Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uncovering the landscape of functional transition and plasticity in the tumor microenvironment

Riddhiman Dhar

riddhiman.dhar@iitkgp.ac.in

IIT Kharagpur https://orcid.org/0000-0003-4642-0492

Subhasis Datta

IIT Kharagpur

Rakesh Mandal

IIT Kharagpur

Sohom Basak

IIT Kharagpur

Article

Keywords: Functional programs, Tumor microenvironment (TME), Functional transitions, Transitory myeloid states, Functional plasticity, Cancer progression

Posted Date: September 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10983198/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Title: Uncovering the landscape of functional transition and plasticity in the tumor microenvironment

Authors: Subhasis Datta[1], Rakesh Mandal[1], Sohom Basak[2], Riddhiman Dhar[1*]

Affiliation: [1]Department of Bioscience and Biotechnology, IIT Kharagpur, India

[2]School of Medical Science and Technology (SMST), IIT Kharagpur, India

*Corresponding author, **email:** riddhiman.dhar@iitkgp.ac.in

Abstract

The tumor microenvironment (TME) harbours multiple cell-types whose functional contributions collectively determine the overall tumor phenotype as well as intra-tumor heterogeneity. These cell types undergo functional and phenotypic transitions that have important implications for cancer progression and therapy response. Intra-tumor functional plasticity has been mapped primarily within individual cell types and these studies, across different cancers, have identified common as well as cancer-specific programs. However, our understanding of less frequent transitory cell states within the TME and their role in cancer progression remains limited. In addition, the landscape of functional plasticity between different cell types within a tumor remains largely unexplored. In the present work, by analysing tumor functional states in lung adenocarcinoma and colorectal cancer, we uncover the occurrence of transitory functional states in myeloid cell populations. These transitory myeloid states occurred preferentially in tumor samples compared to adjacent normal samples and were associated with signatures of immune suppression. We also observed widespread functional plasticity within the TME as different cell types converged to similar functional states across different patients. Taken together, our work illustrates the landscape of functional transition and plasticity within the TME. The discovery of transitory cell states could have important implications for cancer progression and immunotherapy.

Keywords

Functional programs, Tumor microenvironment (TME), Functional transitions, Transitory myeloid states, Functional plasticity, Cancer progression

Running title

Functional transitions and plasticity with cancer progression

Introduction

The tumor microenvironment (TME) is a complex ecosystem where malignant cells coexist with non-malignant epithelial, stromal and tumor infiltrating immune cells (Anderson and Simon, 2020; de Visser and Joyce, 2023). Different cell types in the TME perform a diverse array of functions and their interactions with the malignant cells have important influence on tumor growth, microenvironmental regulation and immune evasion (Quail and Joyce, 2013; de Visser and Joyce, 2023). The cellular composition and spatial organization of the TME varies between patients of a cancer type and also evolves with cancer progression. Further, the cells in the TME show widespread functional plasticity in response to environmental signals (Yuan et al., 2019; Pérez-González et al., 2023). Collectively, functional diversity and plasticity in the TME have major implications for tumor growth, cancer progression and metastasis, and anti-cancer therapy response (Klemm and Joyce, 2015; Pérez-González et al., 2023).

Functional diversity has been shown to emerge even before cancer initiation (Bala et al., 2023), and to persist with cancer progression and metastasis (Marjanovic et al., 2020; Moorman et al., 2025; Buissant des Amorie et al., 2026). Several recent studies have focused on understanding such diversity of functional programs using bulk, single-cell and spatial transcriptomic datasets (Neftel et al., 2019; Hausser et al., 2019; Kinker et al., 2020; Barkley et al., 2022; Gavish et al., 2023; Han et al., 2024). Among these, single-cell studies have elucidated occurrence of a diverse set of functional programs in malignant as well as non-malignant cells (Neftel et al., 2019; Barkley et al., 2022; Gavish et al., 2023). Some of these programs were found across different cancer types, whereas some were cancer-specific (Barkley et al., 2022; Gavish et al., 2023). Further, some functional programs in malignant cells cooccurred with specific functional programs in other cell types in the TME, suggesting close functional interaction between these cell types (Barkley

et al., 2022; Gavish et al., 2023). The functional states and their interactions were also associated with patient survival and therapy response (Kinker et al., 2020; Gavish et al., 2023; Sun et al., 2025).

Functional diversity described in these studies mostly reflect stable cell states within individual cell types and are often represented by a substantial number of cells within the TME. Beyond these, there are likely to be less frequent transitory functional states that could also have important implications for cancer progression and therapy response (Emert et al., 2021; Schaff et al., 2026). In addition, functional diversity and plasticity can also arise through differentiation or dedifferentiation of cell types in presence of specific signals in the TME (Yuan et al., 2019, Sanseviero et al., 2026). However, our understanding of the less frequent or rare functional states remains limited. In addition, it is not clear how functional transitions and plasticity within and across cell types evolve with cancer progression.

In the present work, we characterized the intra-tumor diversity of functional programs in lung adenocarcinoma (LUAD) and colorectal cancer (CRC) patients. We found functional diversity within individual cell types and observed a change in utilization of functional programs with cancer progression. Further, we uncovered transitory functional states – specifically in myeloid cells - that occurred preferentially in tumor samples compared to adjacent normal samples in both LUAD and CRC. These transitory myeloid cells had lower expression of genes associated with antigen presentation, but lacked the expression signatures of typical myeloid- derived suppressor cells. Occurrence of transitory myeloid cell states in patient samples was associated with signatures of immune suppression. In addition, our analysis revealed widespread intra-sample and inter-sample functional plasticity where different cell types converged to similar functional states within and between samples. Taken together, our work uncovers the overall landscape of functional diversity, transition and plasticity in cancer samples. Further, the elucidation of transitory functional states and their association with immune microenvironment have important implications for cancer progression and therapy response.

Results

A Pareto front framework integrated with cell-state stability analysis reveals functional diversification within the tumor microenvironment

To investigate the occurrence of transitory functional cell states and functional plasticity in cancer, we analysed single-cell transcriptomic datasets from lung adenocarcinoma (LUAD) (Kim et al., 2020) and colorectal cancer (CRC) patients (Lee et al., 2020) of different stages along with adjacent normal samples (Table S1). Briefly, we analysed 26 patient samples in LUAD, comprising of 15 tumor samples (8 Stage I, one stage II, two stage III and four stage IV samples) and 11 adjacent normal samples (henceforth termed as normal samples). In CRC, we analysed 23 tumor samples (two stage I, four stage II, 15 stage III and two stage IV samples) and 9 normal samples. In LUAD samples, we considered epithelial, myeloid and T/NK cells for analysis whereas in CRC, we considered epithelial, myeloid, T and stromal cells for analysis, as there were enough number of cells (>100) across majority of the samples for drawing meaningful inferences.

For mapping functional diversity, transitions and plasticity in the tumor microenvironment (TME), we set up a framework consisting of a pareto front analysis coupled with density-based identification of cell-state stability (Fig. 1). In the first step, for each sample, we performed a pareto front analysis on standardized single-cell RNA sequencing data on all cells in the sample using the ParTI framework (Hart et al., 2015). This enabled us to identify the major functional archetypes (specialists and generalists) in the tumor cell population (Fig. 1A). Subsequently, we employed a distance-to-archetype distribution calculation to assign cells to specific archetypes (Fig. 1B). We analysed the gene expression patterns of the specialists and generalist archetypes to check consistent expression pattern within individual archetypes and distinct patterns among archetypes. In the next step, we applied a density-based cell state analysis using *Mellon* (Otto et al., 2024) that enabled identification of stable functional states (Fig. 1C). Majority (>90%) of the cells across samples belonging to specialist and generalist archetypes appeared in stable functional states (Fig. S1), further supporting the identity of the functional archetypes (Fig. 1D).

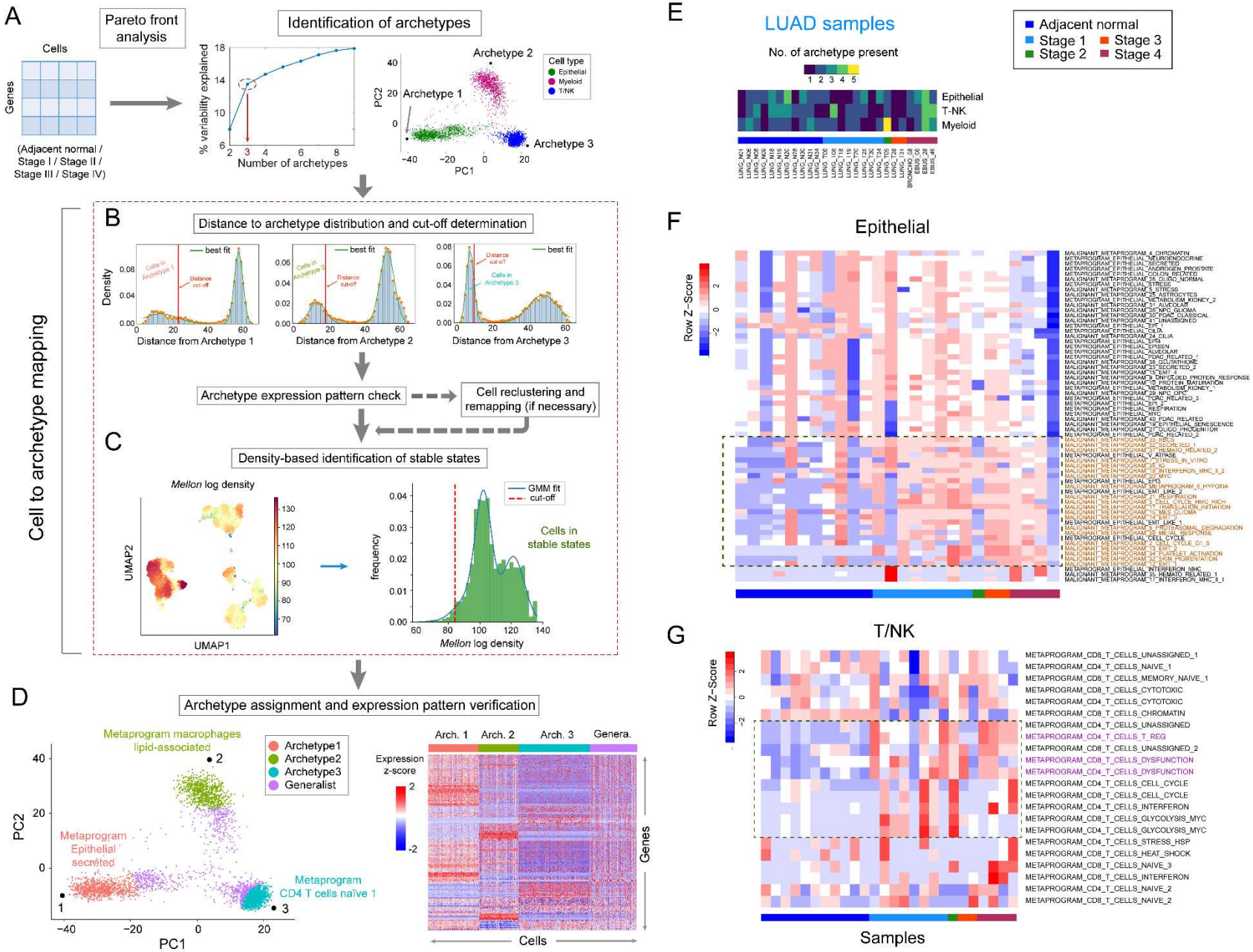

A
Pareto front analysis
Identification of archetypes
Cells
Genes
(Adjacent normal / Stage I / Stage II / Stage III / Stage IV)
% variability explained
Number of archetypes
Archetype 1
Archetype 2
Archetype 3
Cell type
Epithelial
Myeloid
T/NK
PC1
PC2
Cell to archetype mapping
B
Distance to archetype distribution and cut-off determination
best fit
Density
Distance from Archetype 1
Distance from Archetype 2
Distance from Archetype 3
Archetype expression pattern check
Cell reclustering and remapping (if necessary)
C
Density-based identification of stable states
Mellon log density
UMAP1
UMAP2
GMM fit
cut-off
Cells in stable states
frequency
D
Archetype assignment and expression pattern verification
Metaprogram macrophages lipid-associated
Archetype1
Archetype2
Archetype3
Generalist
Metaprogram Epithelial secreted
Metaprogram CD4 T cells naïve 1
Arch. 1
Arch. 2
Arch. 3
Genera.
Expression z-score
Genes
Cells
E
LUAD samples
No. of archetype present
Adjacent normal
Stage 1
Stage 2
Stage 3
Stage 4
Epithelial
T-NK
Myeloid
F
Epithelial
Row Z-Score
G
T/NK
Samples

Figure 1 – Analysis of intra-tumor functional states reveal functional diversification and functional transformation with cancer progression. **(A)** Single-cell gene expression data from individual tumor or adjacent normal sample was first subjected to Pareto front analysis (using *ParTI*) which enabled identification of functional archetypes within a sample**. (B)** Each cell was assigned an archetype (a specialist) or considered a generalist based on a distance to archetype calculation and filtering by a distance cut-off. The expression patterns of cells in each archetype were checked and in case of a heterogeneous expression pattern, re-clustering of cells and remapping to specialist clusters were done. **(C)** In the next step, a density-based cell-state stability analysis (using *Mellon*) was performed to identify cells in stable functional states. **(D)** The cells were assigned archetypes from stability analysis and the expression patterns within each archetype was checked by heatmap. **(E)** Heatmap showing the number of archetypes for which each cell type (Epithelial, T/NK and Myeloid) was part of in the LUAD patient samples, thereby, depicting their functional diversity. **(F)** Heatmap showing changes in utilization of functional metaprograms (as defined by Gavish et al., 2023) in epithelial cells in LUAD samples with cancer progression. The metaprograms highlighted in red show some malignant metaprograms whose utilization increased with cancer progression. The colour bar at the bottom of the heatmap show stages for cancer samples along with adjacent normal samples. **(G)** Heatmap showing changes in utilization of functional metaprograms in T/NK cells in LUAD samples with cancer progression. The metaprograms in purple show functions associated with regulatory T cells and T cell exhaustion.

Next, we assigned functional identity of all specialist archetypes in a sample through a functional enrichment analysis. Briefly, we tested for enrichment of metaprograms, as defined by Gavish et al. (2023), individually for all archetypes. For each archetype and for every metaprogram, we compared the expression of genes associated with the metaprogram in the cells of the archetype and the rest of the cells in the sample (see Methods). Functions with significant enrichment in the archetype (p-adjusted <0.1 after FDR correction) were ranked according to significance and these programs defined functional state of the cells in the archetype.

Individual cell types harbour distinct functional clusters and show signatures of functional transformation with cancer progression

In all cell types across LUAD and CRC, we observed substantial functional diversity and each cell type was part of multiple specialist archetypes across a majority of patient samples (Figures 1E and S2). We did not observe any cancer stage-specific bias in functional diversity and cell types displayed widespread functional diversity even in normal samples in both LUAD and CRC (Figures 1E and S2).

However, the relative ranking of the enriched functional programs within samples and cell types changed with cancer stages in both LUAD and CRC (Figures S3A-C and S4A-D). In epithelial cells in LUAD, the top ranked malignant programs included the metaprograms related to epithelial senescence ('Malignant metaprogram 19 epithelial senescence'), and epithelial secretion ('Malignant metaprogram 23 secreted 2') (Fig. S3A). In CRC, the metaprograms linked to epithelial Myc ('Metaprogram epithelial MYC') and 'Metaprogram epithelial respiration' were among the top ranked metaprograms in the cancer samples but not in adjacent normal samples (Fig. S4A). We observed similar patterns in other cell types. Notably, some of the CD8 T cells programs in T/NK cells in LUAD (Fig. S3C), several macrophage-associated metaprograms ('Macrophages MAC3', 'Macrophages MES glycolysis', 'Macrophages monocyte secreted') in myeloid cells in CRC (Fig. S4B), and several malignant metaprograms (including the metaprogram associated with EMT) in stromal cells in CRC (Fig. S4D) were among the top-ranked programs that were preferentially enriched in cancer samples.

We also observed changes in utilization of functional metaprograms in individual cell types with cancer progression (Figures 1F,G and S5) by quantifying the percentage of specific cell types carrying signature of individual metaprograms. In epithelial cells of both LUAD and CRC, the utilization of several malignant programs increased from normal to tumor samples (Figures 1F and S5A), highlighting the functional transformation of the epithelial cells in the TME. In T/NK cell population in LUAD, the metaprograms associated with T cell exhaustion and regulatory T cells showed increased utilization in cancer samples compared to normal samples (Fig. 1G). In T cells in CRC, the metaprograms associated with T cell exhaustion and regulatory T cells also increased in stage II and III cancer samples compared to normal (Fig. S5B), but were not seen in stage IV samples in this dataset. Taken together, these results highlighted the functional transformation of tumor with cancer stage.

Cell-state stability analysis reveals existence of transitory functional states in both LUAD and CRC

Density-based cell-state stability analysis enabled us to identify low-frequency transitory functional states - that were represented by low-density regions - in each tumor sample. For each cell in transitory state, we assigned the two nearest archetypes based on Euclidean distance to archetypes in Pareto front map (Fig. 2A), and considered these cells to be of intermediate functional phenotype. The transitory functional states were more abundant in tumor samples compared to normal samples in both LUAD and CRC. In LUAD, ~87% of the tumor samples harboured transitory states compared to ~46% of normal

samples (Fig. 2B). Similarly, in CRC, ~91% tumor samples harboured transitory states compared to ~44% normal samples (Fig. 2B).

We also observed two types of transitory functional states. There were transitory states between archetypes harbouring the same cell type (within-cell type transitory states) and between archetypes harbouring different cell types (between-cell type transitory states) (Fig. 2B). Intriguingly, the transitory states between different cell types were more abundant in cancer samples compared to normal samples (Fig. 2B). In LUAD, ~80% tumor samples harboured between-cell type transitory states compared to ~27% normal samples (Fig. 2B). Similarly, in CRC, ~74% tumor samples harboured between-cell type transitory states compared to ~33% adjacent normal samples (Fig. 2B).

We also observed a bias in occurrence of transitory functional states between specific cell types. In both LUAD and CRC, the transitory states between archetypes harbouring predominantly myeloid cells (myeloid-dominant) and archetypes harbouring predominantly T/NK or T cells (T/NK- or T-dominant) were highly enriched in cancer samples (Fig. 2C). In LUAD, ~67% tumor samples harboured transitory functional states between myeloid-dominant and T/NK-dominant archetypes, whereas only ~18% normal samples harboured such transitory states. Similarly, in CRC, ~44% tumor samples harboured transitory states between myeloid and T/NK archetypes compared to ~11% normal samples (Fig. 2C). Further, transitory states between epithelial-dominant and T-dominant archetypes were observed in CRC where ~52% CRC tumor and ~22% normal samples harboured these transitory states (Fig. 2C). These results suggested a potential role of cells residing in between-cell types transitory states and specifically, between myeloid-dominant and T or T/NK-dominant archetypes in cancer progression.

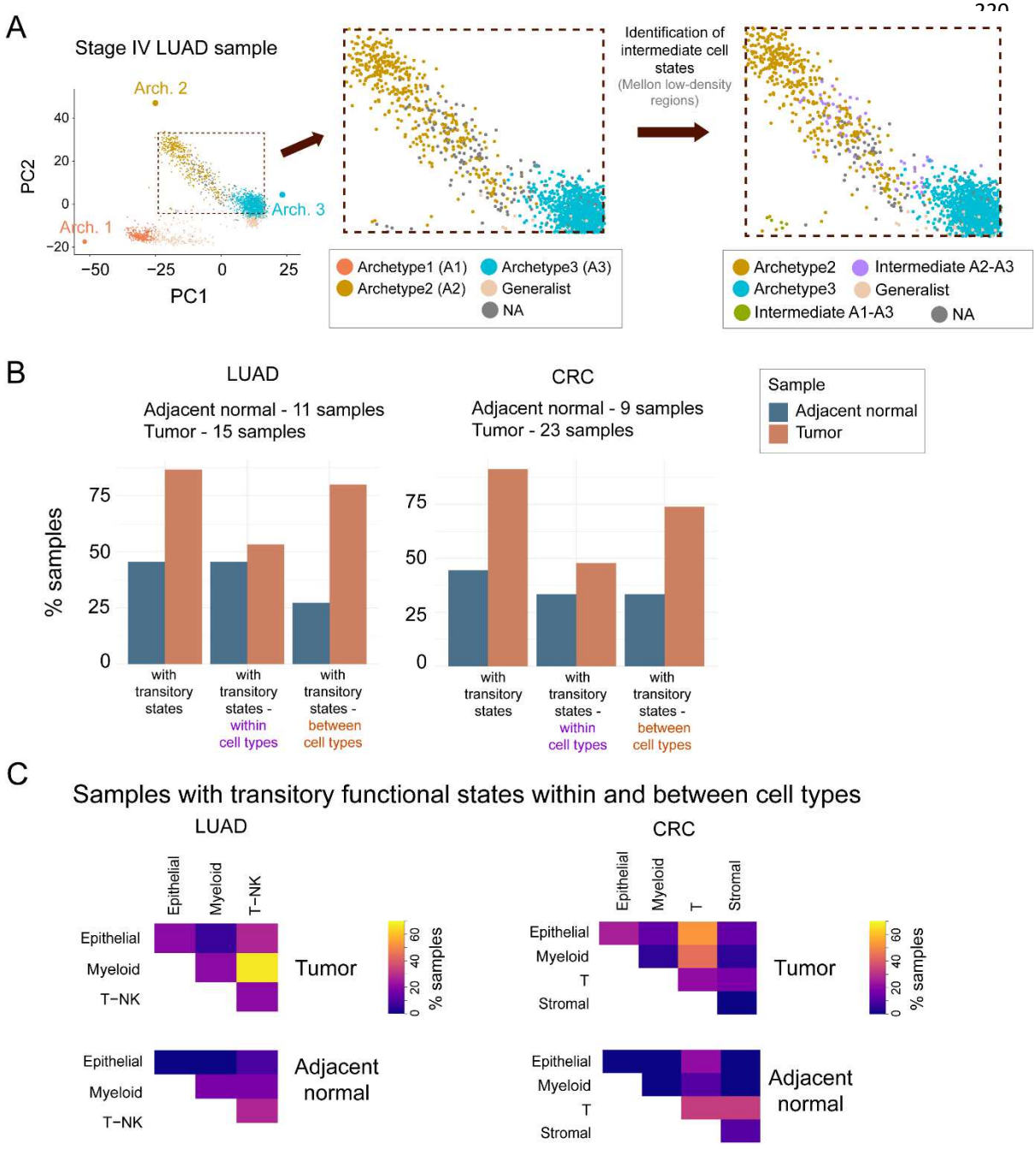
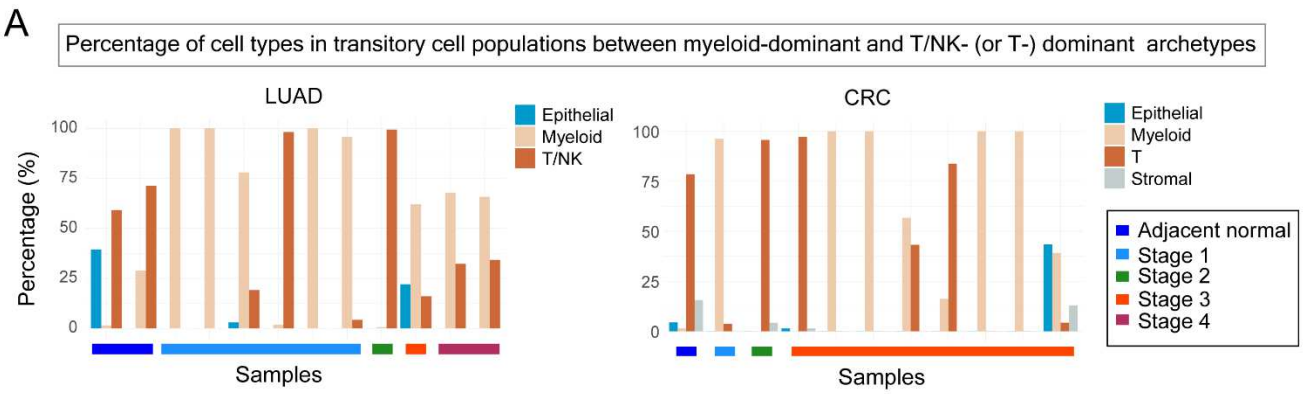


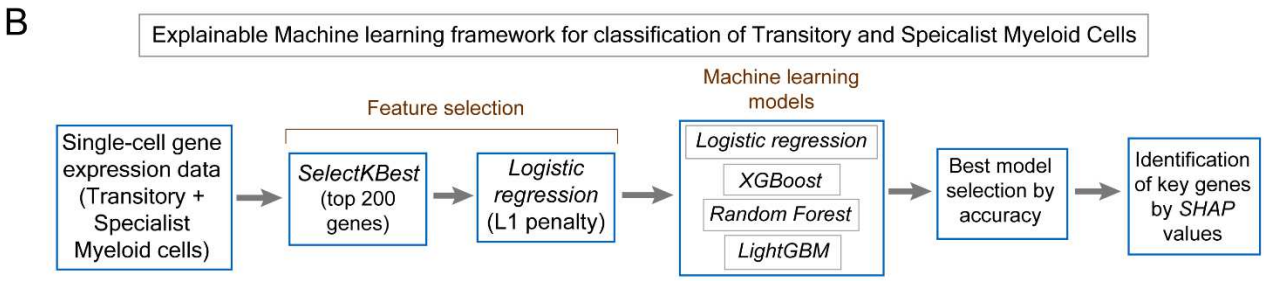
Figure 2 – Transitory functional states are more prevalent in tumor samples compared to adjacent normal samples. (A) The cells in transitory functional states were identified from density-based cell-state stability analysis using *Mellon* (Otto *et al.*, 2024) where cells in low-density regions and not assigned any archetype in Pareto front analysis were considered to be in transitory states. For each transitory cell, the nearest two archetypes were identified from distance calculation to archetypes and the cell was considered to be of intermediate phenotype. **(B)** Percentage of samples harbouring transitory states, transitory states within cell types and transitory states between cell types in LUAD and CRC samples. **(C)** Heatmaps showing the percentage of samples harbouring different within cell type and between cell type transitory states.

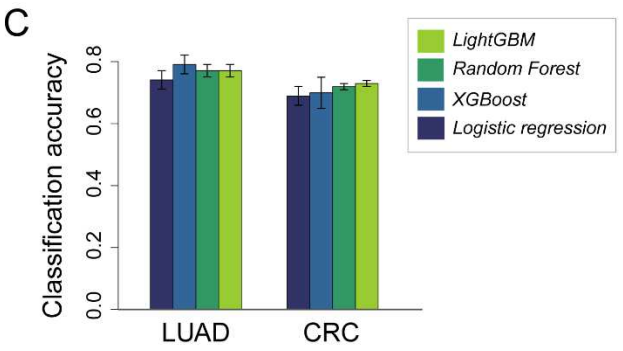
Transitory myeloid states are characterized by distinct gene expression signatures

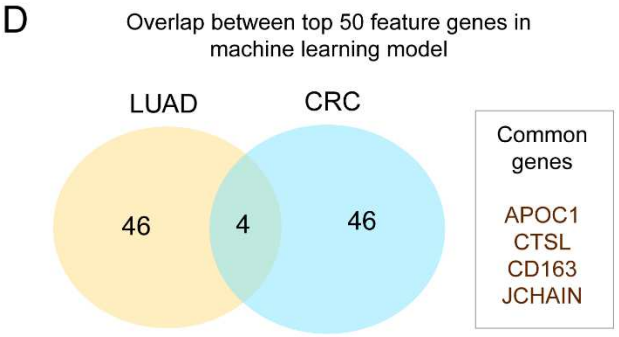
As transitory functional states between myeloid-dominant and T/NK or T-dominant archetypes were enriched in tumor samples of both LUAD and CRC, we specifically investigated the molecular differences between the transitory and specialist myeloid cells. In LUAD, among eight samples with presence of the myeloid-T/NK transitory states, myeloid cells were in majority (>50% frequency in the transitory cell population) in eight samples, whereas T/NK cells were in majority in four samples (Fig. 3A). Similarly, in CRC, among the samples with the presence of myeloid-T transitory states, myeloid cells and T cells were in majority in six and four samples, respectively (Fig. 3A). In LUAD, the transitory myeloid cells ranged up to 5-10% of the total myeloid population of individual samples, with a couple of exceptions where the transitory myeloid cells were ~25-30% of the myeloid population (Fig. S6). In CRC, the transitory myeloid cells were up to 25% of the total myeloid population of individual samples (Fig. S6).

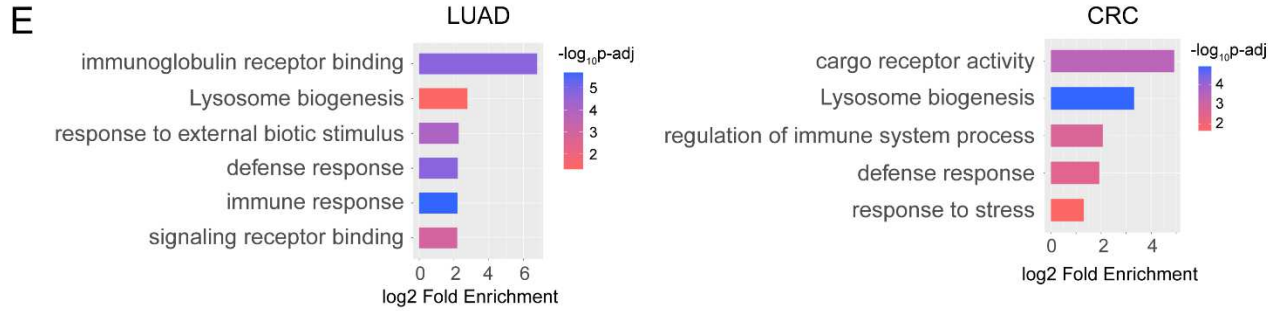
To better understand the functional role of the transitory myeloid cells in the tumor microenvironment, we specifically looked for unique gene expression signatures of the transitory myeloid cells compared to specialist myeloid cells. To do so, we deployed explainable machine learning models to classify myeloid cells into transitory and specialist myeloid states and identified the underlying gene expression signatures that contributed to the classification. Briefly, we started with data containing functional state identity (transitory or specialist myeloid) of the cells and expression data of ~26,000 genes in each cell. We then performed down-sampling of overrepresented classes to reduce data imbalance and subsequently, divided the full data into training (75%) and test (remaining 25% of the full data) data (Fig. 3B). We applied four machine learning models (viz. *Logistic Regression*, *Random Forest*, *XGBoost* and *Light GBM*) and selected the model with the highest accuracy for subsequent inferences (Fig. 3B).


A
Percentage of cell types in transitory cell populations between myeloid-dominant and T/NK- (or T-) dominant archetypes
LUAD
CRC
Percentage (%)
Samples
Epithelial
Myeloid
T/NK
T
Stromal
Adjacent normal
Stage 1
Stage 2
Stage 3
Stage 4


B
Explainable Machine learning framework for classification of Transitory and Speicalist Myeloid Cells
Single-cell gene expression data (Transitory + Specialist Myeloid cells)
Feature selection
SelectKBest (top 200 genes)
Logistic regression (L1 penalty)
Machine learning models
Logistic regression
XGBoost
Random Forest
LightGBM
Best model selection by accuracy
Identification of key genes by SHAP values


C
Classification accuracy
LightGBM
Random Forest
XGBoost
Logistic regression
LUAD
CRC


D
Overlap between top 50 feature genes in machine learning model
LUAD
CRC
46
4
46
Common genes
APOC1
CTSL
CD163
JCHAIN


E
LUAD
immunoglobulin receptor binding
Lysosome biogenesis
response to external biotic stimulus
defense response
immune response
signaling receptor binding
-log10p-adj
log2 Fold Enrichment
CRC
cargo receptor activity
Lysosome biogenesis
regulation of immune system process
defense response
response to stress
-log10p-adj
log2 Fold Enrichment

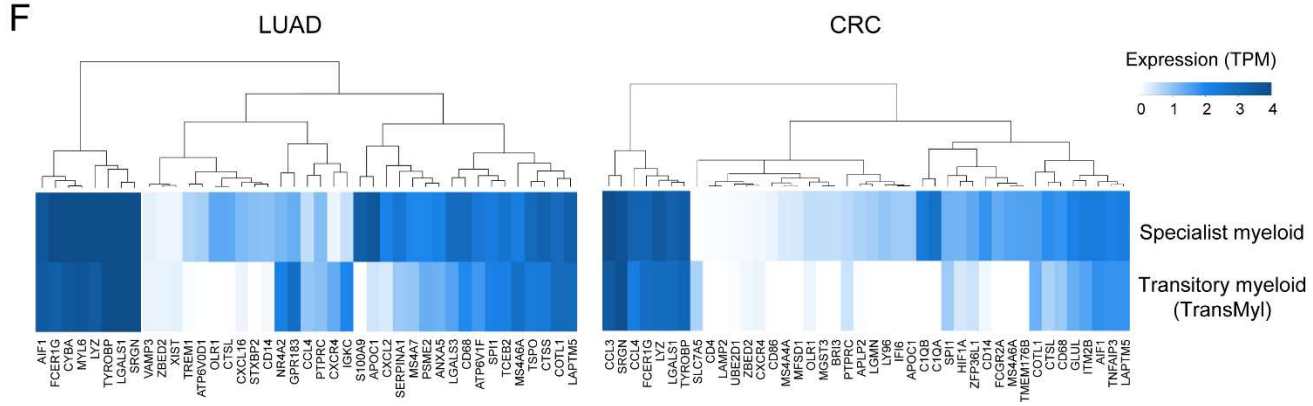

F
LUAD
CRC
Expression (TPM)
Specialist myeloid
Transitory myeloid (TransMyl)

Figure 3 – Transitory myeloid cells show distinct expression patterns compared to specialist myeloid cells. (A) Bar plots showing percentage of different cell types in transitory cell populations between myeloid-dominant and T/NK-dominant archetypes (in LUAD) or myeloid-dominant and T-dominant archetypes (in CRC) in individual samples. The colour bar at the bottom of the heatmap show stages for cancer samples along with adjacent normal samples. **(B)** The explainable machine learning framework for classification of transitory and specialist myeloid cells and for identification of key feature genes that enable this classification. **(C)** Accuracy in classification of transitory and specialist myeloid cells in LUAD and CRC by *Logistic regression, XGBoost, Random Forest* and *LightGBM* models. **(D)** The overlap between the top 50 feature genes identified by machine learning models in LUAD and CRC. **(E)** Enriched functional classes in the top 50 feature genes in LUAD and CRC along with their log2 Fold enrichment (x-axis). The colour indicates the $-\log_{10}$(p-adjusted) value (obtained after FDR correction) for the functional class. **(F)** Heatmaps showing expression of some of the feature genes and some well-known marker genes in specialist and transitory myeloid cells in LUAD and CRC.

In LUAD and CRC, the machine learning models were able to classify cells between transitory and specialist myeloid cells with ~79% and ~73% accuracy with *XGBoost* and *LightGBM* models showing the highest accuracy, respectively (Fig. 3C). These results indicated that the transitory myeloid cells share some specific characteristics that differentiate them from specialist myeloid cells but also harbour heterogeneity among themselves. To better understand the key characteristics of transitory cells, we further analysed these models to identify the key molecular features that distinguish them from specialist myeloid cells. To do so, we evaluated each molecular feature through Shapley Additive Explanations (SHAP) framework (Lundberg and Lee, 2017) where higher SHAP value of a feature indicated higher contribution towards distinguishing functional states (Fig. 3C, D). Functional enrichment of the top 50 feature genes (Table S2) identified several gene ontology functions that were enriched in both LUAD and CRC (p-adjusted < 0.05 after FDR correction). The functions associated with immunoglobulin receptor binding, immune response, and signalling receptor binding were highly enriched with log2 Fold enrichment > 1 (p-adjusted < 0.05) (Fig. 3E). A large majority of these feature genes were downregulated in transitory myeloid cells compared to special myeloid cells (Fig. 3F). Thus, the transitory myeloid cells were likely to have poorer ability in presenting antigens, binding immune receptors and activating immune system.

Occurrence of transitory myeloid states is associated with signatures of T cell exhaustion and immune suppression

To better understand the functional implications of the transitory myeloid states on the tumor microenvironment, we specifically tested whether these cells represented myeloid-derived suppressor

cells (MDSCs) and analysed expression signatures of well-known MDSC marker genes. More specifically, we looked for signatures of monocytic MDSC (M-MDSC) ($CD11b^{+}CD14^{-}CD15^{+}$ (or $CD66b^{+}$)), polymorphonuclear MDSCs (PMN-MDSCs) ($CD11b^{+}CD14^{+}HLA\text{-}DR^{low/-}CD15^{-}$) and early-stage MDSCs (Lin^{-}(CD3/14/15/19/56)/$HLA\text{-}DR^{-}/CD33^{+}$) (Bronte et al., 2016). However, we did not find any match for these expression signatures in neither the transitory nor the specialist myeloid cells (Fig. 4A-C).

In LUAD and CRC, expression of *ITGAM* (encoding for CD11b) and *CD14* genes were lower (~19-41% and ~13-44% respectively for mean TPM value) in the transitory myeloid cells compared to specialist myeloid cells (Fig. 4A), whereas expression of *FUT4* (encoding for CD15) and *CEACAM8* (encoding for CD66b) were extremely low and very similar in transitory and specialist myeloid cells (Fig. 4A). Thus, the gene expression signature for M-MDSCs were not present in the transitory and specialist myeloid cells in LUAD or in CRC. The difference in expression of HLA-DR genes varied in LUAD and CRC (Fig. 4B). While the expression of *HLA-DRA* was marginally higher (~2%) in transitory myeloid cells in LUAD, it was marginally lower (~6%) in transitory myeloid cells in CRC, compared to specialist myeloid cells (Fig. 4B). Thus, these expression patterns also ruled out similarities of transient myeloid cells with PMN-MDSCs. The expression of *CD33* gene was similar between transitory and specialist myeloid cells in LUAD cells, but its expression was very low in myeloid cells in CRC (Fig. 4C). This observation, combined with expression of other marker genes ruled out the presence of signatures of e-MDSCs in the transitory or the specialist myeloid cells (Fig. 4C). Thus, we did not find the presence of signatures of any traditional MDSCs in LUAD and CRC samples included in our analysis. We also noted that the differences in expression of CD14 and CD11b (ITGAM) were not due to inherent differences in the overall expression levels of these genes between the samples, which were largely similar across adjacent normal and tumor samples (Figures S7 and S8). We also tested for other markers of MDSCs reported specifically for non-small cell lung cancer (NSCLC) and colon cancer. However, none of these marker combinations could describe the transitory myeloid cells identified in our analysis neither for LUAD nor for CRC (Figures S9 and S10).

Among the top 50 feature genes identified from the machine learning models, four genes – *APOC1, CTSL, CD163* and *JCHAIN* – were shared between LUAD and CRC data. The genes *APOC1, CTSL,* and *CD163* were all downregulated in the transitory myeloid cells compared to specialists, whereas *JCHAIN* was upregulated in transitory myeloid cells in both LUAD and CRC (Fig. 4D). Intriguingly, the expression of the *LYZ* gene – encoding for lysozyme and a typical marker of myeloid cells – was also lower in the transitory cells (~27% lower in LUAD and ~22% lower in CRC) (Fig. 4E). *LYZ* has an

important role in immune-modulation (Wang et al., 2026) and thus, decreased expression of this gene can lower immune activity. In addition, we observed downregulation of genes encoding for the complement C1Q complex in transitory myeloid compared to specialist myeloid cells (Fig. 4F). C1Q complex has been associated with antigen presentation in myeloid cells and also has an important role in modulation of innate immune memory (van Montfoort 2007; Jonkman et al., 2025). In contrast, the expression of CXCR4 receptor – whose downregulation is associated with increased antitumor immunity (Yang et al., 2018) – showed ~34-54% higher expression in transitory myeloid cells compared to specialists (Fig. 4G). These results hinted at a potential influence of the transitory myeloid cells on the tumor immune microenvironment in both LUAD and CRC.

We further analysed the activity of transcription factors in transitory myeloid cells to identify potential drivers of transitory functional states in LUAD samples. Briefly, we reconstructed the gene regulatory network for transitory myeloid cells and specialist myeloid cells within individual samples following an approach by Ray et al. (2026). For each cell population (transitory or specialist myeloid), we calculated a pair-wise correlation between all genes, performed a filtering of connections from a validated TF-target map, and identified all active regulatory connections through a threshold-based analysis (see Methods). Subsequently, we evaluated the activity of all TFs in individual cells by scoring the activity of the TF-associated regulons using AUCell (Aibar et al., 2017) and identified the top ranked TFs in each subpopulation by their AUCell scores. The transitory and specialist myeloid cells within a sample showed a substantial overlap between the key TFs and their importance (Fig. 4H), but we also found a few TFs which were only active in transitory myeloid cells (Fig. 4H). In addition, comparison of active TFs in transitory myeloid cells across samples revealed minimal overlap between the sets of active TFs in different samples (Fig. 4I). However, a few TFs - including ZNF143, ZBTB4, ZBTB33 and IRF4 - were among the most active TFs in transitory myeloid cells of more than four LUAD samples (Fig. 4I).

To further understand the influence of the transitory myeloid cells on the tumor immune microenvironment, we tested for differences in expression patterns of T/NK cells (or T cells in CRC) in samples harbouring transitory myeloid cells and those without using two complementary approaches. First, we built machine learning models for classification of T/NK (or T) cells into two classes – with and without transitory myeloid states (Fig. 5A). The machine learning models – specifically, *XGBoost,* and *LightGBM* – were able to classify the T cells into two classes with ~85% accuracy in both LUAD and CRC (Fig. 5A), suggesting distinct molecular differences between T/NK or T cell phenotype between the samples that harboured transitory myeloid states and the samples that did not.



Figure 4 – Transitory myeloid cells do not possess the molecular signatures of traditional myeloid derived suppressor cells (MDSCs). (A) Boxplot showing expression of signature genes of monocytic MDSCs (M-MDSCs) in transitory and specialist myeloid cells in LUAD and CRC samples. **(B)** Boxplot showing expression of signature genes of polymorphonuclear MDSCs (PMN-MDSCs) in transitory and specialist myeloid cells in LUAD and CRC samples. **(C)** Boxplot showing expression of signature genes of early-stage MDSCs (e-MDSCs) in transitory and specialist myeloid cells in LUAD and CRC samples. **(D)** Boxplots showing expression of four feature genes that were identified by machine learning models and shared between LUAD and CRC samples (*APOC1, CTSL, CD163* and *JCHAIN*). **(E)** Expression of the myeloid marker gene *LYZ* in transitory and specialist myeloid cells. **(F)** Expression of C1Q complex genes (*C1QA* and *C1QC*) in transitory and specialist myeloid cells. **(G)** Expression of the gene encoding for CXCR4 receptor in transitory and specialist myeloid cells. **(H)** The

activity of the top ranked active TFs – as identified by gene regulatory network reconstruction and AUCell scores – in transitory and specialist myeloid cells of LUNG_T25 (Stage 1), LUNG_T09 (Stage 2), LUNG_T28 (Stage 3) and EBUS_06 (Stage 4) samples. **(I)** Comparison of the activity of the top 10 active TFs in transitory myeloid cells across LUAD samples.

In the next step, we performed differential gene expression analysis between T/NK (or T) cells from samples with and without transitory myeloid states using the MAST framework (Finak et al., 2015). In LUAD, we observed upregulation of 118 genes and downregulation of 128 genes in T/NK cells from samples harbouring transitory myeloid states compared to the T/NK cells from samples without transitory states (Fig. 5B) (Table S3). In CRC, we observed upregulation of 38 genes and downregulation of 17 genes in T cells from samples harbouring transitory myeloid states compared to the T cells from samples without transitory states (Fig. 5B) (Table S4). Among the upregulated genes, 7 were shared and among the downregulated genes, 9 were shared between LUAD and CRC (Fig. 5B,C). In LUAD, the top upregulated gene Regulator of G-receptor signalling 1 (RGS1) has been shown to promote immune evasion in cancer upon upregulation (Huang et al., 2021) and the top downregulated gene – GNLY, encoding for granulysin, is a well-known component of cytotoxic activity of T and NK cells (Huang et al., 2007; Tewary et al., 2010) (Fig. 5D). Similarly, in CRC, the top upregulated gene – HSPA1A, a heat-shock protein, has been linked to poor patient survival in colon cancer (Guan et al., 2021), whereas the top downregulated gene – CREM – is linked with immune modulation in NK cells (Rafei et al., 2025) (Fig. 5D).

Among the downregulated genes, we observed significant enrichment of the functions associated with oxidative phosphorylation and electron transport chain in both LUAD and CRC (log2 Fold enrichment > 1 and p-adjusted <0.05 after FDR correction) (Fig. 5E,F). In addition, in LUAD, we specifically observed downregulation of functions associated with cell killing, T cell activation and immune response (Fig. 5E). In CRC, we observed downregulation of immunoglobulin receptor binding activity (Fig. 5F). Downregulation in activation of oxidative phosphorylation genes in T cells has been linked to T cell exhaustion (Vardhana et al., 2020).

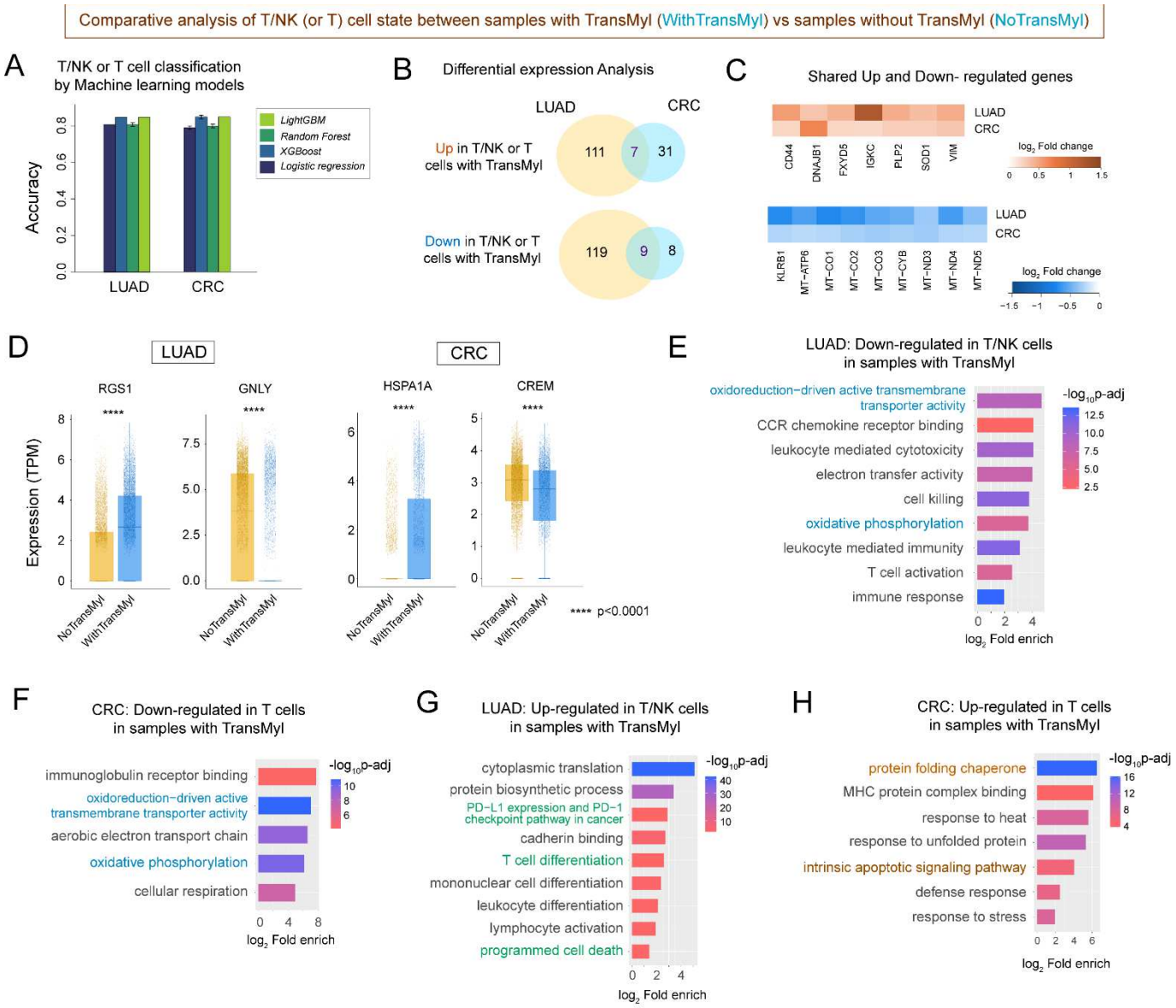


Figure 5 – Tumor samples harbouring transitory myeloid cells show signatures of higher T cell exhaustion and immune suppression. (A) Accuracy of classification between T/NK (or T) cells from samples harbouring transitory myeloid states and T/NK (or T) cells from samples without any transitory myeloid states by machine learning models (*Logistic Regression, XGBoost, Random Forest,* and *LightGBM*). **(B)** Number of up- and down-regulated genes in T/NK (or T) cells from samples harbouring transitory myeloid states compared to T/NK (or T) cells from samples without transitory myeloid states identified through differential gene expression analysis in LUAD and CRC. **(C)** Heatmaps showing expression of up- and down- regulated genes shared between LUAD and CRC samples. **(D)** Expression (TPM) of the top-ranked up- and down-regulated genes in T/NK (or T) cells in samples without transitory myeloid (NoTransMyl) and with transitory myeloid (WithTransMyl) cells in LUAD and CRC. **(E)** Enriched functional classes along with their log2 Functional enrichment (x-axis) among the genes downregulated in T/NK cells in samples with TransMyl in LUAD. **(F)** Enriched functional classes along with their log2 Functional enrichment (x-axis) among the genes downregulated in T cells in samples with TransMyl in

CRC. **(G)** Enriched functional classes along with their log2 Functional enrichment (x-axis) among the genes upregulated in T/NK cells in samples with TransMyl in LUAD. **(H)** Enriched functional classes along with their log2 Functional enrichment (x-axis) among the genes upregulated in T cells in samples with TransMyl in CRC. The colour of the bars in plots F to H indicate $-\log_{10}$(p-adjusted) values obtained after FDR correction.

Among the upregulated genes in T cells in samples with presence of transitory myeloid states, we observed significant enrichment of function associated with programmed cell death and apoptosis in both LUAD and CRC (log2 Fold enrichment > 1 and p-adjusted <0.05 after FDR correction) (Fig. 5G,H). In addition, in LUAD, we observed upregulation of pathway related to PD-L1 expression/PD-1 checkpoint pathway and T-cell differentiation (Fig. 5G). In CRC, we observed upregulation of pathway linked to unfolded protein response and heat response (Fig. 5H). Collectively, these results pointed to a more immunosuppressive microenvironment in tumor samples that harboured transitory myeloid states.

Different cell types across patients show functional plasticity and convergence

Our analysis so far focused on functional archetypes and transitory states for individual patient samples. However, it was not clear whether different cell types across patients of a cancer type largely align together and whether different cell types can perform similar functional role across patients. Thus, to investigate the landscape of functional plasticity across patients we performed a global alignment of functional archetypes and calculated pairwise similarity scores for all archetypes from all patients. The similarity score between two archetypes reflected the extent of similarity between the enriched metaprograms between them. In both LUAD and CRC, we observed distinct functionally similar clusters among archetypes (Figures 6A and S11).

Intriguingly, we found some functional archetypes where more than one cell type was present (Figures 6B and S11B), suggesting functional plasticity and convergence of sub-populations of different cell types within individual samples (intra-sample plasticity). Further, we also observed functional alignment of archetypes represented by different cell types across different samples, suggesting functional convergence across cell types (inter-sample plasticity) (Fig. 6B and S11B). Ranking of metaprograms within one such cluster each in LUAD and in CRC revealed similar importance (and ranking) of metaprograms within the archetypes (Fig. 6C and S11C).

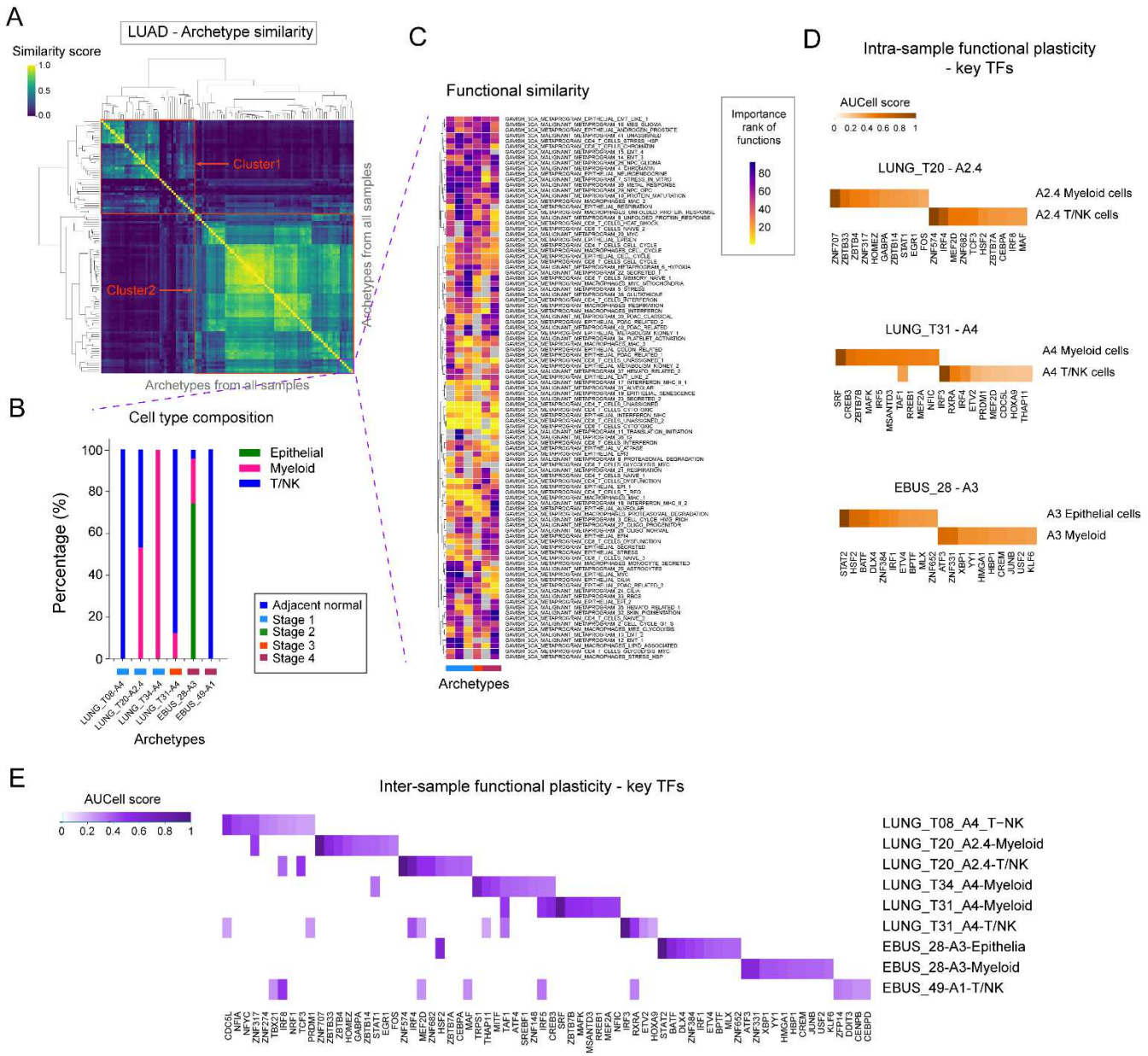


Figure 6 – Intra- and inter-sample functional plasticity and functional convergence of different cell types is driven by distinct regulatory modules. (A) Heatmap depicting pairwise similarity scores between archetypes of all samples obtained after alignment. **(B)** Percentage of different cell types in archetypes that show high similarity scores in LUAD samples. **(C)** Heatmap showing the importance rank of functional metaprograms within the archetypes in (B) that show high similarity score. **(D)** The top-ranked TFs and their activity (as defined by their AUCell scores) in different cell types within individual archetypes in different LUAD samples (Intra-sample plasticity). **(E)** Heatmap showing the top-ranked TFs and their activity in archetypes across different samples that show high similarity score (as identified in (B)).

To test whether intra-sample and inter-sample functional plasticity were driven by similar gene regulatory modules across different cell types, we reconstructed the gene regulatory network for each functional types within each archetype in LUAD samples following an approach by Ray et al. (2026) and evaluated the activity of all TFs and their regulons in each cell within the archetype by AUCell and ranked the top transcription factors based on AUCell scores. Intriguingly, the top 10 TFs among different cell types within an archetype showed no or minimal overlap (Fig. 6D), suggesting functional convergence driven by different regulatory modules. Further, the top ranked TFs driving inter-sample functional plasticity and convergence also showed minimal overlap among different cells even among the same cell types (Fig. 6E). Collectively, these results highlighted existence of intra- and inter-sample functional plasticity and convergence that were driven by distinct regulatory programs in individual patients.

Discussion

Taken together, the present work elucidates the landscape of functional diversity, transition and plasticity in cancer samples. We observed wide-spread functional diversity within individual cell types, in line with earlier findings (Barkley et al., 2022; Gavish et al., 2023). We also observed intra-tumor functional transformation with cancer progression which was reflected in altered utilization of some of the metaprograms in tumor and immune cells. In addition, our work uncovered functional plasticity and convergence among different cell types within and between tumor samples, and showed that distinct regulatory programs can also drive functional convergence between different cell types.

A key finding of our work is the discovery of transitory functional states that were more frequent in tumor samples compared to normal samples. The transitory functional states were observed specifically in myeloid cells in both LUAD and CRC, and these transitory myeloid cells had distinct expression signatures compared to specialist myeloid cells. These cells had lower expression of genes associated with receptor binding and immune response. In addition, occurrence of transitory myeloid cells in tumor samples was associated with a signature of T cell exhaustion and immune suppression in both cancer types. However, these cells did not carry the gene expression signatures of well-known myeloid-derived suppressor cells (MDSCs). Thus, these myeloid cells likely represent a novel or atypical class of myeloid suppressors. These results also revealed the finer differences between adjacent normal and tumor samples, as transitory myeloid cells occurred preferentially in tumor samples.

The gene expression signatures of transitory myeloid cells differed substantially between LUAD and CRC samples with only four shared genes among the top 50 features (Fig. 4D). This points to diverse

molecular origins of the transitory cells in LUAD and CRC. In addition, the molecular origins of transitory cells are likely to be different from one patient to another within a cancer type and thus, the transitory states could actually be an ensemble of many rare states. This was supported by the observation that the transitory myeloid cells in different samples showed diverse active gene regulatory programs and transcription factors with minimal overlap between samples. However, the resolution of our datasets did not allow us to look more deeply into the multitude of transitory states, and analysis of larger samples of single-cell data will be necessary for a better understanding. Moreover, it remains to be seen whether the transitory myeloid cell states carry a unifying set of markers that could be observed across patients and different cancer types, similar to the traditional myeloid-derived suppressor cells. Further, the occurrence of transitory myeloid states was associated with signatures of immune suppression but this association does not establish a causal role. Thus, more in-depth work is required to establish the mechanistic link between the presence of transitory myeloid cells in cancer samples and T cell exhaustion and immune suppression.

Despite these limitations, our work elucidated the landscape of intra-tumor functional transitions and transformation with cancer progression. Our work uncovered rare functional states in the tumor microenvironment in patient samples from two different cancer types and linked them to immune suppression in both of them. These results suggest important implications of these transitory functional states for cancer progression and therapy response. However, understanding their exact role in cancer microenvironment will require more in-depth studies. Although rare cell states have been found to be important for determining therapy response (Emert et al., 2021), most studies have been done in cell culture experiments which fail to capture the complexity of the tumor microenvironment. Thus, future work linking the occurrence of transitory myeloid cells with therapy response in patient samples will be necessary to establish the role of these functional states in determining cancer progression and therapy outcome. Further, analysis across a more diverse cohort of cancer types will reveal a more global picture of the transitory functional states and will enable us to gauge their importance in cancer.

Methods

Datasets

To investigate functional transition and plasticity in the tumor microenvironment with cancer progression, patient derived single cell transcriptomic dataset with cancer stage information as well as cell type information was required. In addition, multiple samples per stage were preferred for more robust inferences.

One such dataset of lung adenocarcinoma (LUAD) was obtained from Kim et al. (2020) (NCBI GEO accession number: GSE131907). The dataset contained single-cell RNA sequencing data of 208,506 cells derived from 58 samples of 44 patients. Among 58 samples, 26 samples of adjacent normal, early-stage, and advanced-stage lung adenocarcinoma from primary sites of the lung were chosen for our analysis. Single-cell RNA sequencing data from cells that are located in the brain, lymph node, or pleural fluids were excluded from this study. This was because gene expression is influenced by the metastasis process as well as by distinct tissue environments, which could confound our analysis. The dataset also contained information on different cell types including Myeloid, Epithelial, B cells, T/NK cells, Fibroblasts, Endothelial and Mast cells. However, some cells in this dataset were unannotated. Thus, the cell type of these cells was predicted by Seurat v4 (Hao et al., 2021) and SciBet (Li et al., 2020), as discussed in Ray et al. (2026). For this study, Epithelial, Myeloid, and T/NK cells were used as they were present in substantial numbers across all 26 patient samples which was required to identify functional specializations (Fig. S12A). Out of the 26 samples collected from lung cancer patients, there were 11 adjacent normal samples, 7 stage I samples, 1 stage II sample, 2 stage III samples, and 4 stage IV samples. “LUNG_T06”, “LUNG_T08”, “LUNG_T18”, “LUNG_T19”, “LUNG_T20”, “LUNG_T25”, “LUNG_T30” and “LUNG_T34” were Stage I samples, “LUNG_T09” was Stage II sample, “LUNG_T28” and “LUNG_T31” were Stage III samples and “EBUS_06”, “EBUS_28”, “EBUS_49” and “BRONCHO_58” were Stage IV samples. “LUNG_N01”, “LUNG_N06”, “LUNG_N08”, “LUNG_N09”, “LUNG_N18”, “LUNG_N19”, “LUNG_N20”, “LUNG_N28”, “LUNG_N30”, “LUNG_N31”, “LUNG_N34” were adjacent normal patient samples and were also analysed for comparison with tumor samples (Table S1).

To test our pipeline, an independent single cell transcriptomic dataset of colorectal cancer (CRC) (Lee et al., 2020) (NCBI GEO id: GSE132465) was used. This dataset had a total of 33 samples among which 10 adjacent normal (SMC01-N, SMC02-N, SMC03-N, SMC04-N, SMC05-N, SMC06-N, SMC07-N, SMC08-N, SMC09-N, SMC10-N), 2 stage I (SMC07-T, SMC24-T), 5 stage II (SMC01-T, SMC05-T, SMC09-T, SMC15-T, SMC18-T), 14 stage III (SMC02-T, SMC03-T, SMC04-T, SMC06-T, SMC08-T, SMC10-T, SMC11-T, SMC14-T, SMC16-T, SMC17-T, SMC19-T, SMC20-T, SMC22-T, SMC23-T) and 2 stage IV (SMC21-T, SMC25-T) samples were present (Table S1). One sample (SMC02-N) failed sub-sequent processing by our pipeline and was removed from our analysis. For this study, Epithelial, Myeloid, T and stromal cells were used as they were present in substantial numbers across all patient samples (Fig. S12B).

Identification of functional archetypes by Pareto front analysis

The TPM normalized UMI count dataset of 26 lung adenocarcinoma (LUAD) patient samples were extracted. Pareto front analysis was performed to identify the functional archetypes using the ParTI framework (Hart et al., 2015). The single-cell RNA sequencing data of each sample was pre-processed as follows. Specifically, the TPM data of every sample was centred**.** This transformation represented the difference in the expression of a given gene in a cell compared to the average expression of that gene across all cells in a sample. For each sample, the centred data matrix was transposed and passed to ParTI, along with gene names and cell type information, for archetype detection. PCHA (Principal Convex Hull Algorithm) (Mørup & Hansen, 2012) was used by ParTI to identify archetypes for the best-fit polytope. This algorithm finds the best-fit polytope whose vertices or archetypes are located on the convex hull of the data. Further, this algorithm constrains the vertices or archetypes as weighted or convex combinations of data points and the datapoints are then approximated as weighted average of archetypes. Along with the constraints imposed, this algorithm also allows some relaxation so that the archetypes can be located within a certain volume around the convex hull (Korem *et al.*, 2015). Further, by comparing the convex hull to principal convex hull, ParTI determines how well this low dimensional shape fits the data and estimates the statistical significance of the low dimensional shape. Thus, it is important to identify the optimum number of archetypes that fits the data well. This was done prior to the final run of ParTI on each patient sample dataset.

To determine the number of archetypes that can fit the data well, the cumulative variance explained (%EV) by the polytope was checked for different numbers of archetypes K, where K varied between 2 and 8. The elbow curve between %EV vs K for 2-8 archetypes**,** and the cumulative variance explained up to 8 principal components for every sample was checked and was used to determine the optimum number of archetypes K^* such that the variance explained by the additional archetypes was minimal. The significance or p-value for polytope with optimum number of archetypes for every sample was checked. Apart from two patient samples LUNG_T08 (P-value = 0.283) and EBUS_28 (P-value = 0.586) in lung adenocarcinoma (LUAD) dataset (Kim et al., 2020) and one sample SMC06-N (P-value = 0.416) in colorectal carcinoma (CRC) dataset (Lee et al., 2020), all samples showed polytope with strong statistical significance. However, the issue with these samples was addressed by a subsequent cell-state density estimation which was integrated in our pipeline for a more robust framework.

For ParTI analysis, the bin size was adjusted according to the total number of cells present in different patient samples in such a way that it lied between 50 to 100. This was done prior to the final run of 'ParTI' function by changing the 'binsize' parameter value between 0-1. Here, the 'binsize' value indicated the fraction of datapoints in a single bin which was required to calculate feature enrichment.

For further analysis, information regarding principal components, archetype coordinates and cell coordinates were collected from the ParTI results.

Identification of specialist and generalist populations using distance distribution

In the next step, each cell was assigned an archetype (or considered a generalist) from a Euclidean distance calculation and filtering. Briefly, Euclidean distance between each cell and each archetype was calculated and the distance distribution of all cells for each archetype were obtained. An earlier study (Nath et al., 2021) had considered cells with archetype score – defined by inverse of Euclidean distance - above 80 percentile as specialists whereas the rest were considered as generalists. Groves et al. (2022) used archetypal weight as a score (weighted average of archetypes measured by ParTI in such a way that the weights for each cell sum up to 1) for individual cells. Cells with a score greater than 0.95 were considered specialists for a single archetype. But taking an arbitrary cut-off like this may not capture true specialists or generalists properly. In the one hand, this cut-off may miss out some crucial specialist cells and on the other hand, this can also include some generalists within specialist subpopulations. Thus, to circumvent this issue, specialist and generalist cell populations were segregated using Euclidean distance distribution as follows.

Euclidean distance of all cells from each archetype was calculated and the overall distance distribution for each archetype was obtained. Subsequently, distance distributions were fitted with multimodal normal distribution. The “normalmixEM” function of “mixtools” package of R (Benaglia et al., 2009) was used to estimate the value of parameters used as initial guesses for automatically fitting distributions with the mixture models of normal distribution. Next, the function “curve_fit” of “scipy.optimize” library in python was used to fit multimodal distributions. A minimum of one mode and a maximum of four modes were observed (In CRC testing dataset, maximum 6 modes were observed in certain archetypes). Among all models starting from unimodal distribution to multimodal distribution, the model which showed the lowest AIC score was considered as the best fit. After model fitting, the specialists belonging to the archetypes and the generalists were identified as follows. For unimodal distributions, first mode – 1×s.d. (standard deviation) and for multimodal distributions, first mode + 1×s.d. were used as cut-off distance for separating specialists from generalists. Specialists that were distributed between two or more archetypes after distance threshold filtering, were not considered as specialists. After cell removal and specialist cell assignments in a sample, remaining cells were considered as generalist cells (Fig. S13). Next, hierarchical clustering was performed across all specialists and generalist in the sample to check

for similarity of expression pattern within specialist cells and distinct expression pattern among different specialists and generalist cell populations.

Re-clustering of samples with mixed expression patterns within archetypes

For some archetypes within some samples, the distance-based cut-off method failed to effectively capture the diverse specialist cell populations. Heatmaps of gene expression patterns of specialists revealed cases where one archetype harboured multiple distinct expression patterns as well as cases where specialists from one or more archetype(s) showed similar expression pattern as the generalists. Thus, for a more robust identification of specialist cell populations from specialists exhibiting mixed expression programs, a UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2018) analysis with hierarchical density-based clustering (HDBSCAN) (McInnes et al., 2017) was performed.

For the archetypes that showed multiple distinct gene expression patterns but did not resemble generalists, re-clustering was performed to identify multiple specialist sub-populations (Fig. S14). Samples that had sub-populations with similar expression pattern as the generalists, re-clustering was done to test whether these sub-populations clustered with generalists. For re-clustering, the TPM count data of the specialist populations under consideration and the generalist populations were first dimensionally reduced by UMAP followed by clustering by HDBSCAN. The “umap.UMAP” function of “umap” library of python was used to perform UMAP operation with the parameter values “n_neighbors” = 30, “min_dist” = 0 and “random_state=42”. Followed by umap operation, “hdbscan.HDBSCAN” function of “hdbscan” library was used for HDBSCAN operation and “min_cluster_size” was set at 3% of the total cell population. Cells that were not part of any clusters (labeled as “-1”) were removed. The goodness of clustering was measured by calculating the “silhouette score” and a score of 0.4 or more was considered as good clustering. A final decision on cluster assignments and specialist cell identifications was verified by a final visual inspection of the expression patterns across specialists and generalist. Cases where specialist subpopulations and generalists were included in the re-clustering process and the silhouette score was less than 0.4, the specialist subpopulations were merged with the generalist cells and were considered as the total generalist population. However, in cases where re-clustering showed a silhouette score of more than 0.4, cluster that contained the highest percentage of generalist cells was considered as generalist population.

Identification of stable functional states and comparison with pareto front analysis

Although specialists and generalist cell populations were identified from distance-based distribution followed by re-clustering and showed distinct gene expression patterns, it was important to test whether these states represented stable functional states. Within the tumor microenvironment, some cells could be stable states, whereas some cells could be in transitory or transient states. A cell-density based analysis framework such as 'Mellon' (Otto et al., 2024) can effectively identify the stable and the transitory states. Mellon estimates cell state density from single-cell transcriptomic data using distance to its nearest neighbour for each cell in diffusion space and converts the distance to physical density using Poisson probability density function followed by maximum likelihood estimation. In this framework, high-density cell states represent stable functional states whereas low-density cell states represent transitory states. Such analysis could also enable a more robust identification of specialist and generalist cells for samples where we observed polytope with low significance in Pareto front analysis (LUNG_T08 & EBUS_28 in LUAD and SMC06-N in CRC).

In Mellon analysis, diffusion map and density estimation were performed. For the diffusion map, n_components **=** 20 was used. The 'mellon.DensityEstimator' function was used followed by the diffusion map. The 'DensityEstimator' class of Mellon was used to estimate the density of each cell, and UMAP was applied to visualize the density landscape on low-dimensional manifolds. After estimating the density for each cell, the distribution of density was checked in every sample. For unimodal density distributions, mean - 3 S.D. (standard deviation) was taken as cut-off, and for multimodal density distributions, mean - 1 S.D. from the first peak was considered as cut-off to separate high- and low-density regions. Subsequently, high-density cells were mapped with specialist and generalist populations obtained from Pareto front analysis and percentage of mapped cells between two frameworks were also estimated for all archetypes. Next, hierarchical clustering was performed again across all specialists and generalist population for every sample to validate the purity and reliability of obtained specialists and generalist cell subpopulation in the tumor microenvironment. The specialist and generalist populations obtained from Mellon were considered the final specialist and generalist subpopulations in the tumor microenvironment and these were used for all subsequent analyses.

Functional enrichment of meta-programs in archetypes

To define the functional identity of specialist cells, functional enrichment of meta-programs - as defined in the MSigDB 3CA (Curated Cancer Atlas) database (Gavish et al., 2023) - in cells of each archetype was performed. For functional enrichment analysis, an approach similar to Gavish et al., 2023 was

followed. For an archetype, the average expression of genes belonging to each metaprogram in each cell belonging to that archetype was estimated. This information was transformed into a matrix where each row represented a cell, each column represented a metaprogram, and the values represented the average expression of the metaprogram in individual cells. The same matrix was created for rest of the population in the sample. In the next step, for every metaprogram, its average expression in the archetype of interest was compared with its expression in the rest of the population using a Mann-Whitney U test. This analysis was performed for all metaprograms and the estimated p-values were corrected using Benjamini-Hochberg method and metaprograms with p-adjusted<0.1 were considered to be enriched. A positive $\log_2$ fold change in expression of the metaprogram in an archetype suggested enrichment of the metaprogram in the archetype, whereas a negative $\log_2$ fold change indicated enrichment of the metaprogram in the rest of the population. The enriched metaprograms within an archetype were ranked according to their p-adjusted values.

Identification of transitory cell states

In Pareto front analysis, many cells across samples fell between two archetypes and did not belong to any of the archetypes. Further, some of these cells also fell in low density regions in the Mellon analysis and were not considered as part of specialist or generalist populations. These cells, although small in number, represent phenotypically plastic states which could have important implications for cancer.

To identify the transitory cells in the pareto front, cells that were not included in any specialist or generalist populations were analysed. To identify the archetypes between which these transitory cells were located, the weightage of all archetypes within the sample on those cells were calculated. The weightage of an archetype for a cell was taken as the inverse of the Euclidean distance of the cell to that archetype. The weights for a cell were normalized to have the sum of the weights as 1. Two archetypes with the top most weights on a cell were identified and they defined the archetypes between which the transitory state lied.

As transitory cells were located in the low-density regions that were connected with two high density regions of two different archetypes (Fig. S15A,B), these cells can drive functional transition from one archetype to another. To understand the functional transitions from one archetype to another, archetypes between which significant numbers of transitory cells (at least 20 cells) present were chosen. ‘Gene change analysis’ of Mellon was performed to do this so that genes driving the transition could be identified. Prior to performing ‘gene change analysis’, a pseudotime analysis using ‘Palantir’ was done (Setty et al., 2019) (Fig. S15B). To run ‘Palantir’, the expression matrix of two different specialist cell

populations and the transitory cells in between these specialists were arranged sequentially and diffusion map was run on the expression matrix using 'palantir.utils.run_diffusion_maps' and 'n_components = 5'. Next, 'palantir.utils.determine_multiscale_space' was used to calculate the multiscale representation of cell states using diffusion maps. Next, to identify the starting cell in an unbiased way for drawing the trajectory, cells with highest value along the first diffusion components were chosen. Next, 'Palantir' core pipeline was run using 'palantir.core.run_palantir' and 'num_waypoints=500' was used. Palantir fate probabilities were checked using 'palantir.presults.select_branch_cells' along different branches and visualized (Fig. S15C). For this, threshold quantile value (q) of 0.01 was used for 'fate probability' and eps = 0.01 was kept.

Genes that showed high change in gene expression in the low-density region in between two high density regions were likely to be the drivers of functional transition along the trajectories. To identify those genes, first local gene variability scores were calculated using 'palantir.utils.run_local_variability'. Next, 'palantir.utils.run_low_density_variability' function was run followed by 'palantir.utils.run_local_variability' which aggregated the local gene expression variability within the low cell density region. This calculated the 'gene change score' of each gene along different branch of trajectories and ranked them based on their score. Genes that showed score more than 95 percentiles were consider as the driving genes for the functional transition from one archetype to another (Fig. S15D).

Machine learning models for classification of transitory and specialist myeloid cells

The transitory cells between myeloid dominant and T/NK dominant (or T dominant) archetypes were abundant in tumor samples in both LUAD and CRC, and these cells were predominantly composed of transitory myeloid cells. Thus, to test whether the transitory myeloid cells had distinct identity compared to specialist myeloid cells and to identify the key molecular features that distinguished them, machine learning models for classifying these two types of myeloid cells (classes) from their gene expression patterns were developed individually for LUAD and CRC. Single-cell gene expression data of all transitory myeloid cells and specialist myeloid cells from the samples with presence of transitory myeloid cells were collated. The data was cleaned to remove genes showing constant values across all cells. Since the specialist myeloid cells greatly outnumbered the transitory myeloid cells, this led to the problem of data imbalance in both LUAD and CRC. Thus, to minimize the effect of data imbalance on the performance of machine learning models, the majority class (specialist myeloid cells) was sampled to include the same number of cells as the transitory myeloid cells for each cancer type.

The sampled data were then divided into training (75%) and test (25%) datasets. The training data was used for hyperparameter tuning through GridSearch process with 5-fold cross-validation. Thus, the training data was divided into training and validation sets for the GridSearch. The training set was scaled and normalized, and subsequently, was used for feature selection using a three-step method. The first step used SelectKBest approach and selected the top 200 features (genes). In the second step, a logistic regression-based feature selection was used which identified the features with significant impact on prediction. In the final step, a recursive feature elimination with cross-validation (rfecv) based on logistic regression was used for selection of final set of features. These features were then used to build classification models using Logistic regression, XGBoost (Chen and Guestrin, 2016), Random Forest (Breiman, 2001) and LigtGBM (Ke et al., 2017) methods. The best set of hyperparameters were chosen by evaluating performance of the model on the validation set. The final accuracy for each model was reported for prediction on the test datasets. Since classification was done on sampled data, the process was repeated five times to ensure that the results obtained were not biased by the sampling step.

The key features (genes) that helped in classification were identified by feature importance ranking using permutation importance approach as well as Shapley additive explanation (SHAP) (Lundberg and Lee, 2017). The features with the highest SHAP scores were further tested for their expression pattern differences between transitory and specialist myeloid cells. The top 50 features were further used for functional enrichment analysis using the gProfiler webserver (Kolberg et al., 2023) with Gene Ontology (GO) terms for Biological Processes (BP) and Molecular Functions (MF) (Ashburner et al., 2000; The GO consortium, 2025) along with the KEGG pathways (Kanehisa et al., 2017).

Gene regulatory network reconstruction and identification of active transcription factors

The contribution of gene regulation programs and specific TFs towards generation of transitory myeloid states and functional plasticity were analysed through gene regulatory network reconstruction following an approach described in Ray et al. (2026) in LUAD samples.

To identify the gene regulatory programs and the transcription factors (TFs) driving generation of transitory myeloid states and functional plasticity, the active gene regulatory network was reconstructed in each cell type within individual LUAD samples. In the first step of gene regulatory network

reconstruction, for each cell type within a sample, a pair-wise expression correlation between all genes were calculated. As expression correlation can also arise due to indirect regulatory connections, a high-confidence TF-target map was used to filter out the correlation values. The TF-target map was constructed using published experimentally validated TF-target relations combined with a second-step screening for binding motifs of TFs on the target promoters. The TF-target correlation distribution for each cell type within each sample was further filtered using mean ± 2×s.d. cut-off values to identify only the active connections.

In the next step, the activity of the identified TF-target connections in the subsets of cells of specific cell types were evaluated. For identifying the active regulons and the key TFs in transitory and specialist myeloid cells, the activity of regulons and TFs in these subpopulations were evaluated. For identifying the active regulons and the key driver TFs of intra-sample/inter-sample functional plasticity, the activity of the regulons was evaluated in the subpopulations of cells showing functional convergence.

The regulon for each TF was constructed by considering all targets that were present in the TF-target map, and the activity of the regulon was evaluated in each cell of the population or subpopulation using the tool AUCell (Aibar et al., 2017). The TFs were then ranked according the AUCell score in the cell subpopulation, representing the most active TFs and regulon within the subset of cells.

Differential expression analysis of T/NK or T cells between samples harboring transitory myeloid states (WithTransMyl) and samples without transitory myeloid states (NoTransMyl)

To investigate the influence of the transitory myeloid cells on the tumor immune microenvironment, a two-pronged approach was taken. In the first approach, machine learning models were built to classify T/NK (or T) cells coming from samples with transitory myeloid states (WithTransMyl) versus the T/NK (or T) cells coming from samples without the transitory states (NoTransMyl). The T/NK (or T) cells belonging to these two different classes were collated individually for the two cancer types (LUAD and CRC), and machine learning models were built following the same steps as described for myeloid cell classification. Briefly, the datasets were divided into training (75%) and test (25%) datasets. The training data was used for hyperparameter training through 5-fold cross validation. All scaling, normalization, feature selection and model fitting (Logistic Regression, XGBoost, Random Forest, and LightGBM) were done on the training set and the best set of hyperparameters were chosen by evaluating performance of the model on the validation set. The final accuracy of each model was reported for prediction on the test datasets. Since classification was done on sampled data, the process was repeated five times to ensure that the results obtained were not biased by the sampling process.

In the second approach, differential expression analysis was performed between T/NK (or T) cells from WithTransMyl samples and the T/NK (or T) cells from NoTransMyl samples to identify the differences in biological processes operating in these two classes of cells. The R package MAST (Finak et al., 2015) was used to identify the differentially expressed genes. Briefly, MAST used a two-step generalized linear model to account for both the fraction of cells that express a gene and the change in expression levels of cells with non-zero expression. The genes with abs(coef) > 0.2 and p-adjusted <0.01 were considered as differentially expressed. The up-regulated (with coef > 0.2) and the down-regulated genes (coef < -0.2) were checked for functional enrichment using gProfiler considering the Gene Ontology (GO) BP and MF terms and KEGG pathways.

Archetype alignment analysis

Analysis on archetype alignment can provide a global view of the functional diversity across patients within a cancer type and can identify the cases of intra-tumor and inter-tumor functional plasticity where different cell types can converge to similar functions. Thus, an alignment of all archetypes, based on their functional identity, across all patients within a cancer type was performed using the following steps. First, the functions that were significantly enriched at each archetype were collated for all patient samples within a cancer type. Next, a pairwise archetype similarity score was calculated between the pairs of archetypes where the values indicated the percentage of functions that were ranked similarly in both these archetypes. Functional enrichment analysis based on meta-programs (Gavish et al., 2023) were used for identifying enriched functions and ranking them. Compilation of all pairwise similarity score led to an archetype similarity matrix across the archetypes. The archetypes without any significant enrichment of functions were removed from the similarity matrix.

Subsequently, a hierarchical clustering was performed on the similarity matrix where both rows and columns were clustered. This identified clusters of archetypes with distinct functional patterns representing functional convergence. Cell types representing those respective archetypes were also identified. Functional clusters where the heatmap showed more than one distinct pattern of small clusters, re-clustering were performed. Prior to perform hierarchical clustering again, row and columns linkage of dendrograms were identified and this information was used to identify the number of small clusters with distinct functional patterns within the bigger functional clusters. To identify the optimum number of smaller clusters within the big clusters, ‘fcluster’ package of ‘scipy.cluster.hierarchy’ was used

and silhouette score were calculated between the range of 2-20 clusters. Number of clusters for which the silhouette score was highest was considered as the optimum number of clusters for hierarchical clustering. The archetypes where multiple cell types were part of the archetypes were considered for further analysis of intra-sample functional plasticity. The archetypes across different samples where different cell types across these samples showed functional convergence were considered for further analysis of inter-sample functional plasticity.

References

Aibar, S. et al. (2017). "SCENIC: Single-Cell Regulatory Network Inference And Clustering." *Nature Methods*, **14**, 1083-1086. doi:10.1038/nmeth.4463.

Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current biology: CB*, *30*(16), R921–R925. https://doi.org/10.1016/j.cub.2020.06.081

Ashburner et al. (2000). Gene ontology: tool for the unification of biology. *Nat Genet.* 25(1):25-9. doi: 10.1038/75556

Bala, P. et al. (2023). Aberrant cell state plasticity mediated by developmental reprogramming precedes colorectal cancer initiation. *Science advances*, *9*(13), eadf0927. https://doi.org/10.1126/sciadv.adf0927

Barkley, D. et al. (2022). Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nature genetics*, *54*(8), 1192–1201. https://doi.org/10.1038/s41588-022-01141-9

Benaglia, T., Chauveau, D., Hunter, D. R., & Young, D. S. (2009). mixtools: An R Package for Analyzing Mixture Models. *Journal of Statistical Software*, *32*(6), 1–29. https://doi.org/10.18637/jss.v032.i06

Breiman, L (2001). Random Forests. *Machine Learning* 45, 5–32. https://doi.org/10.1023/A:1010933404324

Bronte, V. et al. (2016). Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards. *Nat Commun*. 7:12150. doi: 10.1038/ncomms12150.

Buissant des Amorie, J. R. et al. (2026). Emergence of oncofetal plasticity is ubiquitous in early colorectal cancers. *Nature*, *654*(8117), 229–239. https://doi.org/10.1038/s41586-026-10344-7

Chen T, Guestrin C, 2016. XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Pages 785 – 794. https://doi.org/10.1145/2939672.2939785

de Visser, K. E., & Joyce, J. A. (2023). The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer cell*, *41*(3), 374–403. https://doi.org/10.1016/j.ccell.2023.02.016

Emert, B.L. et al. (2021) Variability within rare cell states enables multiple paths toward drug resistance. *Nat Biotechnol.* 39(7):865-876. doi: 10.1038/s41587-021-00837-3.

Finak, G. et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16:278. doi: 10.1186/s13059-015-0844-5.

Gavish, A. et al. (2023). Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature*, *618*(7965), 598–606. https://doi.org/10.1038/s41586-023-06130-4

Groves, S. M. et al. (2022). Archetype tasks link intratumoral heterogeneity to plasticity and cancer hallmarks in small cell lung cancer. *Cell systems*, *13*(9), 690–710.e17. https://doi.org/10.1016/j.cels.2022.07.006

Guan Y, Zhu X, Liang J, Wei M, Huang S, Pan X (2021). Upregulation of HSPA1A/HSPA1B/HSPA7 and Downregulation of HSPA9 Were Related to Poor Survival in Colon Cancer. *Front Oncol.* 11:749673. doi: 10.3389/fonc.2021.749673.

Han, G. et al. (2024). An atlas of epithelial cell states and plasticity in lung adenocarcinoma. *Nature* 627(8004):656-663. doi: 10.1038/s41586-024-07113-9.

Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29 (2021).

Hart, Y. et al. (2015). Inferring biological tasks using Pareto analysis of high-dimensional data. *Nature Methods*, *12*(3), 233–235. https://doi.org/10.1038/nmeth.3254

Hausser, J., Szekely, P., Bar, N. *et al.* (2019). Tumor diversity and the trade-off between universal cancer tasks. *Nat Commun* 10, 5423. https://doi.org/10.1038/s41467-019-13195-1

Huang LP, Lyu SC, Clayberger C, Krensky AM, (2007). Granulysin-mediated tumor rejection in transgenic mice. J Immunol. 178(1):77-84. doi: 10.4049/jimmunol.178.1.77.

Huang, D. et al. (2021). Targeting regulator of G protein signaling 1 in tumor-specific T cells enhances their trafficking to breast cancer. *Nat Immunol.* 22(7):865-879. doi: 10.1038/s41590-021-00939-9.

Jonkman, I. et al. (2025). C1q reprograms innate immune memory. Front Immunol. 2025 May 23;16:1515127. doi: 10.3389/fimmu.2025.1515127.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45(D1): D353-D361. doi: 10.1093/nar/gkw1092.

Ke, G. et al. (2017). LightGBM: a highly efficient gradient boosting decision tree. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Pages 3149 - 3157

Kim, N. et al. (2020). Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16164-1

Kinker, G. S. et al. (2020). Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nature genetics*, *52*(11), 1208–1218. https://doi.org/10.1038/s41588-020-00726-6

Klemm, F., & Joyce, J. A. (2015). Microenvironmental regulation of therapeutic response in cancer. *Trends in cell biology*, *25*(4), 198–213. https://doi.org/10.1016/j.tcb.2014.11.006

Kolberg, L. et al. (2023) g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update) *Nucleic Acids Research*, doi:10.1093/nar/gkad347

Korem, Y et al. (2015). Geometry of the Gene Expression Space of Individual Cells. *PLoS Computational Biology*, 11(7), 1–27. https://doi.org/10.1371/journal.pcbi.1004224

Lee, H.-O. et al. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nature Genetics, 52(6), 594–603. https://doi.org/10.1038/s41588-020-0636-z

Li, C. et al. SciBet as a portable and fast single cell type identifier. *Nat. Commun.* **11**, 1818 (2020).

Lundberg, S., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874 [cs.AI] https://doi.org/10.48550/arXiv.1705.07874

Marjanovic, N. D. et al. (2020). Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. *Cancer cell*, *38*(2), 229–246.e13. https://doi.org/10.1016/j.ccell.2020.06.012

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. The Journal of Open Source Software, 2(11). https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software, 3(29). https://doi.org/10.21105/joss.00861

Moorman, A. et al. (2025). Progressive plasticity during colorectal cancer metastasis. *Nature*, *637*(8047), 947–954. https://doi.org/10.1038/s41586-024-08150-0

Mørup, M., & Hansen, L. K. (2012). Archetypal analysis for machine learning and data mining. Neurocomputing, 80. https://doi.org/10.1016/j.neucom.2011.06.033

Nath, A. et al. (2021). Evolution of core archetypal phenotypes in progressive high grade serous ovarian cancer. *Nature Comms.* 12(1), 1–16. https://doi.org/10.1038/s41467-021-23171-3

Neftel, C. et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell*, *178*(4), 835–849.e21. https://doi.org/10.1016/j.cell.2019.06.024

Otto, D. J., Jordan, C., Dury, B., Dien, C., & Setty, M. (2024). Quantifying cell-state densities in single-cell phenotypic landscapes using Mellon. *Nature Methods*, 21(7), 1185–1195. https://doi.org/10.1038/s41592-024-02302-w

Pérez-González, A., Bévant, K., & Blanpain, C. (2023). Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nature cancer*, *4*(8), 1063–1082. https://doi.org/10.1038/s43018-023-00595-y

Quail, D. F., & Joyce, J. A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nature medicine*, *19*(11), 1423–1437. https://doi.org/10.1038/nm.3394

Rafei, H. et al. (2025). CREM is a regulatory checkpoint of CAR and IL-15 signalling in NK cells. *Nature* 643(8073):1076-1086. doi: 10.1038/s41586-025-09087-8.

Ray, U., Singh, A., Samanta, D., & Dhar, R. (2026). Gene regulatory network transitions reveal the central transcription factors in lung adenocarcinoma progression. Npj Systems Biology and Applications, 12(1), 18. https://doi.org/10.1038/s41540-025-00640-9

Sanseviero, E., Barry, S. T., & Gabrilovich, D. I. (2026). Harnessing myeloid cell plasticity for cancer therapy. *Nature cancer*, 10.1038/s43018-026-01159-6. https://doi.org/10.1038/s43018-026-01159-6

Schaff, D. L. et al. (2026). Pre-existing cell states predict resistance to multiple treatments. *Cell genomics*, *6*(6), 101191. https://doi.org/10.1016/j.xgen.2026.101191

Schmid, M.C. et al. (2018) Integrin CD11b activation drives anti-tumor innate immunity. *Nat Commun.* 9(1):5379. doi: 10.1038/s41467-018-07387-4.

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., & Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnology*, 37(4), 451–460. https://doi.org/10.1038/s41587-019-0068-4

Sun, Y. et al. (2025). Spatial transcriptomics reveals macrophage domestication by epithelial cells promotes immunotherapy resistance in small cell lung cancer. *NPJ precision oncology*, *9*(1), 252. https://doi.org/10.1038/s41698-025-01005-5

Sun, Z. et al. (2023) Cuproptosis and Immune-Related Gene Signature Predicts Immunotherapy Response and Prognosis in Lung Adenocarcinoma. *Life* (Basel). 13(7):1583. doi: 10.3390/life13071583.

Tewary P, Yang D, de la Rosa G et al. (2010). Granulysin activates antigen-presenting cells through TLR4 and acts as an immune alarmin. Blood 116(18):3465-74. doi: 10.1182/blood-2010-03-273953.

The Gene Ontology Consortium (2025). The Gene Ontology knowledgebase in 2026. *Nucleic Acids Res.* gkaf1292. doi: 10.1093/nar/gkaf1292

van Montfoort, N. et al. (2007)/ A novel role of complement factor C1q in augmenting the presentation of antigen captured in immune complexes to CD8+ T lymphocytes. J Immunol.178(12):7581-6. doi: 10.4049/jimmunol.178.12.7581.

Vardhana, S.A. et al. (2020). Impaired mitochondrial oxidative phosphorylation limits the self-renewal of T cells exposed to persistent antigen. *Nat Immunol*. 21(9):1022-1033. doi: 10.1038/s41590-020-0725-2.

Yang, J. et al. (2018). Loss of CXCR4 in Myeloid Cells Enhances Antitumor Immunity and Reduces Melanoma Growth through NK Cell and FASL Mechanisms. *Cancer Immunol Res*. 6(10):1186-1198. doi: 10.1158/2326-6066.CIR-18-0045.

Yuan, S., Norgard, R. J., & Stanger, B. Z. (2019). Cellular Plasticity in Cancer. *Cancer discovery*, *9*(7), 837–851. https://doi.org/10.1158/2159-8290.CD-19-0015

Wang L, Dong Q, He F, Gu Z, Sun A (2026). Beyond bacterial defences: the role of lysozyme in cancer. *Clin Transl Med.* 2026 Feb;16(2):e70575. doi: 10.1002/ctm2.70575.

Wang, Y. et al. (2021). NAD^+ supplement potentiates tumor-killing function by rescuing defective TUB-mediated NAMPT transcription in tumor-infiltrated T cells. *Cell Rep.* 36(6):109516. doi: 10.1016/j.celrep.2021.109516.

Acknowledgement

The authors are grateful to Prof. Ranjit Prasad Bahadur, Department of Bioscience and Biotechnology, IIT Kharagpur for his critical feedback on the work. Work in the RD lab was supported by funding from Department of Biotechnology (DBT), Ministry of Science and Technology grants BT/PR40175/BTIS/137/41/2022 and BT/PR40182/BTIS/137/88/2025; and Ignite Life Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Conceptualization: RD
Methodology: SD, RD
Investigation: SD, RM, SB, RD
Visualization: SD, RD
Funding acquisition: RD
Project administration: RD
Supervision: RD
Writing – original draft: SD, RM, RD
Writing – review & editing: SD, RM, SB, RD

Competing interests

The authors declare that no competing interests exist.

Data availability

All data used in this study are publicly available. The lung adenocarcinoma (LUAD) dataset is available in the NCBI GEO accession number GSE131907. The colorectal cancer (CRC) dataset is available in the NCBI GEO id: GSE132465.

Code availability

All codes for data analysis and models are available in GitHub:
https://github.com/sbliitkgp/Cancer_functional_transition_and_plasticity

Supplementary material

Supplementary tables

Table S1 – Patient sample information for LUAD and CRC

Table S2 – Top 50 feature genes identified by the best performing machine learning models in LUAD and CRC for classification of transitory and specialist myeloid cells.

Table S3 – List of up- and down- regulated genes in T/NK cells from samples harbouring transitory myeloid states compared to T/NK cells in samples without transitory myeloid states in LUAD

Table S4 - List of up- and down- regulated genes in T/NK cells from samples harbouring transitory myeloid states compared to T/NK cells in samples without transitory myeloid states in CRC

Supplementary figures

Figure S1 - Overlap between specialists and generalists identified in Pareto front analysis and high-density regions in cell-density based stability analysis.

Figure S2 - Functional diversity in each cell type in individual patient samples in CRC.

Figure S3 – Changes in top-ranked functional programs with cancer progression in LUAD.

Figure S4 – Changes in top-ranked functional programs with cancer progression in CRC.

Figure S5 – Changes in functional utilization in epithelial and T cells in CRC samples with cancer progression.

Figure S6 – Percentage of transitory myeloid cells out of total myeloid population in samples that harboured transitory myeloid-T/NK (or T) functional states.

Figure S7 – Expression of ITGAM (CD11b), CD14, CD33 and LYZ genes in adjacent normal and tumor samples in LUAD.

Figure S8 – Expression of ITGAM (CD11b), CD14, CD33 and LYZ genes in adjacent normal and tumor samples in CRC.

Figure S9 – Test for expression of NSCLC-specific MDSC markers in transitory myeloid cells in LUAD samples.

Figure S10 - Test for expression of colon cancer-specific MDSC markers in transitory and specialist myeloid cells in CRC samples.

Figure S11 – Different cell types converge to similar functional profiles in CRC samples.

Figure S12 - Percentage of cell types for patient samples in LUAD and CRC.

Figure S13 – Sample-wise percentage of functional specialist, generalist and transitory cells in LUAD and CRC samples.

Figure S14 - An example of refinement of functional specialists through re-clustering.

Figure S15 - Cell state transition trajectory analysis and gene expression change along the trajectory

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigsv6.pdf
- Supplementaryfigsv6.pdf
- Supplementaryfigsv6.pdf
- Supplementaryfigsv6.pdf
- Supplementaryfigsv6.pdf
- Supplementaryfigsv6.pdf
- Supplementaryfigsv6.pdf
- TableS1SampleMetadata.xlsx
- TableS1SampleMetadata.xlsx
- TableS1SampleMetadata.xlsx
- TableS1SampleMetadata.xlsx
- TableS1SampleMetadata.xlsx
- TableS1SampleMetadata.xlsx
- TableS1SampleMetadata.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS2MLmodelTop50featureGenes.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS3DiffExprTNKLUAD.xlsx
- TableS4DiffExprTCRC.xlsx
- TableS4DiffExprTCRC.xlsx
- TableS4DiffExprTCRC.xlsx

- TableS4DiffExprTCRC.xlsx
- TableS4DiffExprTCRC.xlsx
- TableS4DiffExprTCRC.xlsx
- TableS4DiffExprTCRC.xlsx