-specific variation between programs, suggesting that tumor-immune interaction is sensitive to anatomical sites and is associated with clinical response to therapies.

EMT/Invasive and Immune-Interacting tumor programs showed special spatial interaction with exhausted CD8⁺ T cells and monocytes/macrophages, which is worth exploring in subsequent experiments. Together, these findings suggest that the highly heterogeneous melanoma progression is driven by coordinated genomic evolution, phenotypic plasticity, and tumor-immune interactions.

6.2 Melanoma progression and intratumoral heterogeneity in the scope of tumor evolution

Tumor progression is fundamentally an evolutionary process shaped by intratumoral heterogeneity and selective pressures arising internally and externally [citation]. Previous studies have shown that genetic instability, epigenetic regulation, tumor-environment interactions, and clinical interventions collectively generate diverse subclonal populations, enabling tumors to adapt and evolve over time⁹⁹.

Consistent with this framework, our longitudinal analysis captures dynamic clonal selection during melanoma progression. In patient 118, the primary tumor exhibits the highest genomic heterogeneity, characterized by numerous subclonal mutations and mutational clusters that later become clonal in descendant samples. Following local ILP and lymph node excision, the earliest lineage (L1) was eliminated, whereas L2 persisted and L3 emerged at later timepoints (*Figure 1A*). These lineage-level shifts likely reflect selective pressures imposed by treatment and the microenvironment, where only specific clones survive and expand while others become extinct. From an evolutionary perspective, such transitions resemble bottleneck events, which reduce genomic diversity while enriching for resistant or better-adapted clones, and thus provide important insight into mechanisms of treatment response and resistance.

Further insight into intratumoral heterogeneity is revealed through the behavior of subclonal populations across time and lineages. We observed multiple patterns of subclonality in longitudinal samples. First, early subclones can later become dominant, as seen in patient 118 where PyClone Clusters 4 and 11 are initially subclonal in the primary tumor T1 (CCF ~0.2) but later become clonal and define L1. Similarly, Cluster 7 transitions from subclonal in T1 to clonal in a sub-branch of L2 (T6 and T12) (*Figure 1A*). These patterns highlight how minor

subpopulations present early in tumor development can serve as reservoirs for future clonal expansion under selective pressure.

Second, shared subclonality across lineages suggests more complex evolutionary dynamics. For example, T5 in patient 118, the only locoregional lymph node sample, contains subclonal mutations defined in two independent branches with a CCF of 0.5, marking its membership in two distinct tumor clones (*Figure 1A*). This observation is consistent with the role of lymph nodes as hubs for metastatic dissemination, where tumor cells from different origins may converge and mix before distant metastasis¹⁰⁰, supporting a model of polyclonal seeding⁴⁰⁻⁴². Nevertheless, it remains unclear the origin of tumor cells that seeded T5, whether they come from multiple subclones of the primary tumor or independent clones from other metastatic tumors. A related but more complex pattern is observed in patient 135, where CNA events in T12 and T15 (L2) overlap with those found in the L3 brain metastasis branch (T18 and T19), despite no clear direct ancestry (*Figure 1C*). While this could reflect convergent evolution, it also raises the possibility of alternative seeding mechanisms involving contributions from multiple subclonal populations. Resolving this hypothesis would require higher-resolution mapping of CNA breakpoints and integration with external datasets, representing an important direction for future study.

Finally, private subclonal events further contribute to the complexity of tumor evolution. In patient 135, for instance, T4 contains a private subclonal cluster (Cluster 8, $n = 62$, CCF ~ 0.75), while T3 is defined by a private clonal cluster (Cluster 3, $n = 66$, CCF ~ 1) (*Figure 1C*). Although these private clusters do not substantially alter the overall phylogenetic structure, they reflect ongoing diversification within individual lesions. The distinction between clonal and subclonal private events may be influenced by local selective pressures or timing of mutation acquisition. Further investigation into these factors may provide insights into the localized tumor evolution.

6.3 Intertumoral heterogeneity across modalities reveals tumor evolution dynamics and environmental-driven plasticity

Building on the longitudinal genomic and transcriptomic analyses, our results reveal that intertumoral heterogeneity within the same patient evolves differently across molecular layers. While genomic analyses in Aim 1 demonstrate a reduction in mutational heterogeneity over time, reflected by decreasing subclonal mutation proportions, transcriptomic analyses in Aim 2 show a concurrent increase in phenotypic diversity, characterized by

the coexistence of proliferative, invasive, and intermediate melanoma cell states in later samples. In parallel, mutational signature analysis indicates that the diversity of underlying mutational processes increases over time, even as overall genomic diversity decreases. Together, these findings suggest that tumor evolution is not uniformly reflected across genomic and functional dimensions but instead involves a decoupling between genetic diversity and phenotypic complexity.

This apparent paradox can be explained by tumor cell plasticity and microenvironmental influences. As selective pressures such as treatment, immune surveillance, and metastatic niche adaptation reduce genomic diversity through clonal selection, surviving tumor cells may adopt diverse transcriptional states through non-genetic mechanisms, including epigenetic regulation and cell state transitions. Consistent with this model, we observe that tumor cells follow a structured yet non-linear trajectory, transitioning from invasive states at lineage branching points to intermediate and ultimately proliferative states during clonal expansion. This pattern suggests that different evolutionary stages are associated with distinct dominating cell states, while overall heterogeneity is maintained through the coexistence of multiple transitional states within and across lesions. These findings also highlight the importance of integrating multi-modal data to fully capture tumor evolution, as reliance on a single molecular layer may underestimate the complexity of tumor evolution.

6.4 Inter-patient tumor evolution heterogeneity in terms of phylogenetic tree structural variations

Tumor evolutionary trajectories vary substantially across patients, even within the same clinical subgroup. As demonstrated in Chapter 3, patients 118 and 135, both classified as slow progressors, exhibit markedly different genomic landscapes, including differences in the total number of mutations and CNAs, phylogenetic structure, and lineage diversification. Extending this comparison to additional patients within the ITM cohort further highlights this variability. In previously reconstructed phylogenetic trees for patients 136 and 222, work done by the lab post-doc Giuseppe Tarantino and included in a submitted manuscript, Tarantino et al. 2025⁶⁵, both patients display more complex branching structures, with up to seven distinct lineages compared to three in patients 118 and 135 (*Supplementary Figure 12*).

Notably, patient 222, a fast progressor with *NRAS*-mutant acral melanoma, exhibits early and extensive branching events largely independent of treatment timing, with two major diversification waves occurring within the first year after diagnosis (*Supplementary Figure 12A*). This pattern may reflect the distinct biology of acral melanoma, which is less associated with UV exposure and typically exhibits lower tumor mutational burden¹⁰¹. In contrast, patient 136, who harbors a *BRAF V600E* mutation in a lower extremity cutaneous primary tumor, similar to that of patient 135, shows a clearer temporal association between treatment and lineage diversification (*Supplementary Figure 12B*), suggesting that clinical intervention can play a more direct role in shaping tumor evolution in certain contexts.

Despite these differences, some recurring evolutionary patterns can be observed across patients. For example, in both patient 136 and patient 118, early tumor lineages (L1) are eliminated following localized chemotherapy, followed by the emergence of new clones that persist into later disease stages. Similarly, persistent lineages that arise early and survive across multiple timepoints (L2) are observed in both patients, suggesting that certain clones possess intrinsic or acquired resistance to therapy. However, beyond these shared features, phylogenetic structures remain highly variable, particularly in terms of the number of lineages, timing of diversification, and the relationship between branching events and clinical interventions. This variability indicates that even tumors with similar driver mutations or originating from similar anatomical sites can follow distinct evolutionary trajectories.

This high level of inter-patient heterogeneity is further supported by prior cohort-level studies by Spain et al. 2023¹⁰², which demonstrate that melanoma phylogenies can broadly range from linear evolutionary trajectories, consistent with monoclonal seeding⁴³, to highly branched architectures with multiple coexisting subclones, where the number and structure of tumor lineages are independent of melanoma subtype. This implies that factors beyond canonical driver mutations, like microenvironmental context, treatment history, and stochastic evolutionary processes, play critical roles in shaping tumor evolution. Together, these findings highlight that phylogenetic structure is highly patient-specific, and that no single evolutionary model fully captures the diversity of melanoma progression. Understanding this variability will be essential for developing personalized therapeutic strategies and for interpreting tumor evolution in both clinical and research settings.

6.5 Genomic features of melanoma brain metastasis and limitations of SBS mutational signature interpretation

Melanoma brain metastasis is associated with poor prognosis, with median survival of approximately 4-6 months¹⁰³⁻¹⁰⁴, highlighting the importance of understanding the genomic features and risk factors underlying its development. A substantial proportion (40–60%) of melanoma brain metastases harbor *BRAF* mutations¹⁰⁵, and patient outcomes can be improved through combined ICB and BRAF/MEK-targeted therapies¹⁰⁶. Patient 135 from the ITM cohort, who carries a *BRAF V600E* mutation and developed brain metastases at Stage IV disease, is therefore representative of this clinically relevant subgroup.

Notably, *BRAF*-mutant melanomas, especially the ones with *BRAF V600E*, have been reported to be associated with relatively low tumor mutational burden (TMB)¹⁰⁷, and brain tumors more broadly are characterized by predominantly low TMB profiles (~90%)¹⁰⁸. The two brain metastatic samples, T18 and T19, from patient 135 indeed show a low TMB profile (TMB = 5.124 and 6.235, respectively), in accordance with the general expectations given that high TMB is commonly defined by > 10 mut/Mb. However, this observation does not support a model in which patient 135's brain metastasis was driven by the gain of Cluster 0 mutations (n=57) (*Figure 1C*) and a greater mutational distance from L3 to other lineages (*Figure 1D*), suggesting that alternative mechanisms may underlie the genomic features observed in patient 135.

In our analysis, the lineage L3 that contains the brain metastasis in patient 135 is characterized by a distinct mutational signature dominated by polymerase η -associated activity (SBS9) (*Figure 2A*), a process classically linked to somatic hypermutation in B lymphocytes⁹². While this signal is strongly and uniquely restricted to L3 samples, its biological interpretation in the context of melanoma tumor cells is unclear. Polymerase η -associated somatic hypermutation arises from error-prone repair of AID (activation-induced cytidine deaminase)-induced lesions in B cells, generating characteristic mutations enriched at the A or T bases¹⁰⁹. It is not typically associated with melanoma progression or brain metastasis, raising the possibility that this signature may not reflect a tumor-intrinsic mutagenic process.

This observation highlights an important limitation of the SBS mutational signature analysis. Mutational signatures are inferred based on trinucleotide substitution patterns, and multiple biological processes can generate overlapping or highly similar patterns. For example, both UV light exposure and temozolomide-associated signatures are enriched for C>T substitutions in similar trinucleotide contexts and both have a high frequency of

TCC>TTC trinucleotide alterations, making them difficult to distinguish computationally. As a result, certain signatures may be misattributed, particularly in tumors like melanoma where UV-induced mutations dominate the mutational landscape.

Consistent with this limitation, we observed the presence of chemotherapy-associated signatures, including temozolomide (alkylating agent) and chemotherapy exposure signatures (*Figure 2A-B*), in samples collected prior to treatment or in patients who did not receive such therapies. These biologically implausible assignments suggest that some extracted signatures represent artifacts or noise rather than true underlying mutational processes. Similarly, the polymerase η -associated signal observed in L3 may reflect either mis-annotation, contamination from immune cell-derived mutations, or unresolved overlap with other mutational processes.

In addition, the reliability of mutational signature analysis is further constrained by the number of mutations available for inference. Lineage-defining clusters with low mutation counts cannot support robust signature extraction, limiting interpretation for clusters such as Cluster 0, 4, and 5 in patient 118 and Cluster 1 in patient 135 (*Supplementary Figure A, E, Figure 2C-D*). Although Cluster 0 in patient 135 contains sufficient mutations ($n = 57$) for analysis, the resulting signature profile remains difficult to interpret biologically, further emphasizing that even moderate mutation counts may not fully resolve underlying mutational processes.

To mitigate potential mis-annotation, we supplemented computational signature extraction with direct visualization of trinucleotide mutation patterns and compared them to the reference SBS profiles⁹². This step provides an additional layer of validation and helps ensure that inferred signatures are supported by observable mutation patterns rather than purely algorithmic decomposition.

6.6 Additional limitations of the study and future steps

A key limitation of this study is the relatively small sample size for longitudinal genomic analysis. While longitudinal sampling enables detailed reconstruction of tumor evolutionary trajectories within individual patients, acquiring sufficiently dense time-series data across many patients remains clinically challenging. As a result, this study prioritizes depth over breadth, focusing on two representative melanoma patients with extensive sampling.

To address scalability in future work, computational standardization will be critical. Developing an automated and reproducible pipeline that integrates multiple tools for mutation and CNA calling, clonal deconvolution, and phylogenetic reconstruction would enable more consistent cross-patient comparisons. However,

even with improved computational frameworks, cross-study integration remains inherently difficult due to the high degree of inter-patient heterogeneity in tumor evolution. Differences in treatment regimens, timing, dosage, and individual patient response further complicate direct comparisons, limiting the ability to generalize evolutionary patterns across cohorts.

In addition, while the novel tumor programs identified in this study demonstrate strong biological interpretability and potential clinical relevance, further validation is required before broader application. These programs, defined using coordinated enrichment of human hallmark pathways and melanoma signatures, provide a more robust framework for capturing tumor-intrinsic states compared to single gene or solely expression-based approaches. Future work will focus on validating these tumor programs across modalities and in independent cohorts. Due to the availability of spatial-omics data in the HTAN cohort, including MERFISH, CODEX, and Lunaphore, we can validate the identified tumor programs in the spatial context and evaluate their role in the tumor-immune niches and recurrent cellular neighborhoods (RCN). This validation will not only strengthen the biological relevance of these programs but also enable their use in studying tumor-microenvironment interactions *in situ*.

In summary, this thesis work demonstrates that melanoma progression is shaped by the interplay between genomic evolution, phenotypic plasticity, and tumor-immune interactions. By integrating longitudinal genomic profiling with transcriptomic analyses, we revealed that decreasing genomic heterogeneity can coexist with increasing phenotypic and mutational process diversity over time. The identification of biologically interpretable tumor programs further provides a scalable framework for characterizing tumor states across patients and modalities. Together, these findings offer new insights into melanoma evolution and lay the groundwork for future studies for improving therapeutic targeting and patient prognosis.

References

AI Usage: This thesis was revised with the assistance of ChatGPT-5.3 for grammar refinement and clarity enhancement only. All intellectual contributions, including study design, analysis framework, and scientific interpretations, are solely those of the author.

- [1] Rozeman, E. A., Dekker, T. J., Haanen, J. B., & Blank, C. U. (2017). Advanced melanoma: Current treatment options, biomarkers, and future perspectives. *American Journal of Clinical Dermatology*, 19(3), 303–317. <https://doi.org/10.1007/s40257-017-0325-6>
- [2] Grzywa, T. M., Paskal, W., & Włodarski, P. K. (2017). Intratumor and intertumor heterogeneity in melanoma. *Translational Oncology*, 10(6), 956–975. <https://doi.org/10.1016/j.tranon.2017.09.007>
- [3] Caswell, D. R., & Swanton, C. (2017). The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Medicine*, 15(1). <https://doi.org/10.1186/s12916-017-0900-y>
- [4] Gershenwald, J. E., Scolyer, R. A., Hess, K. R., Sondak, V. K., Long, G. V., Ross, M. I., Lazar, A. J., Faries, M. B., Kirkwood, J. M., McArthur, G. A., Haydu, L. E., Eggermont, A. M., Flaherty, K. T., Balch, C. M., & Thompson, J. F. (2017). Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer Eighth Edition Cancer Staging Manual. *CA: A Cancer Journal for Clinicians*, 67(6), 472–492. <https://doi.org/10.3322/caac.21409>
- [5] Tas, F. (2012). Metastatic behavior in melanoma: Timing, pattern, survival, and influencing factors. *Journal of Oncology*, 2012, 1–9. <https://doi.org/10.1155/2012/647684>
- [6] Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J. C., Akerley, W., van den Eertwegh, A. J. M., Lutzky, J., Lorigan, P., Vaubel, J. M., Linette, G. P., Hogg, D., Ottensmeier, C. H., Lebbé, C., ... Urban, W. J. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *New England Journal of Medicine*, 363(8), 711–723. <https://doi.org/10.1056/nejmoa1003466>
- [7] Wolchok, J. D., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J.-J., Cowey, C. L., Lao, C. D., Wagstaff, J., Schadendorf, D., Ferrucci, P. F., Smylie, M., Dummer, R., Hill, A., Hogg, D., Haanen, J., Carlino, M. S., Bechter, O., Maio, M., Marquez-Rodas, I., ... Larkin, J. (2017). Overall survival with combined nivolumab and Ipilimumab in advanced melanoma. *New England Journal of Medicine*, 377(14), 1345–1356. <https://doi.org/10.1056/nejmoa1709684>
- [8] Harris-Bookman, S., Mathios, D., Martin, A. M., Xia, Y., Kim, E., Xu, H., Belcaid, Z., Polanczyk, M., Barberi, T., Theodoros, D., Kim, J., Taube, J. M., Burger, P. C., Selby, M., Taitt, C., Korman, A., Ye, X., Drake, C. G., Brem, H., ... Lim, M. (2018). Expression of lag-3 and efficacy of combination treatment with anti-lag-3 and anti-pd-1 monoclonal antibodies in glioblastoma. *International Journal of Cancer*, 143(12), 3201–3208. <https://doi.org/10.1002/ijc.31661>
- [9] Tawbi, H. A., Schadendorf, D., Lipson, E. J., Ascierto, P. A., Matamala, L., Castillo Gutiérrez, E., Rutkowski, P., Gogas, H. J., Lao, C. D., De Menezes, J. J., Dalle, S., Arance, A., Grob, J.-J., Srivastava, S., Abaskharoun, M., Hamilton, M., Keidel, S., Simonsen, K. L., Sobieski, A. M., ... Long, G. V. (2022). Relatlimab and nivolumab versus nivolumab in untreated advanced melanoma. *New England Journal of Medicine*, 386(1), 24–34. <https://doi.org/10.1056/nejmoa2109970>
- [10] Flaherty, K. T., Infante, J. R., Daud, A., Gonzalez, R., Kefford, R. F., Sosman, J., Hamid, O., Schuchter, L., Cebon, J., Ibrahim, N., Kudchadkar, R., Burris, H. A., Falchook, G., Algazi, A., Lewis, K., Long, G. V., Puzanov, I., Lebowitz, P., Singh, A., ... Weber, J. (2012). Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *New England Journal of Medicine*, 367(18), 1694–1703. <https://doi.org/10.1056/nejmoa1210093>

- [11] Boutros, A., Croce, E., Ferrari, M., Gili, R., Massaro, G., Marconcini, R., Arecco, L., Tanda, E. T., & Spagnolo, F. (2024). The treatment of advanced melanoma: Current approaches and new challenges. *Critical Reviews in Oncology/Hematology*, 196, 104276. <https://doi.org/10.1016/j.critrevonc.2024.104276>
- [12] Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J.-J., Rutkowski, P., Lao, C. D., Cowey, C. L., Schadendorf, D., Wagstaff, J., Dummer, R., Ferrucci, P. F., Smylie, M., Hogg, D., Hill, A., Márquez-Rodas, I., Haanen, J., Guidoboni, M., Maio, M., Schöffski, P., ... Wolchok, J. D. (2019). Five-year survival with combined nivolumab and Ipilimumab in advanced melanoma. *New England Journal of Medicine*, 381(16), 1535–1546. <https://doi.org/10.1056/nejmoa1910836>
- [13] Robert, C., Schachter, J., Long, G. V., Arance, A., Grob, J. J., Mortier, L., Daud, A., Carlino, M. S., McNeil, C., Lotem, M., Larkin, J., Lorigan, P., Neyns, B., Blank, C. U., Hamid, O., Mateus, C., Shapira-Frommer, R., Kosh, M., Zhou, H., ... Ribas, A. (2015). Pembrolizumab versus Ipilimumab in advanced melanoma. *New England Journal of Medicine*, 372(26), 2521–2532. <https://doi.org/10.1056/nejmoa1503093>
- [14] Menzies, A. M., Haydu, L. E., Carlino, M. S., Azer, M. W., Carr, P. J., Kefford, R. F., & Long, G. V. (2014). Inter- and intra-patient heterogeneity of response and progression to targeted therapy in metastatic melanoma. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0085004>
- [15] Marusyk, A., & Polyak, K. (2010). Tumor heterogeneity: Causes and consequences. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1805(1), 105–117. <https://doi.org/10.1016/j.bbcan.2009.11.002>
- [16] Wilmott, J. S., Tembe, V., Howle, J. R., Sharma, R., Thompson, J. F., Rizos, H., Lo, R. S., Kefford, R. F., Scolyer, R. A., & Long, G. V. (2012). Intratumoral molecular heterogeneity in a *braf*-mutant, BRAF inhibitor-resistant melanoma: A case illustrating the challenges for Personalized Medicine. *Molecular Cancer Therapeutics*, 11(12), 2704–2708. <https://doi.org/10.1158/1535-7163.mct-12-0530>
- [17] Nowell, P. C. (1976). The clonal evolution of Tumor Cell Populations. *Science*, 194(4260), 23–28. <https://doi.org/10.1126/science.959840>
- [18] Galon, J., & Bruni, D. (2020). Tumor immunology and tumor evolution: Intertwined histories. *Immunity*, 52(1), 55–81. <https://doi.org/10.1016/j.immuni.2019.12.018>
- [19] Lloyd, M. C., Cunningham, J. J., Bui, M. M., Gillies, R. J., Brown, J. S., & Gatenby, R. A. (2016). Darwinian dynamics of intratumoral heterogeneity: Not solely random mutations but also variable environmental selection forces. *Cancer Research*, 76(11), 3136–3144. <https://doi.org/10.1158/0008-5472.can-15-2962>
- [20] Angelova, M., Mlecnik, B., Vasaturo, A., Bindea, G., Fredriksen, T., Lafontaine, L., Buttard, B., Morgand, E., Bruni, D., Jouret-Mourin, A., Hubert, C., Kartheuser, A., Humblet, Y., Ceccarelli, M., Syed, N., Marincola, F. M., Bedognetti, D., Van den Eynde, M., & Galon, J. (2018). Evolution of metastases in space and time under immune selection. *Cell*, 175(3). <https://doi.org/10.1016/j.cell.2018.09.018>
- [21] Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C. R., ... Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*, 366(10), 883–892. <https://doi.org/10.1056/nejmoa1113205>
- [22] Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., Powderly, J. D., Carvajal, R. D., Sosman, J. A., Atkins, M. B., Leming, P. D., Spigel, D. R., Antonia, S. J., Horn, L., Drake, C. G., Pardoll, D. M., Chen, L., Sharfman, W. H., Anders, R. A., ... Sznol, M. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *New England Journal of Medicine*, 366(26), 2443–2454. <https://doi.org/10.1056/nejmoa1200690>
- [23] Mellman, I., Coukos, G., & Dranoff, G. (2011). Cancer immunotherapy comes of age. *Nature*, 480(7378), 480–489. <https://doi.org/10.1038/nature10673>
- [24] Drake, C. G., Jaffee, E., & Pardoll, D. M. (2006). Mechanisms of immune evasion by tumors. *Advances in Immunology*, 51–81. [https://doi.org/10.1016/s0065-2776\(06\)90002-9](https://doi.org/10.1016/s0065-2776(06)90002-9)
- [25] Mc Adam, A. J., Schweitzer, A. N., & Sharpe, A. H. (1998). The role of B7 Co-stimulation in activation and differentiation of cd4+ and CD8+ T cells. *Immunological Reviews*, 165(1), 231–247. <https://doi.org/10.1111/j.1600-065x.1998.tb01242.x>

- [26] Alegre, M.-L., Frauwirth, K. A., & Thompson, C. B. (2001). T-cell regulation by CD28 and CTLA-4. *Nature Reviews Immunology*, 1(3), 220–228. <https://doi.org/10.1038/35105024>
- [27] Sharpe, A. H., & Pauken, K. E. (2017). The diverse functions of the PD1 inhibitory pathway. *Nature Reviews Immunology*, 18(3), 153–167. <https://doi.org/10.1038/nri.2017.108>
- [28] Pardoll, D. M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer*, 12(4), 252–264. <https://doi.org/10.1038/nrc3239>
- [29] Chambers, C. A., Kuhns, M. S., Egen, J. G., & Allison, J. P. (2001). CTLA-4-mediated inhibition in regulation of T cell responses: Mechanisms and manipulation in tumor immunotherapy. *Annual Review of Immunology*, 19(1), 565–594. <https://doi.org/10.1146/annurev.immunol.19.1.565>
- [30] Jiang, Y., Chen, M., Nie, H., & Yuan, Y. (2019). PD-1 and PD-L1 in cancer immunotherapy: Clinical implications and future considerations. *Human Vaccines & Immunotherapeutics*, 15(5), 1111–1122. <https://doi.org/10.1080/21645515.2019.1571892>
- [31] Tumeh, P. C., Harview, C. L., Yearley, J. H., Shintaku, I. P., Taylor, E. J., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., West, A. N., Carmona, M., Kivork, C., Seja, E., Cherry, G., Gutierrez, A. J., Grogan, T. R., Mateus, C., Tomasic, G., ... Ribas, A. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*, 515(7528), 568–571. <https://doi.org/10.1038/nature13954>
- [32] Sharma, P., Hu-Lieskovan, S., Wargo, J. A., & Ribas, A. (2017). Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell*, 168(4), 707–723. <https://doi.org/10.1016/j.cell.2017.01.017>
- [33] George, S., Miao, D., Demetri, G. D., Adeegbe, D., Rodig, S. J., Shukla, S., Lipschitz, M., Amin-Mansour, A., Raut, C. P., Carter, S. L., Hammerman, P., Freeman, G. J., Wu, C. J., Ott, P. A., Wong, K.-K., & Van Allen, E. M. (2017). Loss of PTEN is associated with resistance to Anti-PD-1 checkpoint blockade therapy in metastatic uterine leiomyosarcoma. *Immunity*, 46(2), 197–204. <https://doi.org/10.1016/j.immuni.2017.02.001>
- [34] Peng, W., Chen, J. Q., Liu, C., Malu, S., Creasy, C., Tetzlaff, M. T., Xu, C., McKenzie, J. A., Zhang, C., Liang, X., Williams, L. J., Deng, W., Chen, G., Mbofung, R., Lazar, A. J., Torres-Cabala, C. A., Cooper, Z. A., Chen, P.-L., Tieu, T. N., ... Hwu, P. (2016). Loss of PTEN promotes resistance to T cell-mediated immunotherapy. *Cancer Discovery*, 6(2), 202–216. <https://doi.org/10.1158/2159-8290.cd-15-0283>
- [35] Zaretsky, J. M., Garcia-Diaz, A., Shin, D. S., Escuin-Ordinas, H., Hugo, W., Hu-Lieskovan, S., Torrejon, D. Y., Abril-Rodriguez, G., Sandoval, S., Barthly, L., Saco, J., Homet Moreno, B., Mezzadra, R., Chmielowski, B., Ruchalski, K., Shintaku, I. P., Sanchez, P. J., Puig-Saus, C., Cherry, G., ... Ribas, A. (2016). Mutations associated with acquired resistance to PD-1 blockade in melanoma. *New England Journal of Medicine*, 375(9), 819–829. <https://doi.org/10.1056/nejmoa1604958>
- [36] Gao, J., Shi, L. Z., Zhao, H., Chen, J., Xiong, L., He, Q., Chen, T., Roszik, J., Bernatchez, C., Woodman, S. E., Chen, P.-L., Hwu, P., Allison, J. P., Futreal, A., Wargo, J. A., & Sharma, P. (2016). Loss of IFN- γ pathway genes in tumor cells as a mechanism of resistance to anti-CTLA-4 therapy. *Cell*, 167(2). <https://doi.org/10.1016/j.cell.2016.08.069>
- [37] Sade-Feldman, M., Jiao, Y. J., Chen, J. H., Rooney, M. S., Barzily-Rokni, M., Eliane, J.-P., Bjorgaard, S. L., Hammond, M. R., Vitzthum, H., Blackmon, S. M., Frederick, D. T., Hazar-Rethinam, M., Nades, B. A., Van Seventer, E. E., Shukla, S. A., Yizhak, K., Ray, J. P., Rosebrock, D., Livitz, D., ... Hacohen, N. (2017). Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01062-w>
- [38] Roh, W., Chen, P.-L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., Gopalakrishnan, V., Wang, F., Cooper, Z. A., Reddy, S. M., Gumbs, C., Little, L., Chang, Q., Chen, W.-S., Wani, K., De Macedo, M. P., Chen, E., Austin-Breneman, J. L., Jiang, H., ... Futreal, P. A. (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Science Translational Medicine*, 9(379). <https://doi.org/10.1126/scitranslmed.aah3560>
- [39] Liu, D., Lin, J.-R., Robitschek, E. J., Kasumova, G. G., Heyde, A., Shi, A., Kraya, A., Zhang, G., Moll, T., Frederick, D. T., Chen, Y.-A., Wang, S., Schapiro, D., Ho, L.-L., Bi, K., Sahu, A., Mei, S., Miao, B., Sharova, T., ... Boland, G. M. (2021). Evolution of delayed resistance to immunotherapy in a melanoma responder. *Nature Medicine*, 27(6), 985–992. <https://doi.org/10.1038/s41591-021-01331-8>

- [40] Heyde, A., Reiter, J. G., Naxerova, K., & Nowak, M. A. (2019). Consecutive seeding and transfer of genetic diversity in metastasis. *Proceedings of the National Academy of Sciences*, 116(28), 14129–14137. <https://doi.org/10.1073/pnas.1819408116>
- [41] Wei, Q., Ye, Z., Zhong, X., Li, L., Wang, C., Myers, R. E., Palazzo, J. P., Fortuna, D., Yan, A., Waldman, S. A., Chen, X., Posey, J. A., Basu-Mallick, A., Jiang, B. H., Hou, L., Shu, J., Sun, Y., Xing, J., Li, B., & Yang, H. (2017). Multiregion whole-exome sequencing of matched primary and metastatic tumors revealed genomic heterogeneity and suggested polyclonal seeding in colorectal cancer metastasis. *Annals of Oncology*, 28(9), 2135–2141. <https://doi.org/10.1093/annonc/mdx278>
- [42] McPherson, A., Roth, A., Laks, E., Masud, T., Bashashati, A., Zhang, A. W., Ha, G., Biele, J., Yap, D., Wan, A., Prentice, L. M., Khattra, J., Smith, M. A., Nielsen, C. B., Mullaly, S. C., Kalloger, S., Karnezis, A., Shumansky, K., Siu, C., ... Shah, S. P. (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nature Genetics*, 48(7), 758–767. <https://doi.org/10.1038/ng.3573>
- [43] Talmadge, J. E., Wolman, S. R., & Fidler, I. J. (1982). Evidence for the clonal origin of spontaneous metastases. *Science*, 217(4557), 361–363. <https://doi.org/10.1126/science.6953592>
- [44] Marine, J.-C., Dawson, S.-J., & Dawson, M. A. (2020). Non-genetic mechanisms of therapeutic resistance in cancer. *Nature Reviews Cancer*, 20(12), 743–756. <https://doi.org/10.1038/s41568-020-00302-4>
- [45] Tsoi, J., Robert, L., Paraiso, K., Galvan, C., Sheu, K. M., Lay, J., Wong, D. J. L., Atefi, M., Shirazi, R., Wang, X., Braas, D., Grasso, C. S., Palaskas, N., Ribas, A., & Graeber, T. G. (2018). Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell*, 33(5). <https://doi.org/10.1016/j.ccell.2018.03.017>
- [46] Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G., ... Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, 352(6282), 189–196. <https://doi.org/10.1126/science.aad0501>
- [47] Rambow, F., Rogiers, A., Marin-Bejar, O., Aibar, S., Femel, J., Dewaele, M., Karras, P., Brown, D., Chang, Y. H., Debiec-Rychter, M., Adriaens, C., Radaelli, E., Wolter, P., Bechter, O., Dummer, R., Levesque, M., Piris, A., Frederick, D. T., Boland, G., ... Marine, J.-C. (2018). Toward minimal residual disease-directed therapy in melanoma. *Cell*, 174(4). <https://doi.org/10.1016/j.cell.2018.06.025>
- [48] Hoek, K. S., Schlegel, N. C., Brafford, P., Sucker, A., Ugurel, S., Kumar, R., Weber, B. L., Nathanson, K. L., Phillips, D. J., Herlyn, M., Schadendorf, D., & Dummer, R. (2006). Metastatic potential of melanomas defined by specific gene expression profiles with no BRAF signature. *Pigment Cell Research*, 19(4), 290–302. <https://doi.org/10.1111/j.1600-0749.2006.00322.x>
- [49] Verfaillie, A., Imrichova, H., Atak, Z. K., Dewaele, M., Rambow, F., Hulselmans, G., Christiaens, V., Svetlichnyy, D., Luciani, F., Van den Mooter, L., Claerhout, S., Fiers, M., Journe, F., Ghanem, G.-E., Herrmann, C., Halder, G., Marine, J.-C., & Aerts, S. (2015). Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the Invasive Cell State. *Nature Communications*, 6(1). <https://doi.org/10.1038/ncomms7683>
- [50] Wouters, J., Kalender-Atak, Z., Minnoye, L., Spanier, K. I., De Waegeneer, M., Bravo González-Blas, C., Mauduit, D., Davie, K., Hulselmans, G., Najem, A., Dewaele, M., Pedri, D., Rambow, F., Makhzami, S., Christiaens, V., Ceysens, F., Ghanem, G., Marine, J.-C., Poovathingal, S., & Aerts, S. (2020). Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nature Cell Biology*, 22(8), 986–998. <https://doi.org/10.1038/s41556-020-0547-3>
- [51] Baron, M., Tagore, M., Hunter, M. V., Kim, I. S., Moncada, R., Yan, Y., Campbell, N. R., White, R. M., & Yanai, I. (2020). The stress-like cancer cell state is a consistent component of tumorigenesis. *Cell Systems*, 11(5). <https://doi.org/10.1016/j.cels.2020.08.018>
- [52] Biermann, J., Melms, J. C., Amin, A. D., Wang, Y., Caprio, L. A., Karz, A., Tagore, S., Barrera, I., Ibarra-Arellano, M. A., Andreatta, M., Fullerton, B. T., Gretarsson, K. H., Sahu, V., Mangipudy, V. S., Nguyen, T. T. T., Nair, A., Rogava, M., Ho, P., Koch, P. D., ... Izar, B. (2022). Dissecting the treatment-naïve ecosystem of human melanoma brain metastasis. *Cell*, 185(14). <https://doi.org/10.1016/j.cell.2022.06.007>

- [53] Jeffs, A. R., Glover, A. C., Slobbe, L. J., Wang, L., He, S., Hazlett, J. A., Awasthi, A., G. Woolley, A., Marshall, E. S., Joseph, W. R., Print, C. G., Baguley, B. C., & Eccles, M. R. (2009). A gene expression signature of invasive potential in metastatic melanoma cells. *PLoS ONE*, 4(12). <https://doi.org/10.1371/journal.pone.0008461>
- [54] Kauffmann, A., Rosselli, F., Lazar, V., Winnepeninckx, V., Mansuet-Lupo, A., Dessen, P., van den Oord, J. J., Spatz, A., & Sarasin, A. (2007). High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene*, 27(5), 565–573. <https://doi.org/10.1038/sj.onc.1210700>
- [55] Widmer, D. S., Cheng, P. F., Eichhoff, O. M., Belloni, B. C., Zipser, M. C., Schlegel, N. C., Javelaud, D., Mauviel, A., Dummer, R., & Hoek, K. S. (2012). Systematic classification of melanoma cells by phenotype-specific gene expression mapping. *Pigment Cell & Melanoma Research*, 25(3), 343–353. <https://doi.org/10.1111/j.1755-148x.2012.00986.x>
- [56] McNeal, A. S., Belote, R. L., Zeng, H., Urquijo, M., Barker, K., Torres, R., Curtin, M., Shain, A. H., Andtbacka, R. H., Holmen, S., Lum, D. H., McCalmont, T. H., VanBrocklin, M. W., Grossman, D., Wei, M. L., Lang, U. E., & Judson-Torres, R. L. (2021). BRAFV600E induces reversible mitotic arrest in human melanocytes via microrna-mediated suppression of AURKB. *eLife*, 10. <https://doi.org/10.7554/elife.70385>
- [57] Ryu, B., Moriarty, W. F., Stine, M. J., DeLuca, A., Kim, D. S., Meeker, A. K., Grills, L. D., Switzer, R. A., Eller, M. S., & Alani, R. M. (2011). Global analysis of brafv600e target genes in human melanocytes identifies matrix metalloproteinase-1 as a critical mediator of melanoma growth. *Journal of Investigative Dermatology*, 131(7), 1579–1583. <https://doi.org/10.1038/jid.2011.65>
- [58] Belote, R. L., Le, D., Maynard, A., Lang, U. E., Sinclair, A., Lohman, B. K., Planells-Palop, V., Baskin, L., Tward, A. D., Darmanis, S., & Judson-Torres, R. L. (2021). Human melanocyte development and melanoma dedifferentiation at Single-Cell resolution. *Nature Cell Biology*, 23(9), 1035–1047. <https://doi.org/10.1038/s41556-021-00740-8>
- [59] Andrews, M. C., Oba, J., Wu, C.-J., Zhu, H., Karpinets, T., Creasy, C. A., Forget, M.-A., Yu, X., Song, X., Mao, X., Robertson, A. G., Romano, G., Li, P., Burton, E. M., Lu, Y., Sloane, R. S., Wani, K. M., Rai, K., Lazar, A. J., ... Woodman, S. E. (2022). Multi-modal molecular programs regulate melanoma cell state. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-31510-1>
- [60] Riesenberger, S., Groetchen, A., Siddaway, R., Bald, T., Reinhardt, J., Smorra, D., Kohlmeyer, J., Renn, M., Phung, B., Aymans, P., Schmidt, T., Hornung, V., Davidson, I., Goding, C. R., Jönsson, G., Landsberg, J., Tüting, T., & Hölzel, M. (2015). MITF and c-jun antagonism interconnects melanoma dedifferentiation with pro-inflammatory cytokine responsiveness and myeloid cell recruitment. *Nature Communications*, 6(1). <https://doi.org/10.1038/ncomms9755>
- [61] Pozniak, J., Pedri, D., Landeloos, E., Van Herck, Y., Antoranz, A., Vanwynsberghe, L., Nowosad, A., Roda, N., Makhzami, S., Bervoets, G., Maciel, L. F., Pulido-Vicuña, C. A., Pollaris, L., Seurinck, R., Zhao, F., Flem-Karlsen, K., Damsky, W., Chen, L., Karagianni, D., ... Marine, J.-C. (2024). A TCF4-dependent gene regulatory network confers resistance to immunotherapy in melanoma. *Cell*, 187(1). <https://doi.org/10.1016/j.cell.2023.11.037>
- [62] Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, 1(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
- [63] Rozenblatt-Rosen, O., Regev, A., Oberdoerffer, P., Nawy, T., Hupalowska, A., Rood, J. E., Ashenberg, O., Cerami, E., Coffey, R. J., Demir, E., Ding, L., Esplin, E. D., Ford, J. M., Goecks, J., Ghosh, S., Gray, J. W., Guinney, J., Hanlon, S. E., Hughes, S. K., ... Zhuang, X. (2020). The Human Tumor Atlas Network: Charting tumor transitions across space and time at single-cell resolution. *Cell*, 181(2), 236–249. <https://doi.org/10.1016/j.cell.2020.03.053>
- [64] de Bruijn, I., Nikolov, M., Lau, C., Clayton, A., Gibbs, D. L., Mitraka, E., Pozhidayeva, D., Lash, A., Sumer, S. O., Altreuter, J., Anton, K., DeFelice, M., Li, X., Lisman, A., Longabaugh, W. J., Muhlich, J., Santagata, S., Nandakumar, S., Sorger, P. K., ... Eddy, J. A. (2025). Sharing data from the human tumor atlas network through standards, infrastructure and community engagement. *Nature Methods*, 22(4), 664–671. <https://doi.org/10.1038/s41592-025-02643-0>

- [65] Tarantino, G., Zaremba, A., Vallius, T., Woodnorth, M., Hua, Y., Pelletier, R., Lopez Leon, M., Shi, Y., Maliga, Z., Makhzami, S., Aprati, T. J., Karlaš, B., Glutsch, V., Schilling, B., Hassel, J., Berking, C., Utikal, J., Meier, F., Meiss, F., ... Schadendorf, D. (2025). *Multitomic Profiling of a Unique In-Transit Melanoma Cohort Identifies Melanoma Differentiation as Predictor of Tumor Progression and Therapy Response*. <https://doi.org/10.1101/2025.10.01.679282>
- [66] Liu, D., Abbosh, P., Keliher, D., Reardon, B., Miao, D., Mouw, K., Weiner-Taylor, A., Wankowicz, S., Han, G., Teo, M. Y., Cipolla, C., Kim, J., Iyer, G., Al-Ahmadie, H., Dulaimi, E., Chen, D. Y., Alpaugh, R. K., Hoffman-Censits, J., Garraway, L. A., ... Van Allen, E. M. (2017). Mutational patterns in chemotherapy resistant muscle-invasive bladder cancer. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-02320-7>
- [67] Van Allen, E. M., Wagle, N., Stojanov, P., Perrin, D. L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D. C., Kryukov, G., Carter, S. L., McKenna, A., Sivachenko, A., Rosenberg, M., Kiezun, A., Voet, D., Lawrence, M., Lichtenstein, L. T., Gentry, J. G., Huang, F. W., ... Garraway, L. A. (2014). Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide Precision Cancer Medicine. *Nature Medicine*, 20(6), 682–688. <https://doi.org/10.1038/nm.3559>
- [68] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- [69] Cibulskis, K., McKenna, A., Fennell, T., Banks, E., DePristo, M., & Getz, G. (2011). Contest: Estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics*, 27(18), 2601–2602. <https://doi.org/10.1093/bioinformatics/btr446>
- [70] Taylor-Weiner, A., Stewart, C., Giordano, T., Miller, M., Rosenberg, M., Macbeth, A., Lennon, N., Rheinbay, E., Landau, D.-A., Wu, C. J., & Getz, G. (2018). Detin: Overcoming tumor-in-normal contamination. *Nature Methods*, 15(7), 531–534. <https://doi.org/10.1038/s41592-018-0036-9>
- [71] GATK. https://software.broadinstitute.org/gatk/documentation/tooldocs/4.0.1.0/picard_fingerprint_CrosscheckFingerprints.php
- [72] GATK. https://software.broadinstitute.org/gatk/documentation/tooldocs/4.0.0.0/picard_analysis_CollectMultipleMetrics.php
- [73] Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S., & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3), 213–219. <https://doi.org/10.1038/nbt.2514>
- [74] Saunders, C. T., Wong, W. S., Swamy, S., Becq, J., Murray, L. J., & Cheetham, R. K. (2012). Strelka: Accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics*, 28(14), 1811–1817. <https://doi.org/10.1093/bioinformatics/bts271>
- [75] Landau, D. A., Carter, S. L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M. S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., Wan, Y., Zhang, W., Shukla, S. A., Vartanov, A., Fernandes, S. M., Saksena, G., Cibulskis, K., Tesar, B., Gabriel, S., ... Wu, C. J. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*, 152(4), 714–726. <https://doi.org/10.1016/j.cell.2013.01.019>
- [76] Ramos, A. H., Lichtenstein, L., Gupta, M., Lawrence, M. S., Pugh, T. J., Saksena, G., Meyerson, M., & Getz, G. (2015). Oncotator: Cancer variant annotation tool. *Human Mutation*, 36(4). <https://doi.org/10.1002/humu.22771>
- [77] Costello, M., Pugh, T. J., Fennell, T. J., Stewart, C., Lichtenstein, L., Meldrim, J. C., Fostel, J. L., Friedrich, D. C., Perrin, D., Dionne, D., Kim, S., Gabriel, S. B., Lander, E. S., Fisher, S., & Getz, G. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research*, 41(6). <https://doi.org/10.1093/nar/gks1443>
- [78] Chakravarty, D., Gao, J., Phillips, S., Kundra, R., Zhang, H., Wang, J., Rudolph, J. E., Yaeger, R., Soumerai, T., Nissan, M. H., Chang, M. T., Chandarlapaty, S., Traina, T. A., Paik, P. K., Ho, A. L., Hantash, F. M., Grupe, A., Baxi, S. S., Callahan, M. K., ... Schultz, N. (2017). OncoKB: A precision oncology knowledge base. *JCO Precision Oncology*, (1), 1–16. <https://doi.org/10.1200/po.17.00011>

- [79] Tordai, H., Torres, O., Csepi, M., Padányi, R., Lukács, G. L., & Hegedűs, T. (2024). Analysis of Alphamissense data in different protein groups and structural context. *Scientific Data*, 11(1). <https://doi.org/10.1038/s41597-024-03327-8>
- [80] Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The Cbio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5), 401–404. <https://doi.org/10.1158/2159-8290.cd-12-0095>
- [81] Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative Analysis of Complex Cancer Genomics and clinical profiles using the cBioPortal. *Science Signaling*, 6(269). <https://doi.org/10.1126/scisignal.2004088>
- [82] de Bruijn, I., Kundra, R., Mastrogiacomio, B., Tran, T. N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., Abeshouse, A., Baiceanu, D., Ciftci, E., Dogrusoz, U., Dufilie, A., Erkoc, Z., Garcia Lara, E., Fu, Z., Gross, B., ... Schultz, N. (2023). Analysis and visualization of longitudinal genomic and clinical data from the AACR project genie biopharma collaborative in cBioPortal. *Cancer Research*, 83(23), 3861–3867. <https://doi.org/10.1158/0008-5472.can-23-0816>
- [83] Akbani, R., Akdemir, K. C., Aksoy, B. A., Albert, M., Ally, A., Amin, S. B., Arachchi, H., Arora, A., Auman, J. T., Ayala, B., Baboud, J., Balasundaram, M., Balu, S., Barnabas, N., Bartlett, J., Bartlett, P., Bastian, B. C., Baylin, S. B., Behera, M., ... Zou, L. (2015j). Genomic classification of cutaneous melanoma. *Cell*, 161(7), 1681–1696. <https://doi.org/10.1016/j.cell.2015.05.044>
- [84] Carter, S. L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P. W., Onofrio, R. C., Winckler, W., Weir, B. A., Beroukhi, R., Pellman, D., Levine, D. A., Lander, E. S., Meyerson, M., & Getz, G. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnology*, 30(5), 413–421. <https://doi.org/10.1038/nbt.2203>
- [85] Dewhurst, S. M., McGranahan, N., Burrell, R. A., Rowan, A. J., Grönroos, E., Endesfelder, D., Joshi, T., Mouradov, D., Gibbs, P., Ward, R. L., Hawkins, N. J., Szallasi, Z., Sieber, O. M., & Swanton, C. (2014). Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discovery*, 4(2), 175–185. <https://doi.org/10.1158/2159-8290.cd-13-0285>
- [86] Gillis, S., & Roth, A. (2020). PyClone-VI: Scalable Inference of clonal population structures using whole genome data. *BMC Bioinformatics*, 21(1). <https://doi.org/10.1186/s12859-020-03919-2>
- [89] Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., & Shah, S. P. (2014). PyClone: Statistical inference of clonal population structure in cancer. *Nature Methods*, 11(4), 396–398. <https://doi.org/10.1038/nmeth.2883>
- [90] Leshchiner, I., Livitz, D., Gainor, J. F., Rosebrock, D., Spiro, O., Martinez, A., Mroz, E., Lin, J. J., Stewart, C., Kim, J., Elagina, L., Bozic, I., Mino-Kenudson, M., Rooney, M., Ou, S.-H. I., Wu, C. J., Rocco, J. W., Engelman, J. A., Shaw, A. T., & Getz, G. (2018). *Comprehensive Analysis of Tumour Initiation, Spatial and Temporal Progression under Multiple Lines of Treatment*. <https://doi.org/10.1101/508127>
- [91] Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S., & Swanton, C. (2016). DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biology*, 17(1). <https://doi.org/10.1186/s13059-016-0893-4>
- [92] Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., ... Forbes, S. A. (2018). Cosmic: The catalogue of somatic mutations in cancer. *Nucleic Acids Research*, 47(D1). <https://doi.org/10.1093/nar/gky1015>
- [93] Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- [94] Hu, M., & Chikina, M. (2023). *InstaPrism: An R Package for Fast Implementation of Bayesprism*. <https://doi.org/10.1101/2023.03.07.531579>

- [95] Chu, T., Wang, Z., Pe'er, D., & Danko, C. G. (2022). Cell type and gene expression deconvolution with BayesPrism enables bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nature Cancer*, 3(4), 505–517. <https://doi.org/10.1038/s43018-022-00356-3>
- [96] Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M., & Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*, 37(7), 773–782. <https://doi.org/10.1038/s41587-019-0114-2>
- [97] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- [98] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with deseq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- [99] Zhang, Y., & Wang, W. (2025). Advances in tumor subclone formation and mechanisms of growth and invasion. *Journal of Translational Medicine*, 23(1). <https://doi.org/10.1186/s12967-025-06486-3>
- [100] Ji, H., Hu, C., Yang, X., Liu, Y., Ji, G., Ge, S., Wang, X., & Wang, M. (2023). Lymph node metastasis in cancer progression: Molecular mechanisms, clinical significance and therapeutic interventions. *Signal Transduction and Targeted Therapy*, 8(1). <https://doi.org/10.1038/s41392-023-01576-4>
- [101] Park, S., & Yun, S.-J. (2022). Acral melanocytic neoplasms: A comprehensive review of Acral Nevus and acral melanoma in Asian perspective. *Dermatopathology*, 9(3), 292–303. <https://doi.org/10.3390/dermatopathology9030035>
- [102] Spain, L., Coulton, A., Lobon, I., Rowan, A., Schnidrig, D., Shepherd, S. T. C., Shum, B., Byrne, F., Goicoechea, M., Piperni, E., Au, L., Edmonds, K., Carlyle, E., Hunter, N., Renn, A., Messiou, C., Hughes, P., Nobbs, J., Foijer, F., ... Turajlic, S. (2023). Late-stage metastatic melanoma emerges through a diversity of evolutionary pathways. *Cancer Discovery*, 13(6), 1364–1385. <https://doi.org/10.1158/2159-8290.cd-22-1427>
- [103] Bander, E. D., Yuan, M., Carnevale, J. A., Reiner, A. S., Panageas, K. S., Postow, M. A., Tabar, V., & Moss, N. S. (2021). Melanoma brain metastasis presentation, treatment, and outcomes in the age of targeted and immunotherapies. *Cancer*, 127(12), 2062–2073. <https://doi.org/10.1002/cncr.33459>
- [104] Siu, T. L., & Huang, S. (2010). Cerebral metastases from malignant melanoma: Current treatment strategies, advances in Novel therapeutics and future directions. *Cancers*, 2(2), 364–375. <https://doi.org/10.3390/cancers2020364>
- [105] Hannan, E. J., O'Leary, D. P., MacNally, S. P., Kay, E. W., Farrell, M. A., Morris, P. G., Power, C. P., & Hill, A. D. K. (2017). The significance of BRAF V600E mutation status discordance between primary cutaneous melanoma and brain metastases. *Medicine*, 96(48). <https://doi.org/10.1097/md.00000000000008404>
- [106] Bander, E. D., Yuan, M., Carnevale, J. A., Reiner, A. S., Panageas, K. S., Postow, M. A., Tabar, V., & Moss, N. S. (2021a). Melanoma brain metastasis presentation, treatment, and outcomes in the age of targeted and immunotherapies. *Cancer*, 127(12), 2062–2073. <https://doi.org/10.1002/cncr.33459>
- [107] Hotz, M. J., O'Halloran, E. A., Hill, M. V., Hayden, K., Zaladonis, A. G., Deng, M., Olszanski, A. J., Reddy, S. S., Wu, H., Luo, B., & Farma, J. M. (2022). Tumor mutational burden and somatic mutation status to predict disease recurrence in advanced melanoma. *Melanoma Research*, 32(2), 112–119. <https://doi.org/10.1097/cmr.0000000000000808>
- [108] Patel, R. R., Ramkissoon, S. H., Ross, J., & Weintraub, L. (2020). Tumor mutational burden and driver mutations: Characterizing the genomic landscape of pediatric brain tumors. *Pediatric Blood & Cancer*, 67(7). <https://doi.org/10.1002/pbc.28338>
- [109] Maul, R. W., & Gearhart, P. J. (2010). Aid and somatic hypermutation. *Advances in Immunology*, 159–191. [https://doi.org/10.1016/s0065-2776\(10\)05006-6](https://doi.org/10.1016/s0065-2776(10)05006-6)

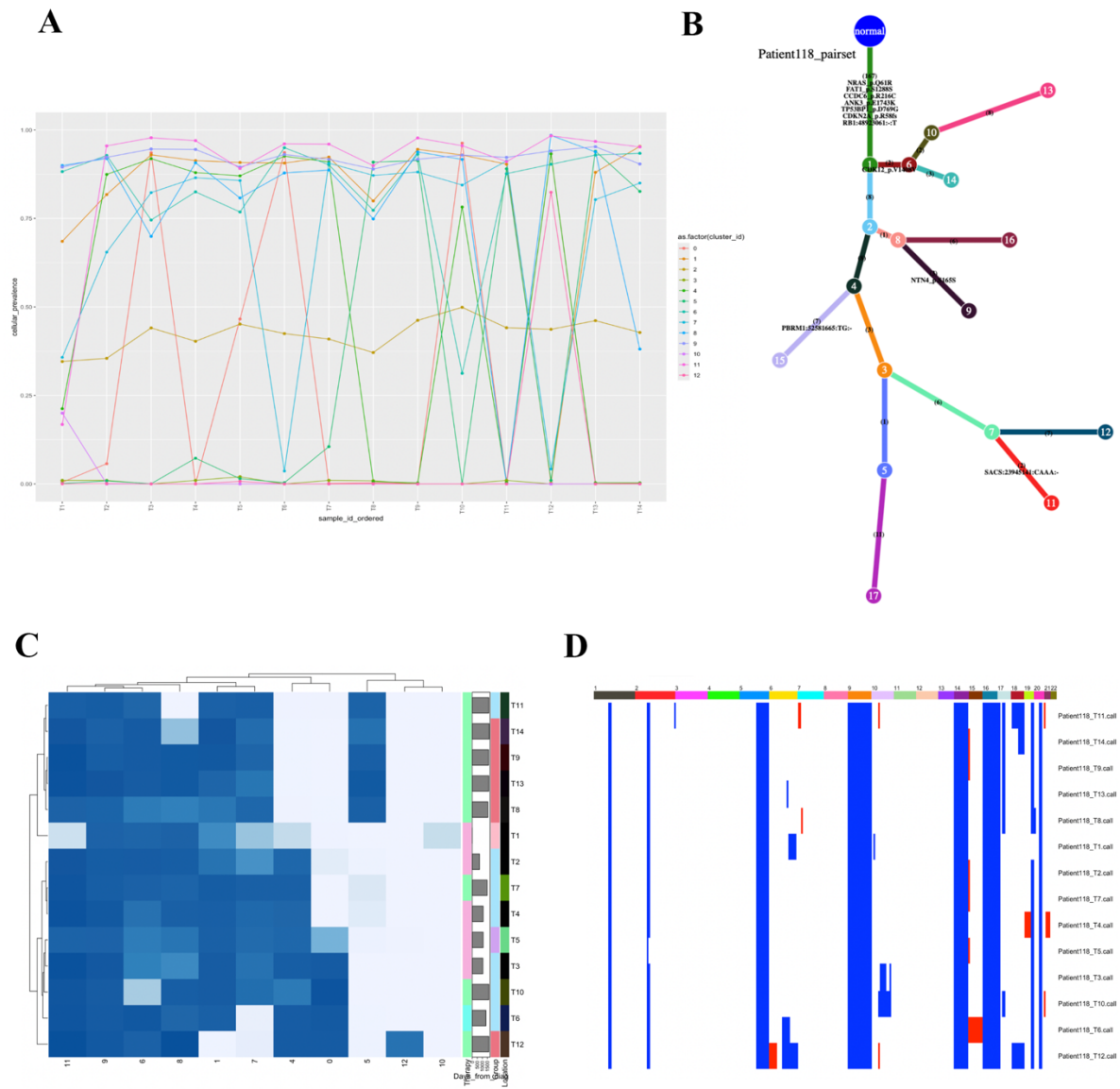
Supplementary Material

Supplementary Figures

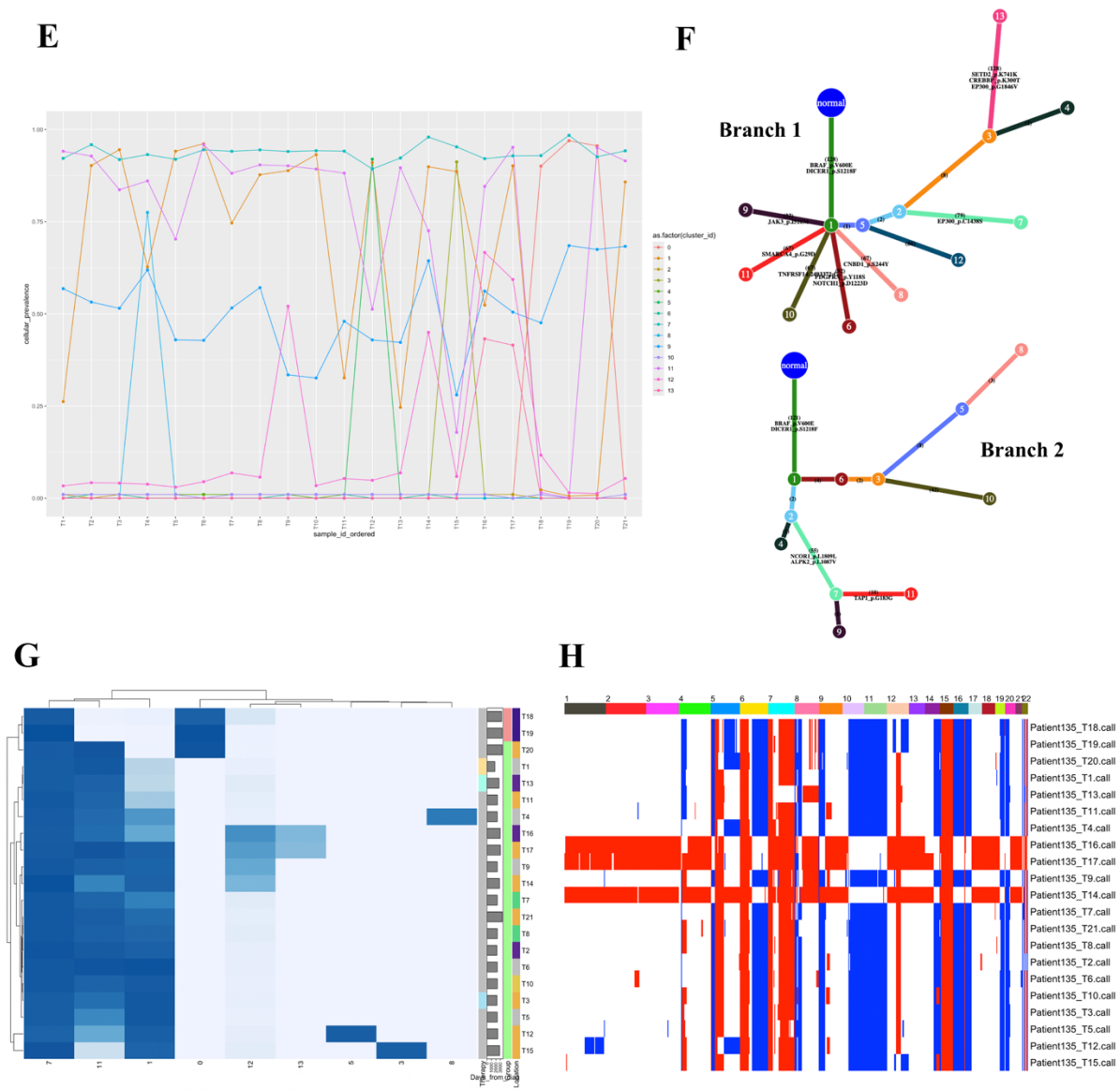
Supplementary Figure 1. Bulk RNA-seq MultiQC summary for patient 118.



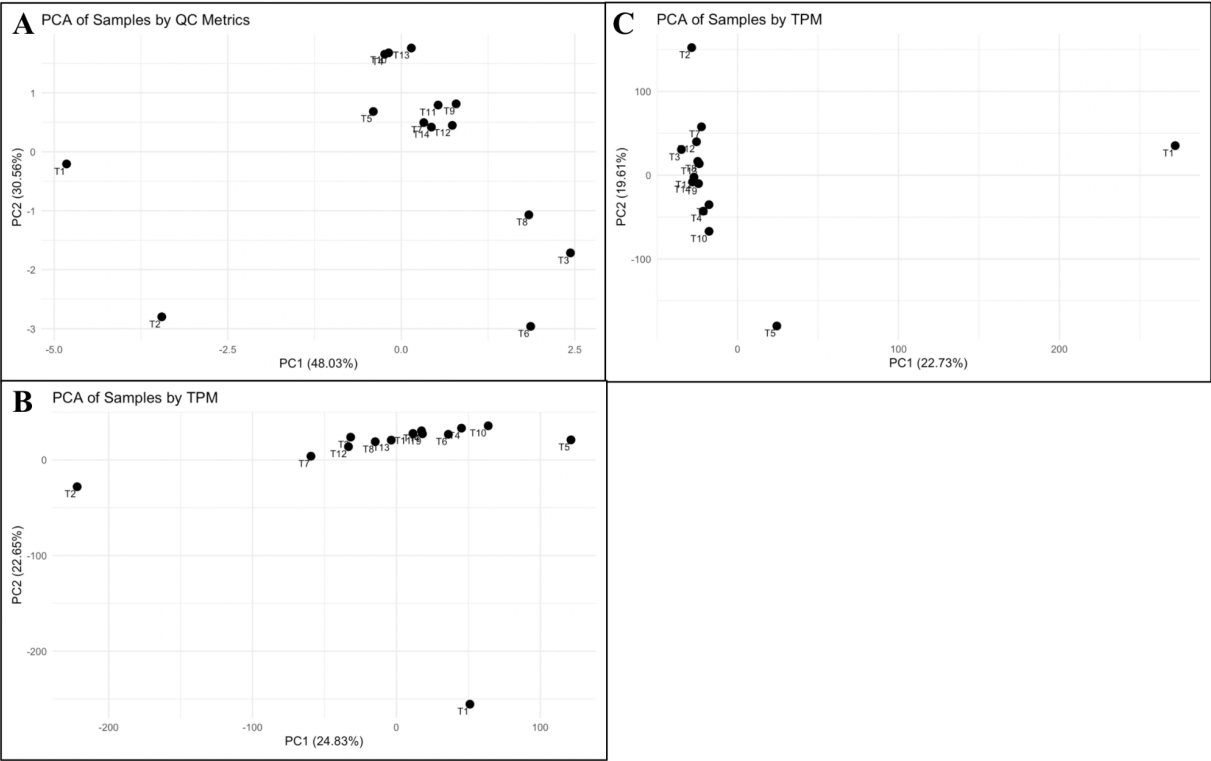
Supplementary Figure 2A-D. Evidence for phylogenetic tree construction for patients 118.



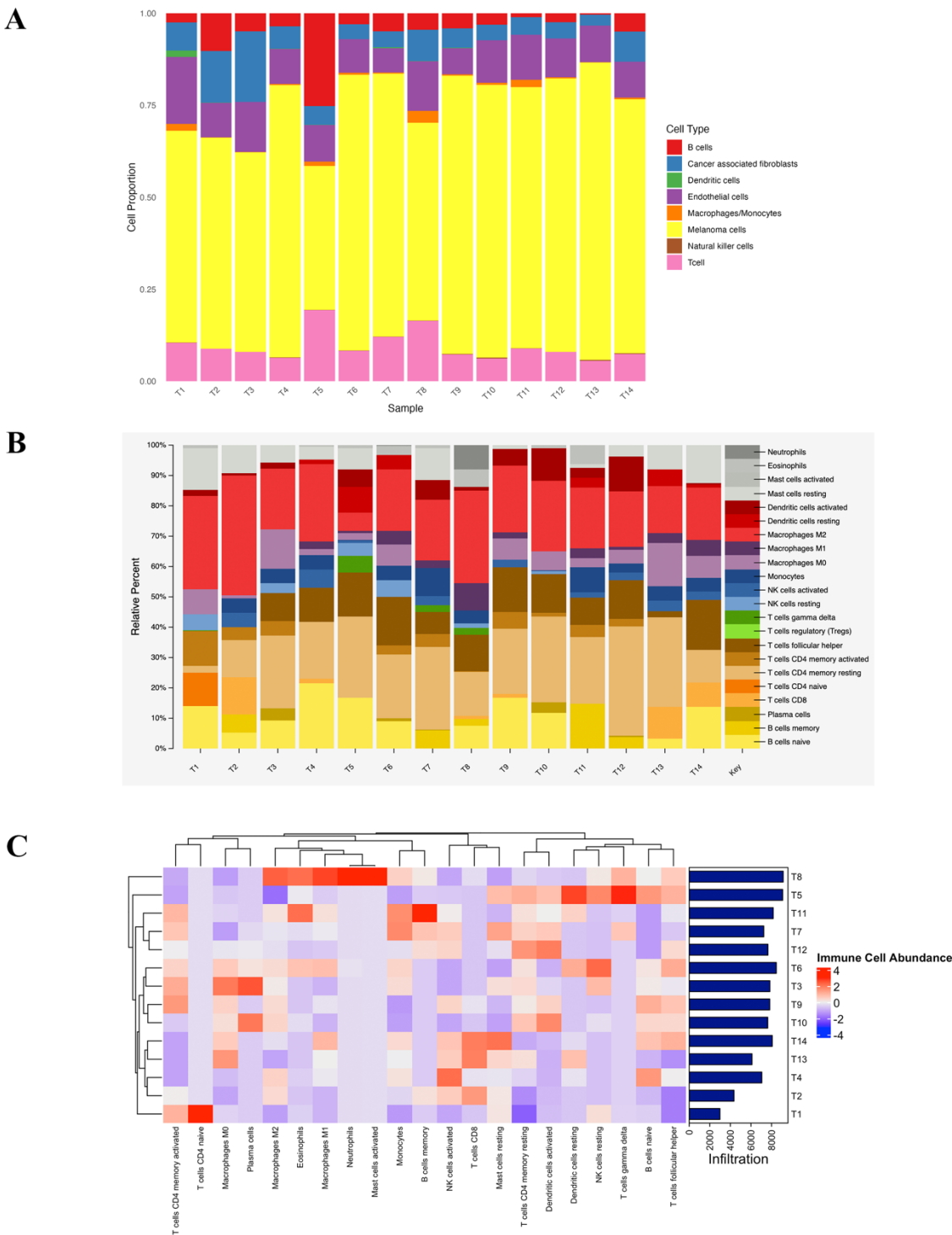
Supplementary Figure 2E-H. Evidence for phylogenetic tree construction for patients 135.



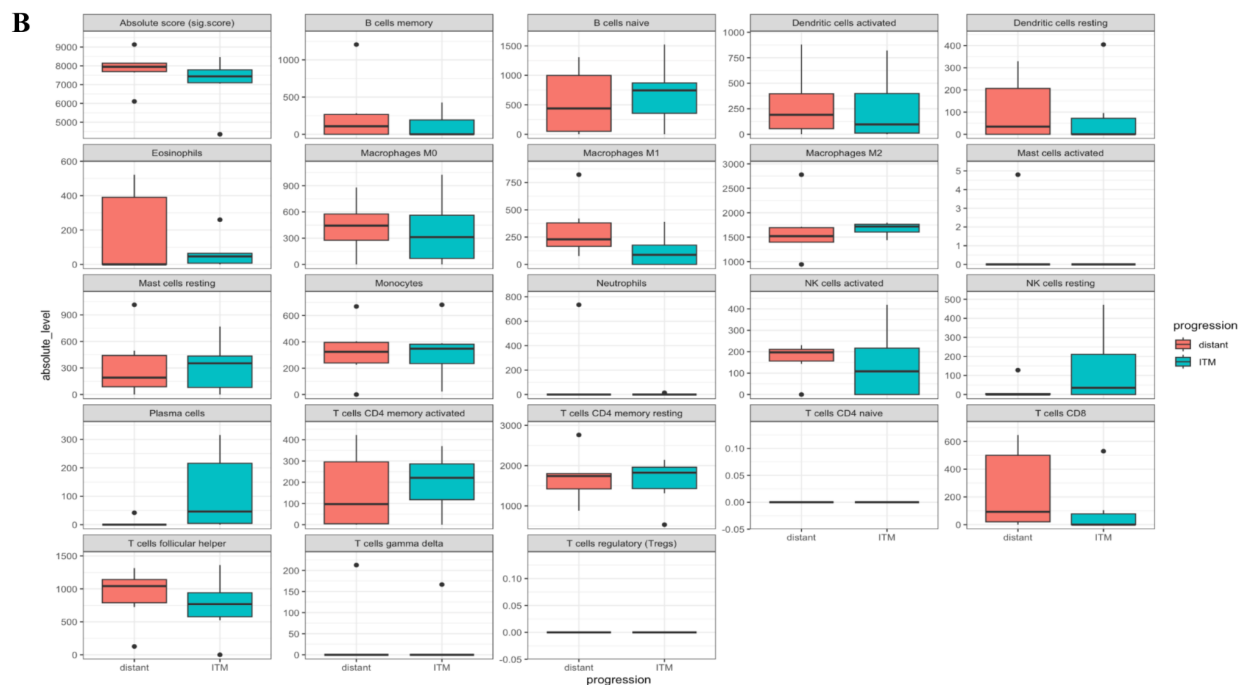
Supplementary Figure 3. Principal component analysis (PCA) of bulk RNA-seq data for patient 118.

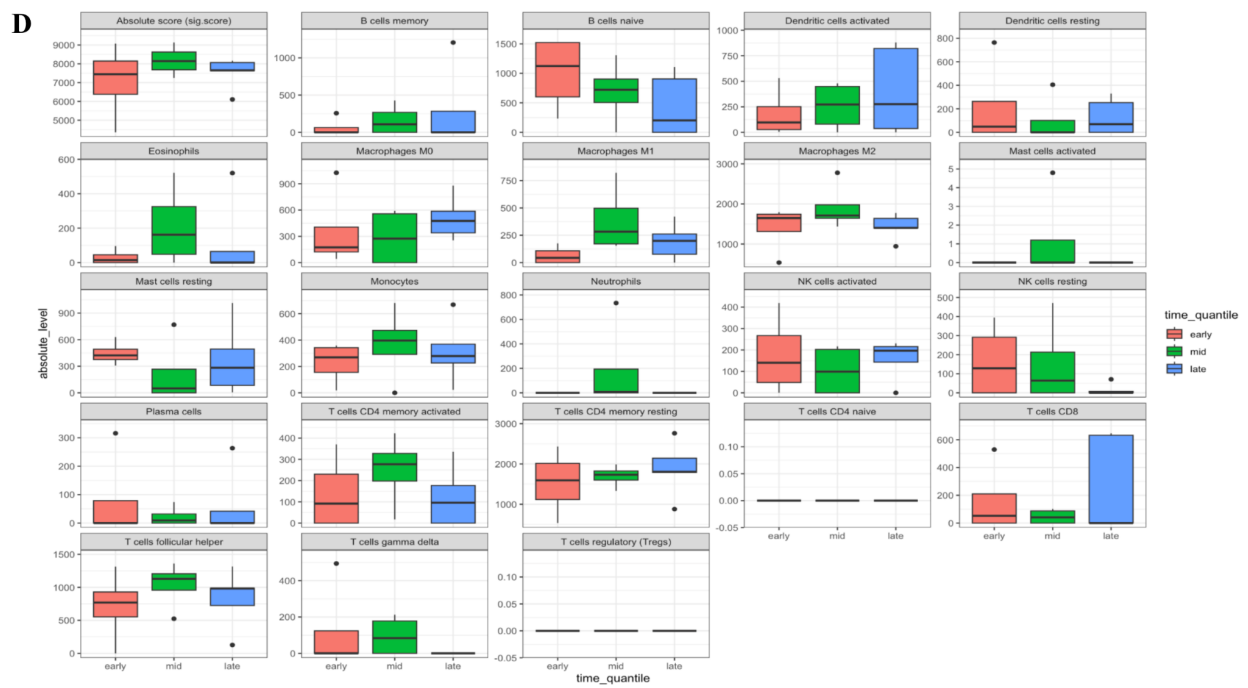
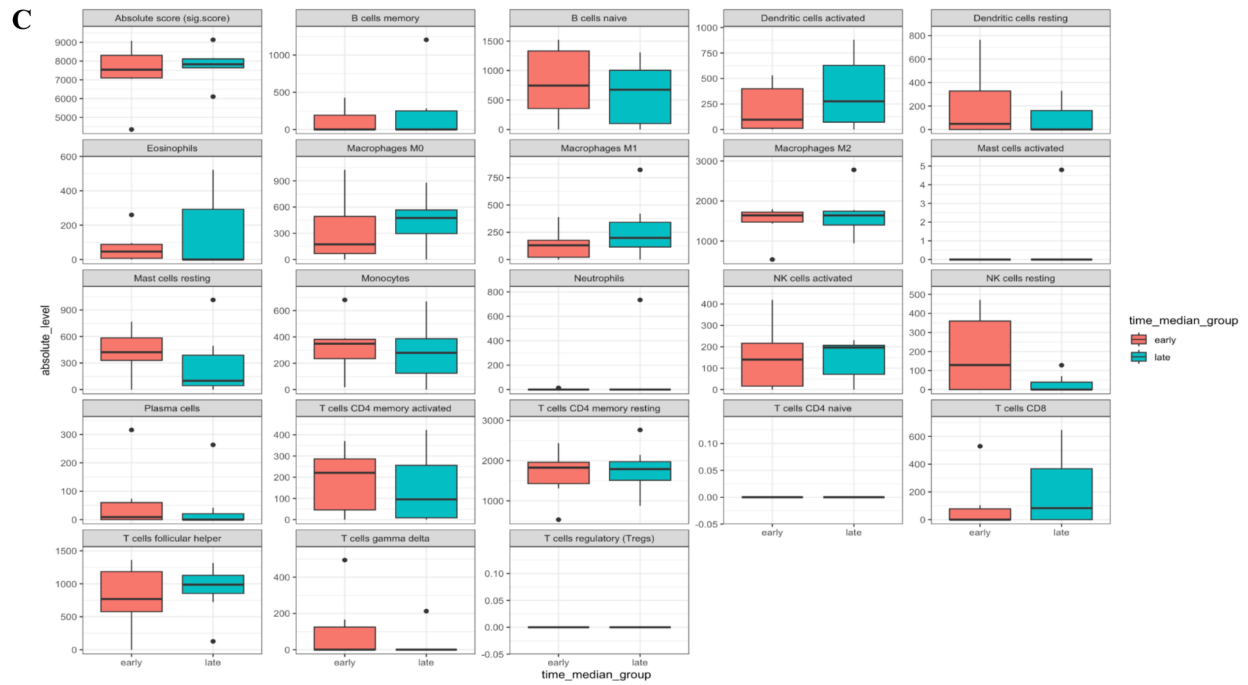


Supplementary Figure 4. Bulk RNA-seq data deconvolution using InstaPrism and CIBERSORT.

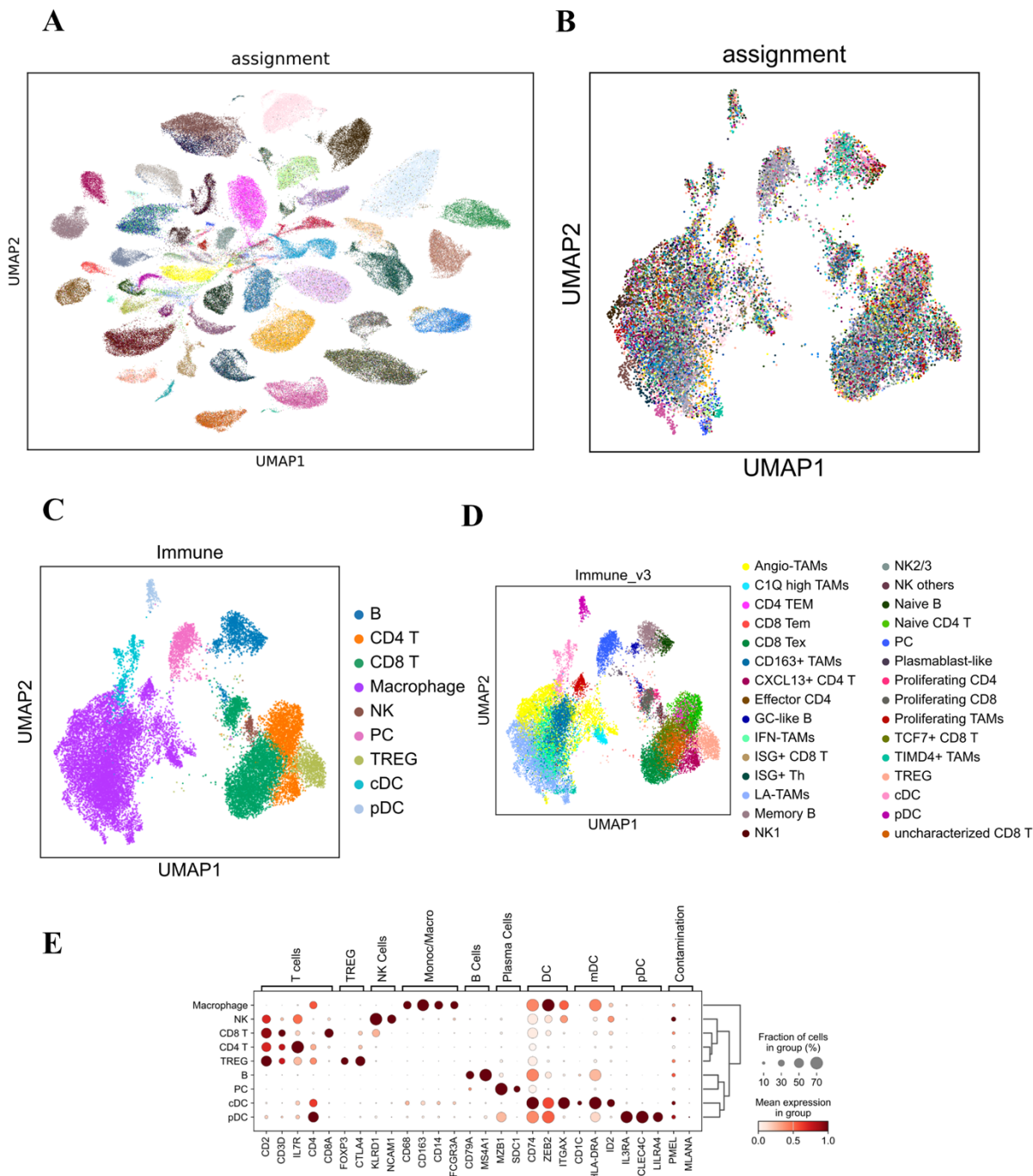


Supplementary Figure 5. Immune infiltration level comparison across clinical and biological sample groups.

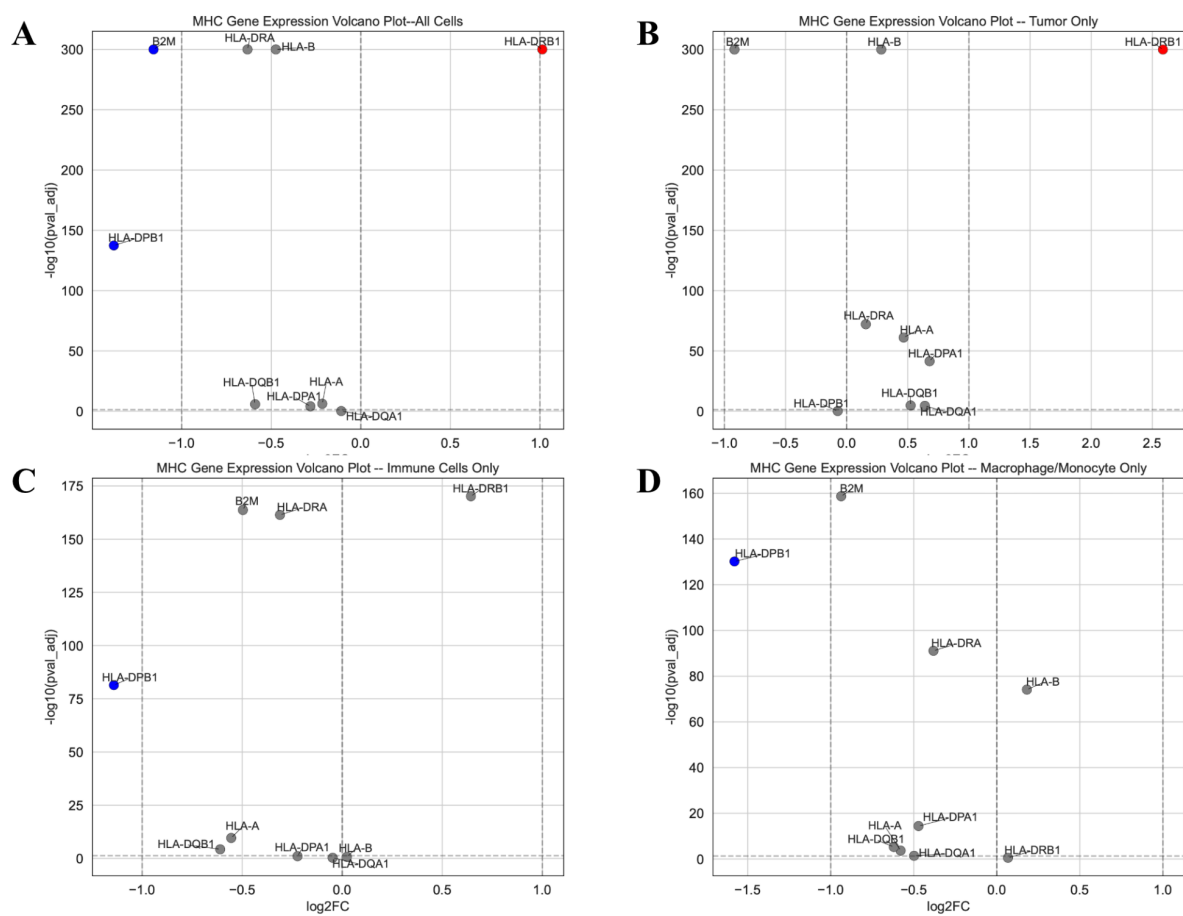




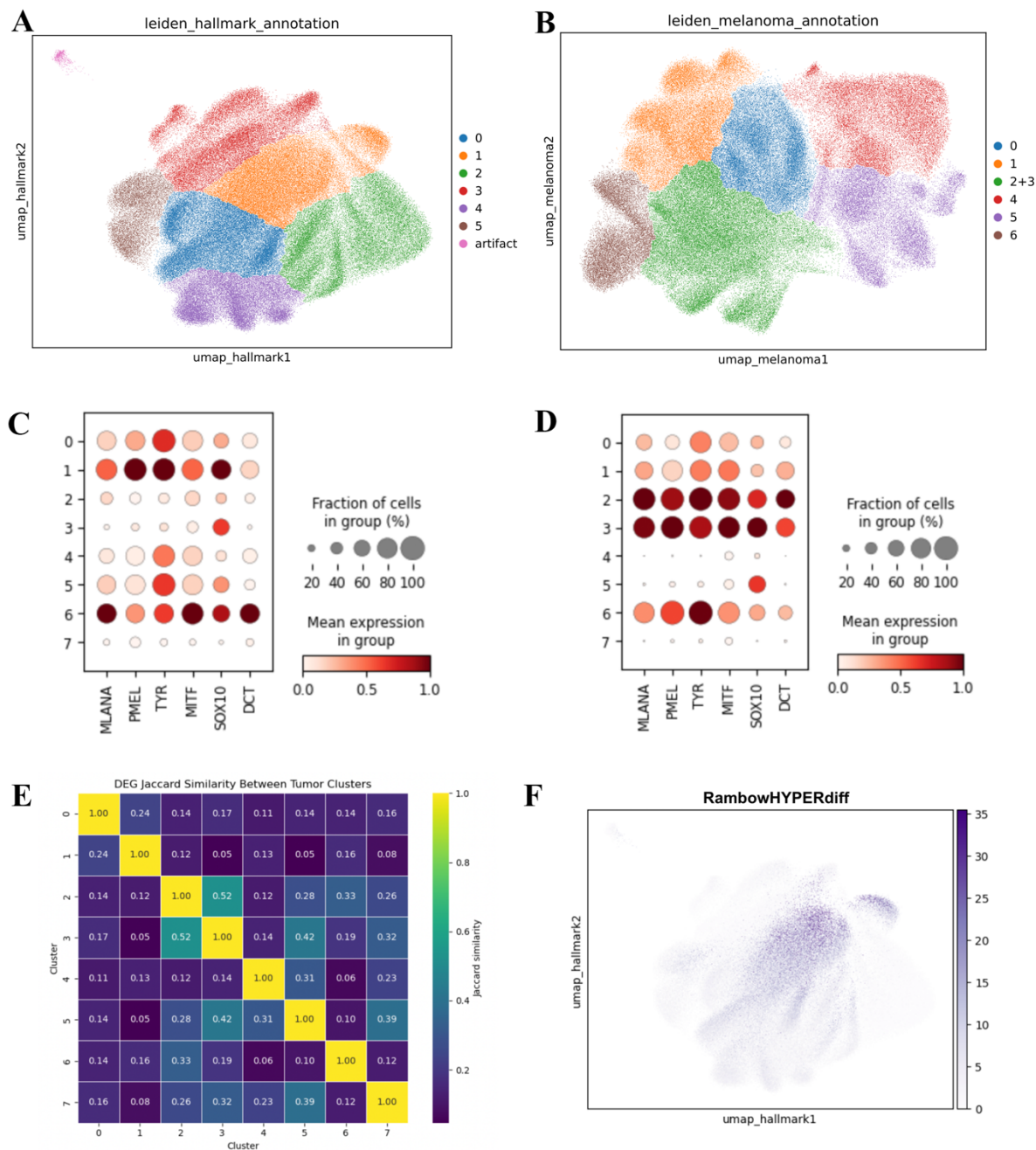
Supplementary Figure 6. HTAN tumor cells characterization and immune cell subtype annotation.



Supplementary Figure 7. Differentially expressed Major Histocompatibility Complex (MHC) genes in MERFISH data.

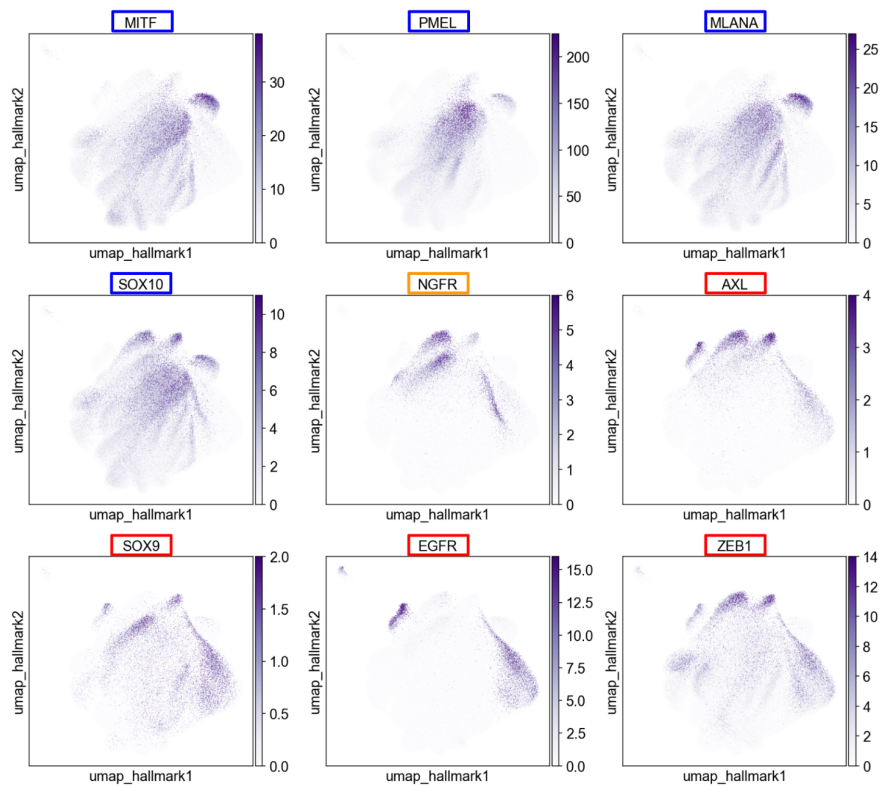


Supplementary Figure 8. Biologically relevant tumor program identification.

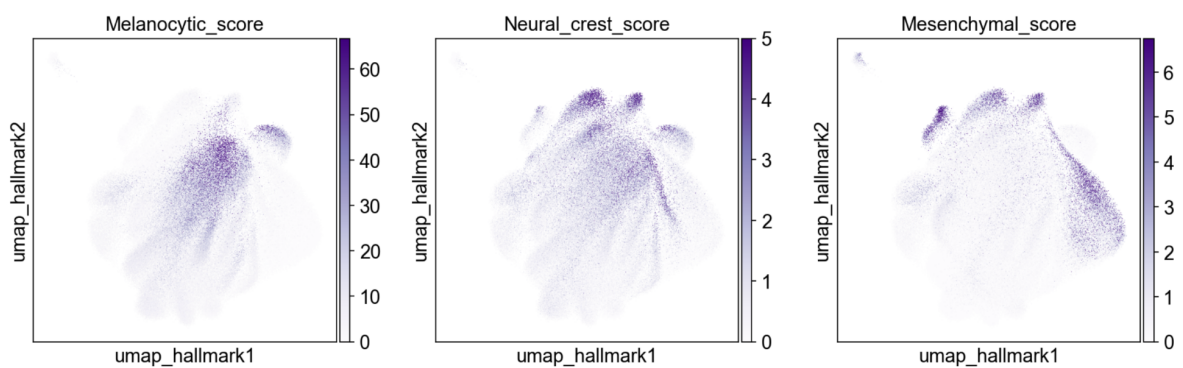


G

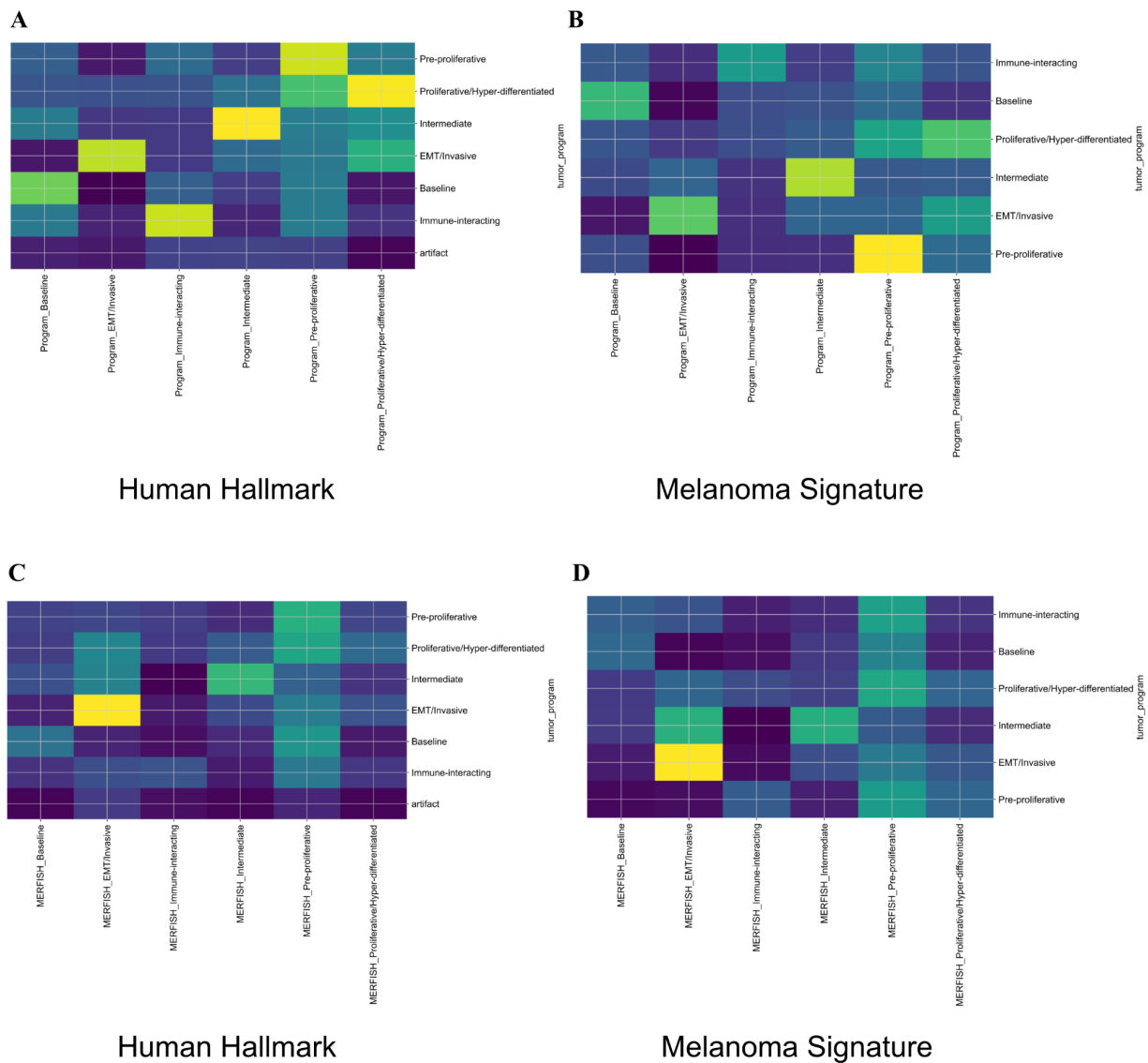
■ Melanocytic
■ Neural-crest-like
■ Dedifferentiated



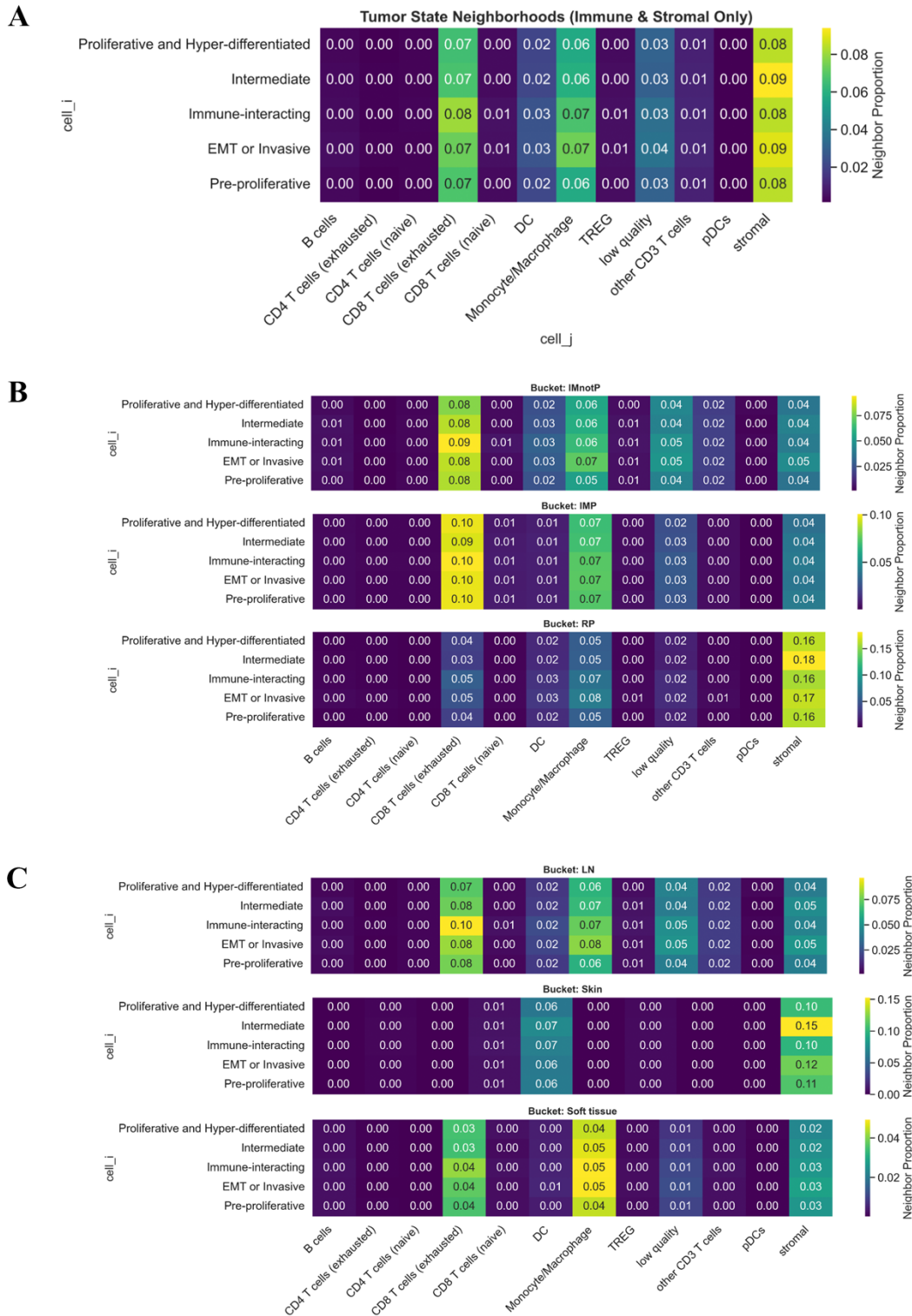
H



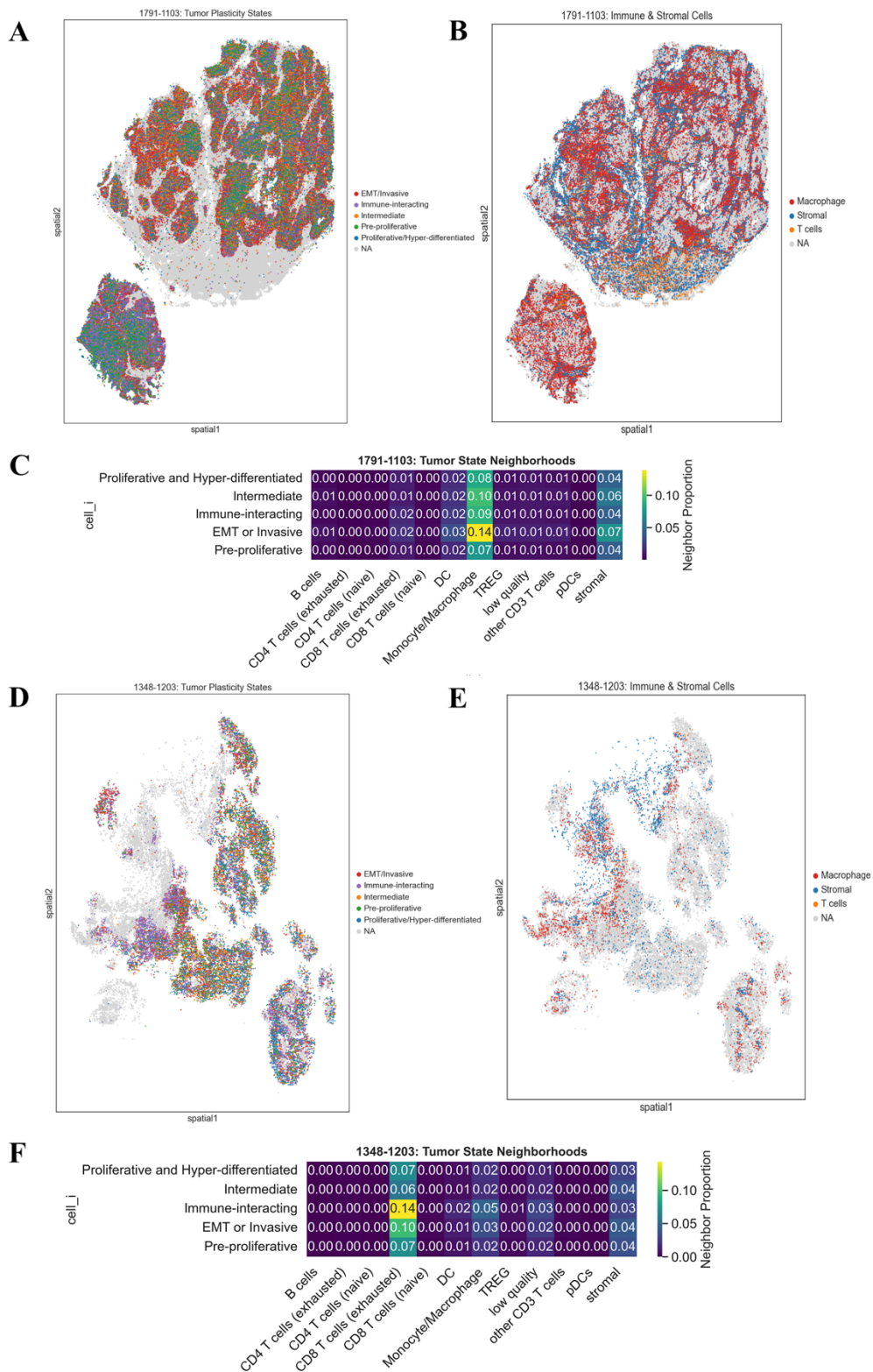
Supplementary Figure 9. Validation of novel tumor program gene markers.



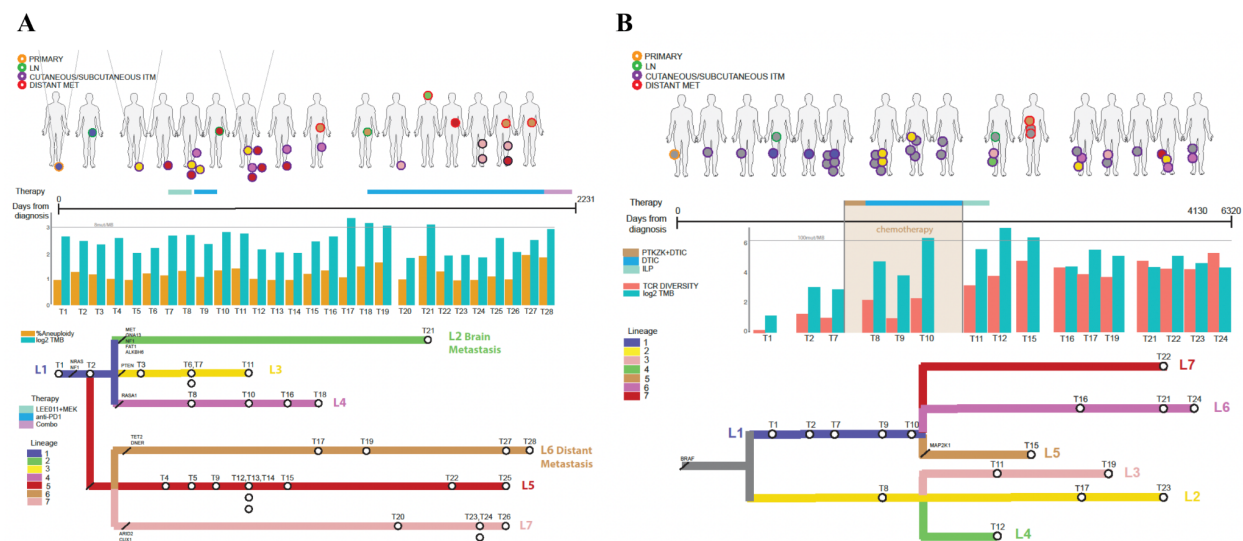
Supplementary Figure 10. Preliminary spatial exploration of novel tumor programs and their immune association in MERFISH data.



Supplementary Figure 11. Examples of tumor programs associating with specific immune cell types in selected MERFISH slides.



Supplementary Figure 12. Phylogenetic analysis of additional melanoma patients from the ITM cohort (patients 136 and 222). ** Directly adapted from Giuseppe et al.⁶⁵.



Supplementary Tables

Supplementary Table 1: ITM patient cohort summary. ** Adapted from Giuseppe et al.⁶⁵.

Characteristic	N	Overall, N = 41 [†]	Distant Progressors, N = 22 [†]	Distant non-progressors, N = 19 [†]
Sex	41			
female		17 (41.5%)	10 (45.5%)	7 (36.8%)
male		24 (58.5%)	12 (54.5%)	12 (63.2%)
Age at diagnosis	41	67 (12)	63 (13)	72 (9)
Localisation primary melanoma	41			
acral		6 (14.6%)	4 (18.2%)	2 (10.5%)
head_neck		2 (4.9%)	0 (0.0%)	2 (10.5%)
lower_extremity		22 (53.7%)	13 (59.1%)	9 (47.4%)
other		1 (2.4%)	1 (4.5%)	0 (0.0%)
trunk		6 (14.6%)	2 (9.1%)	4 (21.1%)
upper_extremity		4 (9.8%)	2 (9.1%)	2 (10.5%)
Histological subtype	41			
ALM		8 (19.5%)	5 (22.7%)	3 (15.8%)
amelanotic		2 (4.9%)	2 (9.1%)	0 (0.0%)
LMM		1 (2.4%)	0 (0.0%)	1 (5.3%)
NMM		18 (43.9%)	8 (36.4%)	10 (52.6%)
other		1 (2.4%)	1 (4.5%)	0 (0.0%)
SSM		4 (9.8%)	2 (9.1%)	2 (10.5%)
unknown		7 (17.1%)	4 (18.2%)	3 (15.8%)
Breslow thickness	41			
T1		2 (4.9%)	0 (0.0%)	2 (10.5%)
T2		4 (9.8%)	3 (13.6%)	1 (5.3%)
T3		17 (41.5%)	9 (40.9%)	8 (42.1%)
T4		17 (41.5%)	9 (40.9%)	8 (42.1%)
unknown		1 (2.4%)	1 (4.5%)	0 (0.0%)
Ulceration	41			
no		13 (31.7%)	9 (40.9%)	4 (21.1%)
unknown		7 (17.1%)	1 (4.5%)	6 (31.6%)
yes		21 (51.2%)	12 (54.5%)	9 (47.4%)
Systemic_therapy	41	30 (73.2%)	14 (63.6%)	16 (84.2%)
Firstline PD-1 yes/no	41	16 (39.0%)	5 (22.7%)	11 (57.9%)
BOR to antiPD-1	16			
PD		6 (37.5%)	2 (40.0%)	4 (36.4%)
PR		4 (25.0%)	1 (20.0%)	3 (27.3%)
SD		5 (31.3%)	2 (40.0%)	3 (27.3%)
UNK		1 (6.3%)	0 (0.0%)	1 (9.1%)
WES	41			
N		4 (9.8%)	2 (9.1%)	2 (10.5%)
Y		37 (90.2%)	20 (90.9%)	17 (89.5%)
RNAseq	41			
N		5 (12.2%)	0 (0.0%)	5 (26.3%)
Y		36 (87.8%)	22 (100.0%)	14 (73.7%)

[†] n (%); Mean (SD)

Supplementary Table 2: HTAN patient cohort summary. ** Table adapted from Giuseppe Tarantino's work.

Characteristic	N	Overall N = 67 [†]	DCB N = 10 [†]	IMnotP N = 12 [†]	IMP N = 23 [†]	RP N = 22 [†]
tx_group	67					
combo_ICB		18 (27%)	5 (50%)	0 (0%)	5 (22%)	8 (36%)
Neovax+combo_ICB		1 (1.5%)	0 (0%)	0 (0%)	0 (0%)	1 (4.5%)
single_aCTLA4		2 (3.0%)	0 (0%)	0 (0%)	1 (4.3%)	1 (4.5%)
single_aPD1		43 (64%)	5 (50%)	12 (100%)	16 (70%)	10 (45%)
single_aPD1_sequential		1 (1.5%)	0 (0%)	0 (0%)	1 (4.3%)	0 (0%)
single_aPD1+urelumab		2 (3.0%)	0 (0%)	0 (0%)	0 (0%)	2 (9.1%)
Tx_setting	67					
Adjuvant		16 (24%)	2 (20%)	7 (58%)	3 (13%)	4 (18%)
metastatic		49 (73%)	8 (80%)	3 (25%)	20 (87%)	18 (82%)
Neoadjuvant		2 (3.0%)	0 (0%)	2 (17%)	0 (0%)	0 (0%)
Subtype	66					
Acral		4 (6.1%)	0 (0%)	2 (17%)	0 (0%)	2 (9.1%)
Cutaneous		46 (70%)	9 (90%)	9 (75%)	16 (73%)	12 (55%)
Mucosal		3 (4.5%)	0 (0%)	0 (0%)	2 (9.1%)	1 (4.5%)
Unknown		13 (20%)	1 (10%)	1 (8.3%)	4 (18%)	7 (32%)
Unknown		1	0	0	1	0
LN_biopsy	67	25 (37%)	3 (30%)	8 (67%)	6 (26%)	8 (36%)
Time_point	67					
Post-tx		29 (43%)	2 (20%)	2 (17%)	14 (61%)	11 (50%)
Pre-tx		38 (57%)	8 (80%)	10 (83%)	9 (39%)	11 (50%)
[†] n (%)						

Characteristic	N	Overall N = 67 ¹	DCB N = 10 ¹	IMnotP N = 12 ¹	IMP N = 23 ¹	RP N = 22 ¹
WES	67					
N		7 (10%)	0 (0%)	1 (8.3%)	5 (22%)	1 (4.5%)
Y		60 (90%)	10 (100%)	11 (92%)	18 (78%)	21 (95%)
snRNAseq	67					
N		1 (1.5%)	0 (0%)	0 (0%)	1 (4.3%)	0 (0%)
Y		66 (99%)	10 (100%)	12 (100%)	22 (96%)	22 (100%)
CODEX	67					
N		45 (67%)	8 (80%)	2 (17%)	19 (83%)	16 (73%)
Y		22 (33%)	2 (20%)	10 (83%)	4 (17%)	6 (27%)
MERFISH	67					
N		49 (73%)	9 (90%)	4 (33%)	20 (87%)	16 (73%)
Y		18 (27%)	1 (10%)	8 (67%)	3 (13%)	6 (27%)
mIF	67					
N		44 (66%)	8 (80%)	2 (17%)	18 (78%)	16 (73%)
Y		23 (34%)	2 (20%)	10 (83%)	5 (22%)	6 (27%)
¹ n (%)						

Supplementary Table 3: Melanoma Signature Gene Set List from Previous Studies

Melanoma Cell State	Melanoma Signature Gene Set Name
Melanocytic/ Proliferative/ Differentiated <i>MITF</i> -high	Rambow_2018_Mitosis; Rambow_2018_MITF_Targets; Kauffmann_2008_DNA_Replication; Tirosh_2016_MITF; Widmer_2012_Proliferative; Jeffs_2009_Proliferation; Verfaillie_2015_Proliferative; Hoek_2006_Proliferative; Tsoi_2018_Melanocytic; Wouters_2020_Melanocytic; Andrews_2022_MEL; Belote_2021_Cmel; Belote_2021_Vmel
Mesenchymal-like/ Invasive/	Rambow_2018_Invasion; Tirosh_2016_AXL; Widmer_2012_Invasive; Jeffs_2009_Invasive; Verfaillie_2015_Invasive; Hoek_2006_Invasive;

De-differentiated/ <i>AXL</i> -high	Tsoi_2018_Undifferentiated; Wouters_2020_Mesenchymal; Andrews_2022_MES; Belote_2021_MSC; Pozniak_2024_Mesenchymal_like
Neural Crest-like/ Transitory/ Intermediate/ Mixed	Rambow_2018_Neuro; Tsoi_2018_Transitory; Tsoi_2018_Neural_Crest; Wouters_2020_Intermediate; Pozniak_2024_Neural_Crest_like
Pigmented/ Hyper-differentiated	Rambow_2018_Pigmentation
Stress response/ Metabolic	Rambow_2018_Hypometabolic; Riesenbergs_2015_TNF_Response; Andrews_2022_NPLAS_Up; Andrews_2022_NPLAS_Down; Belote_2021_ADT; Pozniak_2024_Stress (p53_response); Pozniak_2024_Stress (hypoxia_response); Pozniak_2024_Mitochondrial (low_quality)
Immune-interacting/ Antigen presentation/ Interferon	Pozniak_2024_Interferon_alpha_beta_response; Pozniak_2024_Antigen_presentation
Signaling pathway	Ryu_2011_BRAFV600E_Targets; Verfaillie_2015_TEAD_Targets; McNeal_2021_PMA; McNeal_2021_EDN1
Cell cycle	Kauffmann_2008_DNA_Repair; Pozniak_2024_Mitotic

Supplementary Table 4: Major Histocompatibility Complex (MHC) gene sets enrichment p-values in (A) tumor and (B) immune cells.

MHC Feature	RP_vs_DCB	RP_vs_IMP	RP_vs_IMnotP	DCB_vs_IMP	DCB_vs_IMnotP	IMP_vs_IMnotP	RP_vs_NonProgressor	IMP_vs_NonProgressor	Progressor_vs_NonProgressor
HLA-A	0.783976	0.876167	0.581101	0.997414	0.998983	0.859789	0.517638	0.854962	0.659395
HLA-B	0.68377	0.921844	0.394396	0.997414	0.998983	0.674709	0.316162	0.67347	0.375258
HLA-C	0.901701	0.895613	0.644573	0.997414	0.998983	0.708694	0.653356	0.758609	0.644684
HLA-E	0.939116	0.867203	0.298498	0.997414	0.996	0.32828	0.434943	0.516713	0.359657
HLA-F	0.520296	0.514755	0.226307	0.997414	0.998983	0.786477	0.148289	0.794594	0.360481
TAP1	0.801074	0.914311	0.159698	0.997414	0.998983	0.381872	0.214517	0.550338	0.250962
TAP2	0.792758	0.928767	0.42091	0.997414	0.998983	0.687845	0.415045	0.758609	0.500731
B2M	0.905473	0.447989	0.928429	0.997414	0.998983	0.681309	0.99254	0.778154	0.884828
HLA-DMA	0.928249	0.564479	0.545185	0.997414	0.960628	0.346464	0.759549	0.576881	0.618789
HLA-DMB	0.964025	0.638602	0.479651	0.997414	0.965019	0.336226	0.672491	0.547339	0.546146
HLA-DOA	0.691817	0.44652	0.910544	0.997414	0.971634	0.664228	0.545577	0.282925	0.667473
HLA-DOB	0.81941	0.165889	0.705864	0.997414	0.99971	0.898616	0.614015	0.865546	0.957236
HLA-DPA1	0.921592	0.34113	0.707144	0.997414	0.998983	0.318822	0.855894	0.453697	0.659395
HLA-DPB1	0.961075	0.380224	0.562001	0.997414	0.998983	0.215053	0.734834	0.360062	0.510513
HLA-DQA1	0.863335	0.401738	0.655598	0.997414	0.939596	0.317921	0.278944	0.454904	0.244949
HLA-DQA2	0.631703	0.656893	0.196123	0.715162	0.998983	0.104466	0.167349	0.074018	0.06987
HLA-DQB1	0.793441	0.130934	0.823504	0.997414	0.973369	0.21074	0.376158	0.278843	0.226083
HLA-DQB2	0.953688	0.185006	0.853846	0.997414	0.998983	0.821943	0.847518	0.716024	0.527409
HLA-DRA	0.918008	0.274874	0.519699	0.997414	0.931941	0.123388	0.721444	0.230104	0.446534
HLA-DRB1	0.955443	0.251859	0.877416	0.997414	0.998983	0.43237	0.948715	0.469965	0.750516
HLA-DRB5	0.956365	0.472533	0.669614	0.997414	0.998983	0.395714	0.717317	0.427114	0.534699

MHC Feature	Macrophage_vs_NonProgressor	CD163TAM_Progressor_vs_NonProgressor	Immune_Progressor_vs_NonProgressor
HLA-A	0.999934	0.999916	1
HLA-B	0.999934	0.999916	1
HLA-C	0.867978	0.999916	1
HLA-E	0.999934	0.999916	1
HLA-F	0.999934	0.999916	1
TAP1	0.867978	0.999916	1
TAP2	0.999934	0.999916	1
B2M	0.999934	0.999916	1
HLA-DMA	0.999934	0.999916	1
HLA-DMB	0.999934	0.999916	1
HLA-DOA	0.999934	0.999916	1
HLA-DOB	0.999934	0.999916	1
HLA-DPA1	0.999934	0.999916	1
HLA-DPB1	0.999934	0.999916	1
HLA-DQA1	0.999934	0.999916	1
HLA-DQA2	0.999934	0.999916	1
HLA-DQB1	0.999934	0.999916	1
HLA-DQB2	0.999934	0.999916	1
HLA-DRA	0.999934	0.999916	1
HLA-DRB1	0.999934	0.999916	1
HLA-DRB5	0.999934	0.999916	1

Supplementary Table 5A: Tumor programs markers.

Tumor Program	Marker Genes
Baseline	<i>ABCB5, ABHD15-AS1, AC006059.1, AC008991.1, AC009562.1, AC011092.2, AC012409.2, AC015923.1, AC022146.2, AC098829.1, ADAMTSL1, AL031428.1, AL139383.1, AL139807.1, ANKUB1, AP001011.1, AP002336.2, AP003086.2, ATP10A, BCAN, CASC15, CCDC171, CCDC26, CCSER1, CHSY3, COL24A1, CYP7B1, DCT, DLGAP1, DMD, DOK6, ERVMER61-1, ESRRG, FAM178B, FAM78B, FBXL7, GLRX, GRIK2, GRK3, GULP1, HPN-AS1, HPSE2, IPO9-AS1, KCNIP4, KCNQ5, LINC00578, LINC00923, LINC01608, LINGO2, LSAMP, MEGF10, MIR99AHG, MSC-AS1, MTSSI, NCKAP5, NLGN1, NR2F2-AS1, NRG1, OBSCN, PALMD, PDE10A, PDE3A, PDZRN3, PHACTR1, PIR, PKNOX2, PLCG2, PLCL1, PPARGC1A, PPP2R2B, PRDM16, RARB, RBFOX1, RBMS3, RIMS2, RNF150, RNF219-AS1, ROBO2, RTTN, SDK1, SEMA6D, SHC4, SNTG1, SOX4, SPDYE3, TTC23L, XRR41</i>
Pre-proliferative	<i>AC008991.1, AC068599.1, AC092546.1, AC099520.1, AC116049.2, ACTG1, ADCY2, ADGRG1, AEBP1, AL137220.1, AP003049.2, APCDD1, APOD, APOE, ARHGAP24, BCHE, BSG, CA14, CADPS, CASP1, CAV1, CD81, CDH1, CTNND2, CTSA, CTSF, CTSH, CTSL, CYP2U1, DCLK1, DPP6, EEF1A1, ENOX1, FAM198B-AS1, FHOD3, FSTL5, FTH1, GABRG3, GALNTL6, GMPR, HHATL, HLA-E, HPGD, HSP90B1, IGFBP7, ITGB1, ITM2B, KCNN2, KCNQ5, LAPT4A, LGALS3BP, LINC00473, LINC00681, LINC00871, LINC01266, LURAP1L-AS1, MAP2, MCOLN3, MRGPRX3, MT-ATP6, MT-CO3, MT-CYB, MT-ND1, MT-ND3, MT-ND4, MT-ND4L, NR3C2, OCA2, PDIA3, PLCH1-AS1, PLPPR4, PRKG2, PRUNE2, PSAP, PURPL, QPCT, RGMB, RPL13, RPL3, RPL41, RPLP1, RPS3A, SEMA3C, SERPINE2, SERPINF1, SLC24A4, SLC24A5, SLC35B2, SLC45A2, SLC6A15, SOD3, SPARCL1, SPP1, ST8SIA6, SYTL2, TMEM59, TRIM48, USP53, VAV3, VEPH1, WDR63, XIST</i>
Proliferative/Hyper-differentiated	<i>ACP5, ALDOA, APOC1, ARL6IP5, ATP5F1E, ATP6AP1, ATP6AP2, ATP6V0C, BCAP31, BST2, CCND1, CFL1, CHCHD2, COX6A1, COX6C, COX7C, CSPG4, CSTB, CTSA, CTSK, CTSZ, CYB5R3, CYP51A1, DDIT4, DDOST, DHRS7, ECM1, EEF1B2, EIF1, EIF5A, EMP3, ENO1, ERP29, FAU, FOS, GAPDH, GNG5, GSTP1, H2AFJ, H2AFZ, HSPA8, HSPB1, ICAM1, IFITM3, KDELR1, KDELR2, KRTCAP2, LAPT4A, LDHA, LGALS3, LINC00473, LINC00504, MARCKSL1, MCAM, MIF, MXRA7, MYL6, NCL, NDUFA4, NPC2, OAZ1, OST4, PDIA4, PEBP1, PFN1, PGRMC1, PHB, PKM, PLAT, PLP1, POLR2L, PRDX4, RAN, RPL14, RPL15, RPL19, RPL23A, RPL24, RPL29, RPL35A, RPL36A, RPL38, RPL39, RPN1, RPS12, RPS13, RPS15, RPS15A, RPS21, RPS26, RPS27, RPS27A, RPS28, RPS3, RPS4X, S100A1, SCD, SEC11C, SLC25A5, SLC2A3, SPCS1, SRP14, SSR4, TAPI, TMBIM6, TMED2, TMED9, TOMM7, TP11, TSPO, TUBA1B, UBC, UQCRLB, VAT1</i>
Intermediate	<i>AC016687.3, AC092691.1, AC109466.1, ADGRB3, ADGRL3, AL033504.1, AL929601.1, ALDH1A2, BCAT1, C22orf34, C5orf17, CACNA1C, CADM1, CCSER1, CDH11, CNTN4, COL1A1, COL1A2, COL3A1, COL5A1, CPED1, DACH1, DIRC3, DPP10, EDN3, EGFR,</i>

	<i>EGR1, EMCN, EMILIN1, ERG, EYA1, FAM110B, FBN2, FGF13, FGFR1, FXP2, GLIS3, GPC5, GPC6, GRID2, GRM7, H3F3A, HDAC9, HSP90AB1, ID3, IGFBP2, ITGA8, KALRN, KCNT2, KIF26B, LAMB1, LINC01297, LINC01524, LINGO2, LRRC4C, MDK, MEG3, MFAP2, MMP16, NLGN4Y, NRXN1, NRXN3, PALLD, PBX1, PCDH15, PDE7B, PDZD2, PHLDB2, PID1, PLK2, PRKG1, PTN, PTPRD, PXDN, RNF150, ROR2, RPL11, RPL13, RPL14, RPL15, RPL18A, RPL19, RPL21, RPL28, RPL3, RPL30, RPL32, RPL34, RPL37, RPL4, RPL8, RPL9, RPS12, RPS14, RPS16, RPS18, RPS19, RPS2, RPS23, RPS27, RPS3A, RPS6, RPS8, RPS9, RUNX1T1, SAMD5, SEMA3A, SETBP1, SLIT2, SORCS2, SOX4, SULF1, TCF4, TOX, TSHZ2, TUBB, UNC5C, VCAN, WISP1</i>
EMT/Invasive	<i>AC074327.1, ADAMTS16, AHNAK, AIM2, ALDH1A3, ANKH, ANXA1, ARL4C, AXL, CD109, CD9, CDH11, CDH2, CLIC1, COL12A1, COL18A1, COL4A1, COL6A1, COL6A2, COL6A3, COL7A1, COL8A1, CRYBG1, CST3, DAG1, DCBLD2, DPYD, ECE1, EDIL3, ENO1, FABP5, FAM20C, FAP, FLRT3, FN1, FRMPD4, FXYD5, GPC1, HDAC9, HLA-DPA1, HLA-DRA, HLA-DRB1, HLA-DRB5, HSPA5, IGFBP5, IL1RAPL1, ITGA3, ITGA6, ITGB8, KCNMA1, KCNN4, LINC02257, LOXL3, LRP1B, MAP1B, MT2A, MYDGF, MYOF, NEFL, NGFR, NID1, NPTX2, NT5E, NTRK3, P4HA2, PAPP, PDIA4, PERP, PLAUR, PLEC, PLOD1, PLP2, PMEP1, PTMA, RAB27B, RCN1, RFX8, RNF128, RUNX2, S100A10, S100A11, S100A13, S100A16, S100A4, S100A6, SEMA3B, SERPINE1, SH3BGRL3, SLC16A3, SLC20A1, SLC9A7, SPARC, SPOCK1, SRPX, SRPX2, TAGLN2, TENM3, TGFB1, THBS2, TIMP1, TIMP3, TMEM158, TMEM178B, TMSB10, TNFRSF12A, TRAF1, VEGFA</i>
Immune-interacting	<i>A2M, ABCA8, ABCB4, AC007319.1, AC011294.1, AC016687.3, AC023824.3, AC079793.1, AC090502.1, AC090673.1, AC116049.2, ADAM12, ADCY1, AGMO, AKAP6, AL390957.1, ANKRD30B, AP000561.1, AP003400.1, ATP10B, ATP1B1, B2M, BAALC, BIRC3, C4orf19, CA12, CADPS2, CCDC144NL-AS1, CEACAM1, CFH, CHL1, CHRM3, COL22A1, CRTAC1, CSPG4, CTNNA3, CUBN, CXCL10, DOK5, ENOX1, EPSTI1, EYA1, FRMD5, GBP1, GBP4, GBP5, GNGT1, HHATL, HLA-B, HLA-C, HORMAD1, IDO1, IFI27, IFI44L, IFI6, IGFBP7, IRF1, KCNMB2-AS1, KIAA1549L, KIAA1755, L3MBTL4, LINC01876, LINC02456, LRAT, LRP2, LRRTM4, MAMDC2, MIR2052HG, MPPED2, MRGPRX3, MSR1, MT-ND2, MX1, NELL1, NOX4, PDE7B, PDK4, PII5, PLA1A, POSTN, PRELID2, PTPRZ1, PYHIN1, RNASE1, RUNX1, SCIN, SEMA5A, SERPINB9, SFRP1, SHC4, SLC5A4-AS1, SLC6A15, SLC9A3, SOD3, SORBS2, SPON2, SPP1, STXBP5-AS1, STXBP5L, TAPI, TDRD9, TMEM140, TMEM163, TRIM22, TTN, VEGFD, WARS, XAF1</i>

Supplementary Table 5B: Selected tumor programs contained in MERFISH gene.

Tumor Program	Marker Genes
Baseline	<i>PHACTR1</i>
Pre-proliferative	<i>AEBP1, CTSL, IGFBP7, ITM2B, SERPINE2, SOD3, SPARCL1</i>
Proliferative/Hyper-differentiated	<i>APOC1, CTSK, EIF1, LGALS3, RPS15A, RPS4X, S100A1, TUBA1B</i>
Intermediate	<i>CDH11, COL1A1, COL1A2, COL3A1, RPL11, TCF4, TOX, TUBB, VCAN</i>
EMT/Invasive	<i>ALDH1A3, ANXA1, CDH11, COL6A1, COL6A2, COL6A3, FN1, HLA-DPA1, HLA-DRA, HLA-DRB1, IGFBP5, MT2A, NGFR, PTMA, RUNX2, S100A11, S100A13, S100A4, SERPINE1, SPARC, TIMP1, TIMP3, VEGFA</i>
Immune-interacting	<i>A2M, B2M, BIRC3, CEACAM1, HLA-B, IFI27, IGFBP7, SOD3, SPON2</i>