Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

Cancer Cell

CellPress

Perspective

Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors

Itay Tirosh[1,*] and Mario L. Suva[2,*]
[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 761001, Israel
[2]Department of Pathology and Krantz Family Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
*Correspondence: itay.tirosh@weizmann.ac.il (I.T.), suva.mario@mgh.harvard.edu (M.L.S.)
https://doi.org/10.1016/j.ccell.2024.08.005

SUMMARY

Human tumors are intricate ecosystems composed of diverse genetic clones and malignant cell states that evolve in a complex tumor micro-environment. Single-cell RNA-sequencing (scRNA-seq) provides a compelling strategy to dissect this intricate biology and has enabled a revolution in our ability to understand tumor biology over the last ten years. Here we reflect on this first decade of scRNA-seq in human tumors and highlight some of the powerful insights gleaned from these studies. We first focus on computational approaches for robustly defining cancer cell states and their diversity and highlight some of the most common patterns of gene expression *intra*-tumor heterogeneity (eITH) observed across cancer types. We then discuss ambiguities in the field in defining and naming such eITH programs. Finally, we highlight critical developments that will facilitate future research and the broader implementation of these technologies in clinical settings.

INTRODUCTION

Single-cell RNA-sequencing (scRNA-seq) has revolutionized our understanding of complex biological systems. In cancer, a combination of intrinsic genetic and epigenetic variation with extrinsic and spatial factors creates a paragon of biological complexity. Diverse genetic clones of malignant cells exist in a complex tumor microenvironment (TME) of immune, stromal and vascular cells, where cell-cell interactions and nutrient and oxygen availability affect tumor growth, differentiation, invasion, metastasis and response to therapy. scRNA-seq technologies identify programs that represent an integration of all these layers of variation, defining a fundamental aspect of cancer biology.

The initial studies of human cancer specimens by scRNA-seq were performed by specialized laboratories a decade ago.[1–5] These early studies profiled single cells that had been isolated by flow cytometry into wells or in distinct chambers by microfluidic devices. While these studies provided detailed insights into expression programs of cancer and immune cells, they were limited by scalability, typically to a few dozens or hundreds of cells per sample. The development of high-throughput methods such as microdroplets offered impressive scalability at reduced costs (Macosko et al.[6] and Klein et al[7]; also commercialized later by 10× Genomics) by enabling the 3′-end or 5′-end sequencing of thousands of cells in parallel.

Microdroplet-based systems have been established as the most widely used platforms for high-throughput scRNA-seq. Additional developments in scale were enabled by combinatorial barcoding (SPLiT-seq),[8] bypassing the need for separation by leveraging the cells themselves as compartments, further reducing costs. Nanowells-based systems offer interesting portability and cell imaging possibilities.[9] While flow cytometry-based sorting remains limited in scale compared to droplet microfluidics, it remains the method of choice for cell selection, multi-omic profiling (e.g., RNA + DNA), and full-length sequencing, although long reads may offer full-length sequencing from droplets, directly from cDNA. Overall, these developments made high-throughput scRNA-seq accessible for most labs and led to a gradual increase in the number of scRNA-seq cancer studies that peaked around 2020 and remains high ever since, generating a vast amount of data and literature on tumor heterogeneity at the transcriptional level.

Given the very rapid expansion of the field and the wide adoption of these technologies in many laboratories across the world, here we reflect on some of the lessons learned in the past ten years, current challenges for the field and future developments for clinical application. We focus primarily on cancer cell states and their diversity within tumors, as observed by scRNA-seq.

DISTINGUISHING *INTER*-TUMOR VS. *INTRA*-TUMOR EXPRESSION HETEROGENEITY

Single-cell expression profiling of a cohort of tumors can be used to define both *inter*-tumor and *intra*-tumor heterogeneity for each of the observed cell types. *Inter*-tumor expression heterogeneity has been extensively studied previously by bulk profiling—for example through The Cancer Genome Atlas (TCGA) datasets—although this approach has limitations associated with tumor purity and TME composition. The study of *intra*-tumor heterogeneity for cancer cell states requires single cell data and therefore has expanded dramatically in the last decade. Nevertheless, when it comes to the malignant cell expression programs (but not typically for immune and stromal cells), *inter*-tumor heterogeneity tends to be the dominant component in unbiased analyses. For example, expression clustering of all malignant cells from a cohort of tumors would

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

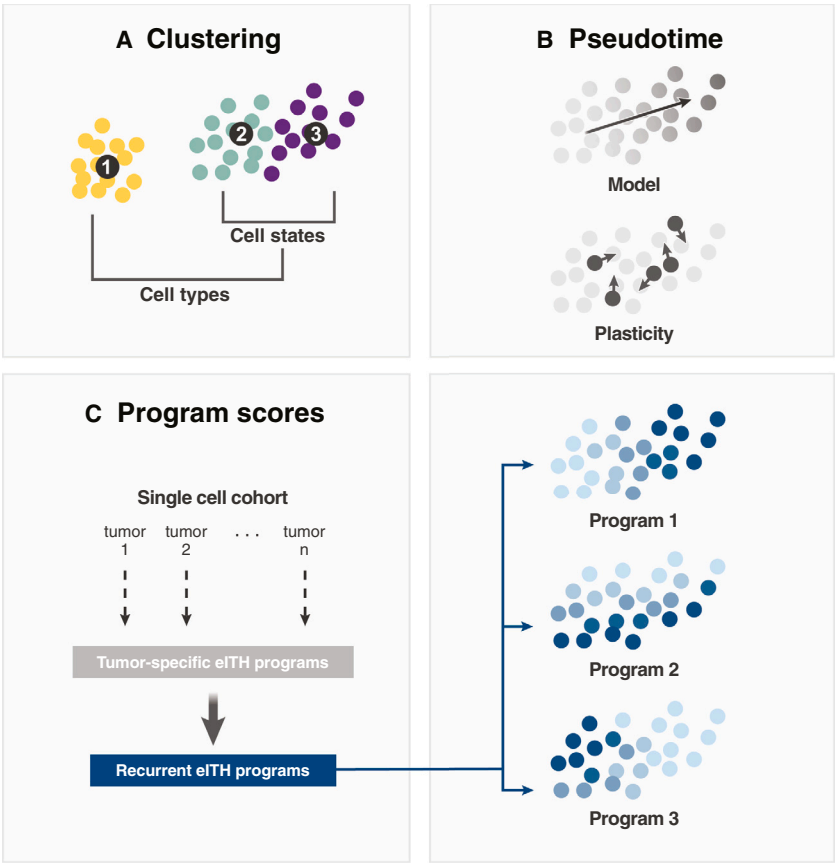

Figure 1. Approaches for eITH analysis

(A) Cell clustering, showing a toy example with three identified clusters as marked by colors and cluster-numbers in black. One of the clusters (#1) is highly distinct from the other cells, reflecting a distinct cell type. The other two clusters (#2 and #3) reflect a continuum of cellular states that is broken into clusters in a manner that is method-dependent. The cells of clusters #2 and #3 are also depicted in the subsequent sections in which their diversity is described by other approaches.

(B) Top panel: pseudotime analysis orders the cells by an inferred trajectory (from white to black as also indicated by an arrow). While such simple trajectories may accurately describe certain processes, cancer cells often demonstrate plasticity such that their state may progress in multiple "directions," including in opposing manners, as depicted in the bottom panel.

(C) Cells may be described by their scores for multiple eITH programs, while prioritizing programs that recur across tumors. Left: identification of recurrent eITH programs. First, eITH programs are identified per tumor, for example by non-negative matrix factorization. Second, the tumor-specific eITH programs are compared to one another to identify the recurrent programs. In the toy example 3 recurrent programs are identified and used to score the cells in the right panel. Right: cells (as shown in A and B) colored by their scores for three recurrent eITH programs that together describe their variation in gene expression.

typically result in separate cluster(s) for each tumor in the cohort.[1,2,10,11] This highlights the very high degree of *inter*-tumor expression heterogeneity, which reflects each tumor's unique genetics along with other potential effects such as the cell-of-origin, spatial location and history of the tumor and various other patient characteristics. Accordingly, direct analysis of cell states from scRNA-seq dataset of a cancer cohort would likely be dominated by tumor-specific patterns and could potentially mask the diversity within tumors.

The large contribution of *inter*-tumor heterogeneity may be decreased by the use of integration methods that are designed to remove technical batch effects.[12,13] However, since each tumor is processed separately, this approach implicitly assumes that observed tumor-specific patterns reflect a batch effect that needs to be removed and hence may distort the data, while eliminating genuine biological signals.[14,15] Integration methods may retain some amount of patient-specific signals while removing others, making it difficult to interpret the remaining differences between tumors. Thus, apart from the two approaches noted previously (direct combination of data from multiple patients or their integration with dedicated methods), a third approach involves separate analysis of each tumor in order to directly define *intra*-tumor heterogeneity without confounding it with *inter*-tumor heterogeneity, thereby considering the two as distinct modes of diversity. In this approach, *intra*-tumor heterogeneity is first defined separately within each tumor and only subsequently, the resulting patterns may be compared across tumors in order to identify recurrent patterns or those that are unique to specific tumors or cancer types.

Given this complex relationship between *inter*- and *intra*-tumor heterogeneity, we focus subsequent sections on *intra*-tumor patterns at the level of gene expression, which we term expression *intra*-tumor heterogeneity (eITH).

DEFINING eITH FROM scRNA-SEQ: CLUSTERS, PSEUDOTIME, AND PROGRAM SCORES

Based on scRNA-seq data, how many distinct subpopulations of cancer cells are found in one tumor sample? This expected and seemingly simple question turns out to be difficult to address. Many studies used clustering methods to report specific numbers of clusters, but the results often appear to depend on the specific methodology applied and to be threshold-dependent. Even defined clusters can usually be subjected to another round of clustering to define meaningful sub-clusters, limiting the relevance of any reported number of clusters. The underlying reason for this complexity is that rather than a well separated set of clusters, malignant cells typically span a continuum of transcriptional programs (Figure 1A). Such continuums reflect a range of *cell states*, as opposed to discrete clusters that may be better described as *cell types*.[16] Continuums may of course be broken into a discrete number of clusters, which simplifies the analysis and provides an (often desired) cluster definition. Yet, the observed continuums suggest that other approaches should also be considered.

Expression continuums raise the possibility that cells may dynamically change their state over time and hence a population

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

of cells may cover a range of observed cellular states. Such dynamic behavior may be unidirectional, such as in lineage differentiation, or may be multi-directional, a concept often referred to as cancer cell plasticity.[17,18] Such plasticity was demonstrated in many contexts, including melanoma,[19,20] glioma,[21,22] pancreatic cancer,[23] and prostate cancer.[24] Accordingly, cancer cell plasticity has gained increasing attention in recent years, including its definition as a hallmark of cancer[25] and as one of cancer's grand challenges (https://cancergrandchallenges.org/). Plasticity implies that even if we successfully target a given state of cancer cells, this state might be regenerated by other cancer cells, calling for new therapeutic strategies based on an improved understanding of the dynamics and determinants of cell states.

With this appreciation of the dynamic behavior of cancer cells, a common alternative to clustering is to order the cancer cells along observed continuums and associate each cell with its position along that ordering (Figure 1B). Such ordering, usually referred to as pseudotime, is based on methods that were initially designed for developmental contexts,[26,27] in which cells indeed progress along a defined path of states, reflecting their differentiation or maturation. Related methods, such as velocity,[28] try to directly infer dynamic cellular trajectories by comparison of recently generated (unspliced) RNA to the pool of mature RNAs.

The pseudotime approach is well suited for continuums but faces two added challenges in the cancer context. First, unlike developmental contexts, it is unclear if cancer cells indeed progress along one direction of a continuum or whether they could advance in multiple directions. Thus, the assumption that cells move from low to high pseudotime should be questioned in a cancer context. Second, while differentiation implies the progression of cells along a defined set of states, in the cancer context cells may simultaneously traverse along multiple axes. For example, cells may increase or decrease their expression of proliferation-related programs and at the same time increase or decrease stress-related programs, immune-related programs and others. By "program," we refer to sets of genes whose expression is correlated among the cancer cells in a given tumor and that are often linked to a particular cellular function or process. An attempt to combine these disparate processes into a single pseudotime may oversimplify the patterns of diversity and mask some of these axes.

Therefore, an alternative to pseudotime involves "scoring" of cancer cells for multiple expression programs[10,21,29] (Figure 1C). Relevant expression programs may first be defined as those that vary recurrently in multiple tumors of a given type, hence reflecting a consistent feature of eITH in that context. Each cell may then be described by a set of program scores that could be subjected to further analysis, including clustering if desired. This approach removes the underlying assumption of directionality and we argue is better suited for the multidimensional dynamics that exist in cancer.

This approach of program scores raises several questions. For example, which programs vary within a given tumor? Which programs vary consistently across many tumors? How does the identity of variable programs depend on the cancer type? These questions could also be phrased in the language of clusters and more generally relate to the fundamental question of which processes and phenotypes tend to vary among cancer cells in a given tumor. In the subsequent sections, we address these questions based on the many recent scRNA-seq tumor studies and their meta-analysis,[29,30] which highlight similarities in eITH patterns across tumors and provides a framework to discuss the relative frequencies of the various eITH programs.

MOST COMMON eITH PROGRAMS: CELL CYCLE AND STRESS

Arguably, the most fundamental property of cancer cells is their increased rate of proliferation, such that any tumor is expected to contain proliferating cancer cells. Since scRNA-seq captures a static snapshot, it is further expected that only a minority of cancer cells will be profiled during their cell cycle. Accordingly, the most commonly observed pattern of eITH is a distinction between cycling and non-cycling cells. The fraction of cycling cells varies extensively between tumors, from rare (<1%) in some low-proliferation tumors[2] to the majority of cells in some rapidly proliferating tumors[11] and in typical cell line models,[31] where proliferation is enriched relative to *in vivo* tumors.

Cycling cells can further be separated into distinct phases of the cell cycle.[2,32,33] In particular, there are two prominent cell cycle programs that are highly consistent between diverse types of cancer and non-cancer cells, largely reflecting core gene-sets of the G1/S and G2/M phases, respectively. Cells may be assigned to distinct phases along the cell cycle based on the relative activation of these two programs. Cells that do not express these two programs may be considered non-cycling, although it is difficult to distinguish if they are only non-cycling at the time of sampling (e.g., in early phase of G1) or whether they do not proliferate for a longer period (quiescent or senescent cells, i.e., in G0).

Some studies exclude cycling cells from their datasets. However, the amount of cycling cells and their distribution across cell states (as defined by processes other than the cell cycle itself) are important tumor features that may only be ascertained if the cycling cells are retained for analysis. Among cycling cells, much of the transcriptome may be devoted to proliferation, and signals for other processes may appear weaker than in non-cycling cells, but multiple studies showed that cycling cells are still associated with a diversity of states, although particular states are often enriched. For example, in gliomas, progenitor states are enriched for cycling cells, while more differentiated cells are depleted in proliferative potential, but such associations are partial and may vary between specimens.[2,21] Therefore, methods that attempt to normalize scRNA-seq data for the phase of the cell cycle assume a precise relationship between cell cycle and cell state, an oversimplification that may distort the inferred state of proliferating cells and mask a more complex pattern of cancer cell proliferation.

The second most common pattern of ITH is the so-called cellular stress-related programs.[29,34] These include hypoxia-related programs, in which the top genes reflect central mechanisms by which cancer cells cope with hypoxia such as angiogenesis (*VEGFA*), HIF inhibition (*EGLN3*) and increased glucose uptake (*SLC2A1*). Heat-shock response and unfolded-protein response are additional examples of common cellular stress pathways, and there are various other stress-related expression profiles in tumors.

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

Despite the diversity of physiological stresses, the observed stress-related programs typically share a common set of genes, thereby defining a generic stress program that appears to be induced by multiple stimuli and hence is often difficult to interpret. This program may be linked to drug resistance[34] and appears to be orchestrated by multiple transcription factors of the AP-1 family (*JUN*/*FOS*) along with *ATF3* and further includes heat-shock proteins, DNA damage genes and other stress-related genes.[29,34] This generic program can be induced by tumor dissociation and hence in some cases may be artificially inflated,[35,36] but it is also observed frequently in studies that avoided dissociation (such as spatial transcriptomics or single-nucleus RNA-seq [snRNA-seq] data) and hence is a genuine program of eITH that is broadly observed across all major cancer types.[29,34]

While stress-related programs are often difficult to connect to a specific source of stress, they suggest a prominent role of the location of cells within the tumor, with different tumor regions subjecting the cells to different stresses such as hypoxia, nutrient limitation due to competition, heat, acidity and others. Indeed, recent studies have begun to demonstrate that cancer cell states tend to be spatially segregated,[30,37] reflecting the rapidly expanding field of spatial omics.[38–40] For example, in glioblastoma, hypoxia serves as a driver of tumor spatial organization.[41]

eITH PROGRAMS PARTIALLY REFLECT DEVELOPMENTAL AND PHYSIOLOGICAL PROCESSES

It has long been appreciated that tumors co-opt processes from development and physiology.[42] In response to stimuli from their environment, or even simply as a stochastic process,[43] cancer cells may induce programs that have evolved for a developmental or physiological purpose. The recent explosion in tumor scRNA-seq studies enables a renewed analysis of this concept.

Probably the most well-known example of a developmental (and physiological) process that is co-opted in cancer is that of epithelial-mesenchymal transitions (EMT). EMT has been studied extensively in the cancer context, where it is thought to endow cancer cells with increased metastatic potential, drug resistance and possibly immunosuppression.[44] Nevertheless, the role of EMT in cancer has often been shrouded by debates around both the existence of EMT and its importance for metastasis and drug resistance. Numerous scRNA-seq studies have uncovered eITH programs enriched with mesenchymal-related genes, supporting the broad importance of EMT in cancer. In fact, such programs were observed across all types of epithelial cancers that were studied extensively by scRNA-seq.[29,30] Moreover, mesenchymal-related programs were also detected in various types of non-epithelial cancer types,[29] such as melanoma,[45,46] glioma,[21] neuroblastoma,[47] and sarcoma[48] (where in principle these programs should not be referred to as EMT). Thus, like the capacity to proliferate and to activate stress responses, also the ability to mount mesenchymal-related responses appears to be a common characteristic of cancer cells across diverse types of cancer.

While scRNA-seq studies provide strong evidence for the existence of EMT-related programs in cancer, they also may help to explain the debate around EMT. Most identified mesenchymal-related programs may be better described as representing a hybrid or partial EMT,[10,49,50] in which cells maintain their expression of epithelial markers along with only a partial activation of a subset of the mesenchymal markers that may be expected by classical description of EMT. For example, some EMT-related programs identified by scRNA-seq lack any of the core EMT transcription factors that were traditionally deemed essential for EMT. Thus, the annotation of such programs as EMT-related is itself under debate, with many researchers not considering a program as EMT-related only due to the enrichment of mesenchymal genes. In summary, mesenchymal-related eITH programs are common in cancer, but they appear to be much more limited than those in development and in physiology such that even their designation as EMT is questionable. Some of the studies that detected such programs also provided evidence for their association with metastasis, drug resistance or reduced survival,[10,45,46,50] although most scRNA-seq studies are statistically underpowered to examine such associations directly from clinical specimen. Experimental models coupled to scRNA-seq may offer a granular approach to dissect the establishment of resistance. For example, in ovarian cancer models, single-cell analysis suggests that EMT facilitates the adaptation to treatment and plays a role in the establishment of a resistance continuum of states.[51]

The EMT example highlights a wider phenomenon: eITH tends to resemble developmental and physiological processes, but often reflects only a partial version of these processes that may lack canonical markers and whose phenotypic consequences are difficult to assess. Another example of this concept involves cellular senescence: multiple epithelial cancer types, and especially squamous cell carcinoma, often harbor subpopulations of cells that express a program with high similarity to that of senescent epithelial cells.[29,31] These cells do not seem to activate classical markers of senescence (e.g., p16 and p21) and they appear to maintain a proliferative capacity, albeit reduced compared to other cells in the same tumor.

Further examples are seen in multiple other cancer types. In glioma, most cancer cells activate programs that are reminiscent of one of several lineages of neurodevelopmental cell types—oligodendrocyte progenitors, neural progenitors and astrocytes.[21] But glioma cells that induce those programs express distorted versions of them and remain transcriptionally more similar to one another than to their respective normal neurodevelopmental cells. In melanoma, cancer cells express melanocyte-like, neural crest-like, and mesenchymal-like programs[1,52,53] but the signal for such states may be subtle and difficult to distinguish within a given melanoma tumor and these states usually do not imply the phenotypes of respective normal cells (e.g., pigmentation). In chronic lymphocytic leukemia (CLL), malignant cells recapitulate B cell programs, but with decreased coordination of epigenetic layers that leads to increased transcriptional heterogeneity.[54] In lung cancer, subsets of cells express an alveolar-like program that resembles AT2 cells,[55,56] but once again this program lacks features of the normal AT2 cells such as activation of major histocompatibility complex class II (MHC-II) genes.[29]

In colorectal cancer, multiple developmental programs were detected in cancer cells, including colonic stem/transit-amplifying-like cells, and neuroendocrine-like cells.[57–59] Notably, the latter program is scarce within primary colorectal tumors but

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

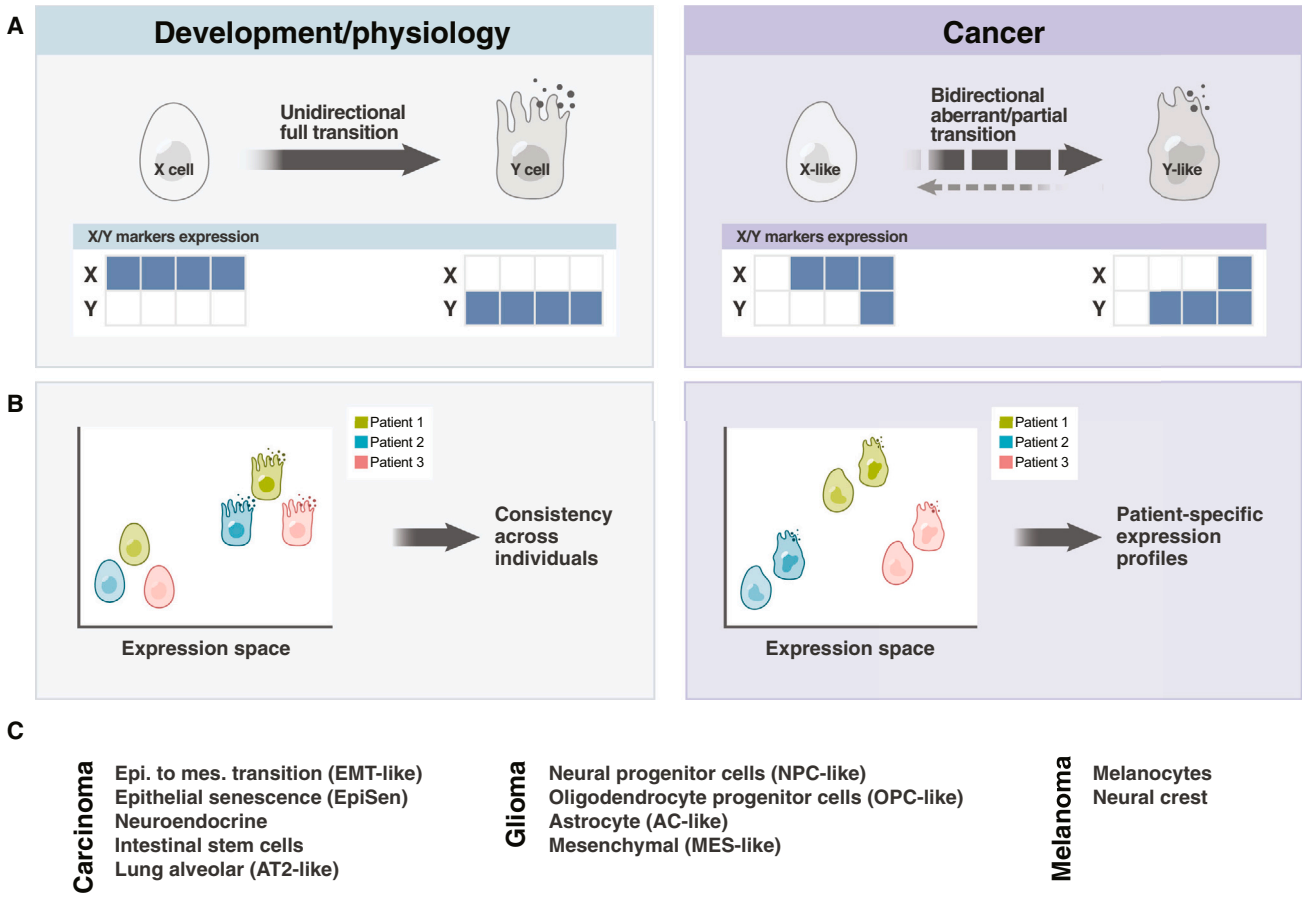

Figure 2. Cancer eITH programs often reflect limited versions of developmental or physiological programs

(A) Left: schematic depiction of a cellular transition/differentiation from cell type X to cell type Y, which is unidirectional, is accompanied by morphological changes and a complete transition from expression of only X markers to only Y markers. Right: a similar transition is found in cancer (from X-like to Y-like cells) where it is potentially bidirectional and is accompanied by partial morphological changes and a partial transition from expression of mostly X markers to mostly Y markers.

(B) Left: in development/physiology, X and Y cells from different individuals cluster together in expression space by cell type rather than by the individual they are derived from. Right: in cancer, X-like and Y-like cells cluster together in expression space by the patient they are derived from rather than by their cell-type; cell type does have a consistent influence on gene expression (i.e., Y-like cells are shifted to the top-right of the X-like cells from the same patient), but the patient-specific effect is larger than the effect of the cellular transition from X-like to Y-like.

(C) List of developmental or physiological cell types and cell states for which a similar state was observed in cancer and the transition to or from that state may fit the scenario described in (A) and (B).

becomes more prominent in liver metastases, suggesting that it either facilitates metastasis and/or is stimulated by the metastatic niche. Interestingly, neuroendocrine programs were seen not only in colorectal cancer metastases but also in post-treatment lung and prostate cancer, where they have been linked to drug resistance to inhibition of EGFR and androgen receptor, respectively.[60,61] Thus, programs that mimic normal development or physiology are seen as the core component of eITH in some cancer types (e.g., glioma and melanoma) and as acquired components of eITH that are linked to metastasis or drug resistance in other cancer types (colorectal, lung and prostate).

Taken together, scRNA-seq studies support the claim that cancer cells often induce distorted versions of programs from normal development or physiology and highlight that cancer programs tend to reflect limited versions of their normal counterparts, hence complicating their detection and interpretation (Figure 2). Moreover, the cancer eITH programs are embedded on top of the unique genetics and transcriptome of each tumor, resulting in patient-specific profiles (Figure 2B). Accordingly, it would be important to explore whether the restricted and distorted cancer versions are associated with unique phenotypes such as those expected by the normal processes. In some cases, the normal process involves a terminal differentiation such that the ability to further induce that differentiation may limit tumor proliferation and provide a clinical benefit. For example, in IDH-mutant glioma, which harbors astrocyte-like cells, further induction of astrocytic differentiation appears to decrease the proliferation of cancer cells and provide a clinical benefit.[62]

eITH PROGRAMS LINKED TO INTERACTIONS WITH IMMUNE CELLS

With the prevalent interest in immunotherapies, much of the attention in tumor scRNA-seq studies has been placed on

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

immune cells, especially T cells. Diversity among immune cells is reviewed elsewhere,[63,64] but even when focusing on the cancer cells, some eITH programs may have implications for immunotherapies. For example, several studies reported on cancer cell programs associated with the exclusion of immune cells from the core of the tumor, such as in synovial sarcoma.[48] In colorectal cancer, three immune-related programs expressed by cancer cells, especially in mismatch repair-deficient (MMRd) tumors, may mediate interactions with the immune system.[65] EMT programs were also linked to immunosuppression.[44,46,66] Interestingly, this association appears to be context-dependent, as mesenchymal programs in glioma are associated with increased, rather than decreased, immune activity.[67,68] This positive association with immune activity appeared to be mediated by the upregulation of MHC-I and MHC-II genes by the mesenchymal-like cancer cells.

Expression of MHC-I genes is known to be disrupted in some cancers,[69] reflecting a mechanism of immune evasion, but these genes are only rarely induced in eITH programs. In contrast, MHC-II genes are commonly induced by cancer cells, often together with classical interferon-response genes (e.g., *ISG15*, *IFIT3*, *STAT1*, and *OAS1*).[29,30,46,70] In fact, in most cancer types, interferon-response genes are highly correlated with MHC-II genes such that detected eITH programs typically contain both of these sets of genes (interferon-response + MHC-II genes). This tight coupling may suggest that interferon is secreted by T cells and other immune cells, and then induces the coordinated expression of these sets of genes by adjacent cancer cells.

The induction of MHC-II genes is not restricted to subsets of cancer cells but is also observed in various cell types within the TME. Beyond macrophages and dendritic cells—the professional antigen-presenting cell (APCs)—high MHC-II was observed in subsets of fibroblasts and endothelial cells.[29,71,72] However, in those cell types MHC-II expression was decoupled from interferon-responses, suggesting distinct modes of regulation of MHC-II genes and interferon response in non-epithelial cell types.[29]

MHC-II and many interferon-response genes are sometimes considered as immune genes and even as markers of APCs, as they are traditionally expressed by immune cell types, but as noted previously, they are also induced in subsets of cancer cells and other TME cell types. These observations suggest caution when interpreting the expression profiles of such genes in bulk samples and when applying deconvolution to assess immune cell type frequencies.

LEVERAGING SINGLE-CELL SIGNATURES FOR DECONVOLUTION OF BULK PROFILES

The wealth of published tumor scRNA-seq datasets implies that most cancer-related cell types and cell states may have already been described. The corresponding signatures of cell types and cell states may be leveraged for improved deconvolution analysis of bulk RNA-seq. Deconvolution refers to the estimation of cell type and cell state frequencies in a sample such that the combined expression profiles of all those cells would provide the best possible fit for the observed bulk RNA-seq profile. This approach has become widely used in recent years,[73,74] reflecting another outcome of the single-cell revolution. We expect this approach to continue to be widely used and to facilitate consideration of the entire tumor ecosystem even from bulk RNA-seq datasets.

However, it is important to also recognize the limitations of deconvolution approaches. As demonstrated by scRNA-seq datasets, the genes considered as markers of a certain cell type (or cell state) are almost never fully restricted to these cells, and instead are also expressed (albeit at lower levels) by multiple other cell types/states. The exact degree of such "non-canonical" expression varies significantly between cell types, samples and studies in a manner that cannot be fully predicted by deconvolution approaches and that therefore limits their accuracy. The impact of non-canonical expression scales with the frequency of the associated cell types/states. For example, when inferring the frequencies of an abundant cell type (A) and a rare cell type (B), non-canonical expression of markers of A by cells of B will have minimal influence, while non-canonical expression of markers of B by cells of A will have a much larger influence. Thus, frequency estimation of rare cell types is particularly prone to errors.

Such errors may be even more pronounced when estimating the frequency of cell states, which tend to have signatures with lower specificity that are also expressed by other states of the same cell type or of other cell types. Cell cycle, stress responses, mesenchymal programs (e.g., EMT), interferon responses, and antigen presentation, are all examples of expression programs that may be activated by multiple cell types and often are concomitantly activated by multiple cell types.[29] Hence, the corresponding signatures could not be confidently ascribed to a particular cell type. In summary, deconvolution is a highly useful approach to extract more information from bulk profiles but the results should be interpreted with great caution, especially for rare cell types/states and for those with limited specificity of the marker genes. An accurate and nuanced analysis of cell states will therefore continue to require single cell measurements despite advances with bulk deconvolution.

CHALLENGES IN ANNOTATION OF eITH PROGRAMS

We discussed previously the most common eITH programs including cell cycle, stress, mesenchymal, senescence, immune-related (MHC-II and interferon-response), and various developmental lineage-related programs. Other eITH programs not discussed previously but still commonly observed include those associated with global protein regulation, respiration, MYC activation, and secretion.[29]

However, it is important to consider the ambiguity in defining and naming eITH programs which implies that programs reported in many studies might seem distinct and could be difficult to fit into any of those categories. For example, consider a tumor with cancer cells in two equally common lineage-related states (A and B) and in which 90% of the cycling cells are in state A. Standard analysis would likely define a signature of A cells that includes both genuine A markers as well as generic cell cycle markers (that would also be expressed by cycling B cells). Such combined signature may be named as proliferating A cells and hinder the detection of pure signatures of A cells and of the cell cycle. Similarly, if stress is primarily activated in cells of one lineage, then stress genes may be contained within a signature

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

CellPress

of that lineage rather than discerned as a separate signature and cellular state. These examples of proliferation and stress response accompanying another cell state, reflect a more general phenomenon—that cells may turn on programs of multiple states simultaneously, thereby complicating the detection and definition of discrete signatures. The degree to which distinct cellular processes are lumped together in state signatures varies across studies, depending on their methodology and the coordination among such processes in the respective cohorts.

Even for a well-defined set of signature genes, there is often enrichment with various functional gene-sets (e.g., linked to developmental, mechanistic, and metabolic processes), and accordingly the signature could be named differently by distinct researchers, depending on their biological perspective and exact methodology. These caveats highlight the need to converge to consistent definitions of cellular states and eITH programs and at the very least to acknowledge the ambiguities and the similarities across studies and state definitions.

FUTURE DEVELOPMENTS: EXPANDING THE REACH OF SINGLE CELL PROFILING

While scRNA-seq of tumor samples has already matured, we view the next frontier as a broader incorporation of these approaches in the clinic, enabling the study of both larger cohorts as well as of highly informative samples, such as on-treatment samples, pre-neoplastic lesions, and exceptional responders. This is greatly facilitated by recent technological developments. snRNA-seq methods[75] have opened the possibility of profiling snap-frozen specimens (either from biobanks or prospectively collected). Indeed, the collection of fresh tumor tissue for scRNA-seq in the clinic presents unique challenges, such as the need for rapid tissue dissociation.[76] snRNA-seq bypasses such needs, as the nuclear membrane remains intact during freezing. While snRNA-seq provides lower gene counts and somewhat different gene representation compared to scRNA-seq, the biological processes highlighted are consistent between snRNA-seq and scRNA-seq, offering a practical and scalable solution to profiling large clinical collections of tissue specimens.[75]

Another key recent development is the possibility of using formalin-fixed or even FFPE (formalin-fixed, paraffin-embedded) blocks for scRNA-seq and spatial profiling. These samples have been extremely challenging to profile as formalin fragments and crosslinks RNA molecules and degradation of RNA may also occur prior to fixation of the clinical specimen due to delays in collection times. Despite these limitations, FFPE represents the most common archival source of clinical material and offers exquisite tissue morphology. Recent methods (snPATHO-seq,[77] snFFPE-seq,[78] and snRandome-seq[79]) have improved FFPE profiling by optimizing RNA extraction and/or leveraging random primers (instead of oligo-dTs that fail on degraded RNA). We view these methods as critical to the broader implementation of sc/snRNA-seq workflows in cancer medicine.

BEYOND RNA PROFILING

Technological developments are ushering a new era in which single cell methods extend from RNA profiling to the joint profiling of multiple modalities (genome, epigenome, transcriptome, proteome, and/or metabolome) from the same cell. For example, multi-omic methods can link the genotypes to the transcriptome, empowering the integration of genetic and non-genetic mechanisms underlying eITH programs and cancer evolution.[80–82] Other methods jointly interrogate DNA methylation and RNA,[83] and have been leveraged to glean insights into dysregulated epigenetic mechanisms in gliomagenesis.[84] In these approaches, genotypes may either be fully sequenced, or may be captured by target-specific primers that facilitate the combination with measurements of mRNA (GoT[81]), protein (GoT-Splice[85]), or chromatin-accessibility profiles (GoT-ChA[86]). While many multi-omic modalities are still being developed or are confined to specialized laboratories, their broader implementation is expected to empower the investigation of the link between genotypes, epigenomes, and phenotypes. By directly comparing different clones in the same tumor, such approaches also offer an ideal setting to study the impact of somatic mutations on eITH while controlling for other variables such as patient-specific effects and TME.

In addition to the modalities noted previously, spatial profiling methods are particularly poised to revolutionize cancer research.[38–40] The exact location of cells in each particular state will help to uncover mechanisms that drive this state; proximity between different cell types and cell states will reveal important cell-cell interactions; and higher-order tumor organization will provide insights into additional tumor phenotypes such as the degree of immune infiltration and the presence of immune hubs that correlate with an effective immune response.[41,87,88]

Since 2015–2016, when the first spatial transcriptomics studies were published,[89,90] the field of spatial transcriptomics has been expanding dramatically, with multiple methods that provide much higher resolution, which are either imaging-based (typically hybridization probe-based, e.g., CosMx and Xenium) or sequencing based (after *in situ* indexing, e.g., Slide-Tags[91] and Visium HD). The application of spatial transcriptomics and proteomics to tumors is still in its infancy and is expected to expand in the next few years. Methods such as Slide-Tags are compatible with single-cell multi-omics workflows and may provide a very compelling approach to interrogate genetic and non-genetic cancer mechanisms *in situ.* High-resolution spatial methods may also play an important role in the detection and characterization of pre-neoplastic lesions, by complementing histopathology with detailed, spatially resolved, molecular information.[92] As many of these techniques are non-destructive, they provide a unique opportunity for direct implementation into existing clinical workflows.

ACKNOWLEDGMENTS

This work was supported by an ERC (European Research Council) consolidator grant 101044318 (I.T.), the MGH Research Scholars Award (M.L.S.), and grant from the Mark Foundation Emerging Leader Award (M.L.S.). I.T. is the incumbent of the Dr. Celia Zwillenberg-Fridman and Dr. Lutz Zwillenberg Career Development Chair, and is supported by the Zuckerman STEM Leadership Program. We thank members of the Tirosh and Suva labs for helpful discussions.

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

DECLARATION OF INTERESTS

M.L.S. is equity holder, scientific co-founder and advisory board member of Immunitas Therapeutics. I.T. is an advisory board member of Immunitas Therapeutics.

REFERENCES

1. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.

2. Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature *539*, 309–313.

3. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

4. Miyamoto, D.T., Zheng, Y., Wittner, B.S., Lee, R.J., Zhu, H., Broderick, K.T., Desai, R., Fox, D.B., Brannigan, B.W., Trautwein, J., et al. (2015). RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science *349*, 1351–1356.

5. Kim, K.T., Lee, H.W., Lee, H.O., Song, H.J., Jeong, D.E., Shin, S., Kim, H., Shin, Y., Nam, D.H., Jeong, B.C., et al. (2016). Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. Genome Biol. *17*, 80.

6. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

7. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

8. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.

9. Gierahn, T.M., Wadsworth, M.H., 2nd, Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods *14*, 395–398.

10. Puram, S., Tirosh, I., Parikh, A., Patel, A., Yizhak, K., Gillespie, S., Rodman, R., Luo, C., Mroz, E., Emerick, K., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell *171*, 1611–1624.e24.

11. Puram, S.V., Mints, M., Pal, A., Qi, Z., Reeb, A., Gelev, K., Barrett, T.F., Gerndt, S., Liu, P., Parikh, A.S., et al. (2023). Cellular states are coupled to genomic and viral heterogeneity in HPV-related oropharyngeal carcinoma. Nat. Genet. *55*, 640–650.

12. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., and Theis, F.J. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nat. Methods *19*, 41–50.

13. Argelaguet, R., Cuomo, A.S.E., Stegle, O., and Marioni, J.C. (2021). Computational principles and challenges in single-cell data integration. Nat. Biotechnol. *39*, 1202–1215.

14. Emmanúel Antonsson, S., and Melsted, P. (2024). Batch correction methods used in single cell RNA-sequencing analyses are often poorly calibrated. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.19.585562.

15. Tyler, S.R., Guccione, E., and Schadt, E.E. (2023). Erasure of Biologically Meaningful Signal by Unsupervised scRNAseq Batch-correction Methods. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.15.468733.

16. Zeng, H. (2022). What is a cell type and how to define it? Cell *185*, 2739–2755.

17. Perez-Gonzalez, A., Bevant, K., and Blanpain, C. (2023). Cancer cell plasticity during tumor progression, metastasis and response to therapy. Nat. Can. (Ott.) *4*, 1063–1082.

18. Davies, A., Zoubeidi, A., Beltran, H., and Selth, L.A. (2023). The Transcriptional and Epigenetic Landscape of Cancer Cell Lineage Plasticity. Cancer Discov. *13*, 1771–1788.

19. Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435.

20. Rambow, F., Marine, J.C., and Goding, C.R. (2019). Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities. Genes Dev. *33*, 1295–1318.

21. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835–849.e21.

22. Liau, B.B., Sievers, C., Donohue, L.K., Gillespie, S.M., Flavahan, W.A., Miller, T.E., Venteicher, A.S., Hebert, C.H., Carey, C.D., Rodig, S.J., et al. (2017). Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. Cell Stem Cell *20*, 233–246.e7.

23. Burdziak, C., Alonso-Curbelo, D., Walle, T., Reyes, J., Barriga, F.M., Haviv, D., Xie, Y., Zhao, Z., Zhao, C.J., Chen, H.A., et al. (2023). Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis. Science *380*, eadd5327.

24. Chan, J.M., Zaidi, S., Love, J.R., Zhao, J.L., Setty, M., Wadosky, K.M., Gopalan, A., Choo, Z.N., Persad, S., Choi, J., et al. (2022). Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. Science *377*, 1180–1191.

25. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov. *12*, 31–46.

26. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

27. Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. *37*, 547–554.

28. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

29. Gavish, A., Tyler, M., Greenwald, A.C., Hoefflin, R., Simkin, D., Tschernichovsky, R., Galili Darnell, N., Somech, E., Barbolin, C., Antman, T., et al. (2023). Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. Nature *618*, 598–606.

30. Barkley, D., Moncada, R., Pour, M., Liberman, D.A., Dryg, I., Werba, G., Wang, W., Baron, M., Rao, A., Xia, B., et al. (2022). Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. Nat. Genet. *54*, 1192–1201.

31. Kinker, G.S., Greenwald, A.C., Tal, R., Orlova, Z., Cuoco, M.S., McFarland, J.M., Warren, A., Rodman, C., Roth, J.A., Bender, S.A., et al. (2020). Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. Nat. Genet. *52*, 1208–1218.

32. Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. *25*, 1860–1872.

33. Guo, X., and Chen, L. (2024). From G1 to M: a comparative study of methods for identifying cell cycle phases. Briefings Bioinf. *25*, bbad517.

34. Baron, M., Tagore, M., Hunter, M.V., Kim, I.S., Moncada, R., Yan, Y., Campbell, N.R., White, R.M., and Yanai, I. (2020). The Stress-Like Cancer Cell State Is a Consistent Component of Tumorigenesis. Cell Syst. *11*, 536–546.e7.

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

35. Neuschulz, A., Bakina, O., Badillo-Lisakowski, V., Olivares-Chauvet, P., Conrad, T., Gotthardt, M., Kettenmann, H., and Junker, J.P. (2023). A single-cell RNA labeling strategy for measuring stress response upon tissue dissociation. Mol. Syst. Biol. *19*, e11147.

36. van den Brink, S.C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C., and van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat. Methods *14*, 935–936.

37. Hunter, M.V., Moncada, R., Weiss, J.M., Yanai, I., and White, R.M. (2021). Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. Nat. Commun. *12*, 6278.

38. Bressan, D., Battistoni, G., and Hannon, G.J. (2023). The dawn of spatial omics. Science *381*, eabq4964.

39. Seferbekova, Z., Lomakin, A., Yates, L.R., and Gerstung, M. (2023). Spatial biology of cancer evolution. Nat. Rev. Genet. *24*, 295–313.

40. Elhanani, O., Ben-Uri, R., and Keren, L. (2023). Spatial profiling technologies illuminate the tumor microenvironment. Cancer Cell *41*, 404–420.

41. Greenwald, A.C., Darnell, N.G., Hoefflin, R., Simkin, D., Mount, C.W., Gonzalez Castro, L.N., Harnik, Y., Dumont, S., Hirsch, D., Nomura, M., et al. (2024). Integrative spatial analysis reveals a multi-layered organization of glioblastoma. Cell *187*, 2485–2501.e26.

42. Pierce, G.B., and Speers, W.C. (1988). Tumors as caricatures of the process of tissue renewal: prospects for therapy by directing differentiation. Cancer Res. *48*, 1996–2004.

43. Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

44. Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. Nat. Rev. Mol. Cell Biol. *20*, 69–84.

45. Pedri, D., Karras, P., Landeloos, E., Marine, J.C., and Rambow, F. (2022). Epithelial-to-mesenchymal-like transition events in melanoma. FEBS J. *289*, 1352–1368.

46. Pozniak, J., Pedri, D., Landeloos, E., Van Herck, Y., Antoranz, A., Vanwynsberghe, L., Nowosad, A., Roda, N., Makhzami, S., Bervoets, G., et al. (2024). A TCF4-dependent gene regulatory network confers resistance to immunotherapy in melanoma. Cell *187*, 166–183.e25.

47. Thirant, C., Peltier, A., Durand, S., Kramdi, A., Louis-Brennetot, C., Pierre-Eugène, C., Gautier, M., Costa, A., Grelier, A., Zaïdi, S., et al. (2023). Reversible transitions between noradrenergic and mesenchymal tumor identities define cell plasticity in neuroblastoma. Nat. Commun. *14*, 2575.

48. Jerby-Arnon, L., Neftel, C., Shore, M.E., Weisman, H.R., Mathewson, N.D., McBride, M.J., Haas, B., Izar, B., Volorio, A., Boulay, G., et al. (2021). Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. Nat. Med. *27*, 289–300.

49. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468.

50. Simeonov, K.P., Byrns, C.N., Clark, M.L., Norgard, R.J., Martin, B., Stanger, B.Z., Shendure, J., McKenna, A., and Lengner, C.J. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell *39*, 1150–1162.e9.

51. Franca, G.S., Baron, M., King, B.R., Bossowski, J.P., Bjornberg, A., Pour, M., Rao, A., Patel, A.S., Misirlioglu, S., Barkley, D., et al. (2024). Cellular adaptation to cancer therapy along a resistance continuum. Nature *631*, 876–883.

52. Rambow, F., Rogiers, A., Marin-Bejar, O., Aibar, S., Femel, J., Dewaele, M., Karras, P., Brown, D., Chang, Y.H., Debiec-Rychter, M., et al. (2018). Toward Minimal Residual Disease-Directed Therapy in Melanoma. Cell *174*, 843–855.e19.

53. Karras, P., Bordeu, I., Pozniak, J., Nowosad, A., Pazzi, C., Van Raemdonck, N., Landeloos, E., Van Herck, Y., Pedri, D., Bervoets, G., et al. (2022). A cellular hierarchy in melanoma uncouples growth and metastasis. Nature *610*, 190–198.

54. Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K., Risso, D., Kim, K.T., Pastore, A., et al. (2019). Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature *569*, 576–580.

55. Bischoff, P., Trinks, A., Obermayer, B., Pett, J.P., Wiederspahn, J., Uhlitz, F., Liang, X., Lehmann, A., Jurmeister, P., Elsner, A., et al. (2021). Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. Oncogene *40*, 6748–6758.

56. Maynard, A., McCoach, C.E., Rotow, J.K., Harris, L., Haderk, F., Kerr, D.L., Yu, E.A., Schenk, E.L., Tan, W., Zee, A., et al. (2020). Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell *182*, 1232–1251.e22.

57. Becker, W.R., Nevins, S.A., Chen, D.C., Chiu, R., Horning, A.M., Guha, T.K., Laquindanum, R., Mills, M., Chaib, H., Ladabaum, U., et al. (2022). Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. Nat. Genet. *54*, 985–995.

58. Moorman, A.R., Cambuli, F., Benitez, E.K., Jiang, Q., Xie, Y., Mahmoud, A., Lumish, M., Hartner, S., Balkaran, S., Bermeo, J., et al. (2023). Progressive Plasticity during Colorectal Cancer Metastasis. Preprint at bioRxiv. https://doi.org/10.1101/2023.08.18.553925.

59. Canellas-Socias, A., Cortina, C., Hernando-Momblona, X., Palomo-Ponce, S., Mulholland, E.J., Turon, G., Mateo, L., Conti, S., Roman, O., Sevillano, M., et al. (2022). Metastatic recurrence in colorectal cancer arises from residual EMP1(+) cells. Nature *611*, 603–613.

60. Rubin, M.A., Bristow, R.G., Thienger, P.D., Dive, C., and Imielinski, M. (2020). Impact of Lineage Plasticity to and from a Neuroendocrine Phenotype on Progression and Response in Prostate and Lung Cancers. Mol. Cell. *80*, 562–577.

61. Romero, R., Chu, T., Gonzalez-Robles, T.J., Smith, P., Xie, Y., Kaur, H., Yoder, S., Zhao, H., Mao, C., Kang, W., et al. (2024). The neuroendocrine transition in prostate cancer is dynamic and dependent on ASCL1. Preprint at bioRxiv. https://doi.org/10.1101/2024.04.09.588557.

62. Spitzer, A., Gritsch, S., Nomura, M., Jucht, A., Fortin, J., Raviram, R., Weisman, H.R., Gonzalez Castro, L.N., Druck, N., Chanoch-Myers, R., et al. (2024). Mutant IDH inhibitors induce lineage differentiation in IDH-mutant oligodendroglioma. Cancer Cell *42*, 904–914.e9.

63. Ren, X., Zhang, L., Zhang, Y., Li, Z., Siemers, N., and Zhang, Z. (2021). Insights Gained from Single-Cell Analysis of Immune Cells in the Tumor Microenvironment. Annu. Rev. Immunol. *39*, 583–609.

64. Davis-Marcisak, E.F., Deshpande, A., Stein-O'Brien, G.L., Ho, W.J., Laheru, D., Jaffee, E.M., Fertig, E.J., and Kagohara, L.T. (2021). From bench to bedside: Single-cell analysis for cancer immunotherapy. Cancer Cell *39*, 1062–1080.

65. Pelka, K., Hofree, M., Chen, J.H., Sarkizova, S., Pirl, J.D., Jorgji, V., Bejnood, A., Dionne, D., Ge, W.H., Xu, K.H., et al. (2021). Spatially organized multicellular immune hubs in human colorectal cancer. Cell *184*, 4734–4752.e20.

66. Taki, M., Abiko, K., Ukita, M., Murakami, R., Yamanoi, K., Yamaguchi, K., Hamanishi, J., Baba, T., Matsumura, N., and Mandai, M. (2021). Tumor Immune Microenvironment during Epithelial-Mesenchymal Transition. Clin. Cancer Res. *27*, 4669–4679.

67. Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell *39*, 779–792.e11.

68. Chanoch-Myers, R., Wider, A., Suva, M.L., and Tirosh, I. (2022). Elucidating the diversity of malignant mesenchymal states in glioblastoma by integrative analysis. Genome Med. *14*, 106.

69. Dhatchinamoorthy, K., Colbert, J.D., and Rock, K.L. (2021). Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. Front. Immunol. *12*, 636568.

Please cite this article in press as: Tirosh and Suva, Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors, Cancer Cell (2024), https://doi.org/10.1016/j.ccell.2024.08.005

70. Izar, B., Tirosh, I., Stover, E.H., Wakiro, I., Cuoco, M.S., Alter, I., Rodman, C., Leeson, R., Su, M.J., Shah, P., et al. (2020). A single-cell landscape of high-grade serous ovarian cancer. Nat. Med. *26*, 1271–1279.

71. Lavie, D., Ben-Shmuel, A., Erez, N., and Scherz-Shouval, R. (2022). Cancer-associated fibroblasts in the single-cell era. Nat. Can. (Ott.) *3*, 793–807.

72. Macy, A.M., Herrmann, L.M., Adams, A.C., and Hastings, K.T. (2023). Major histocompatibility complex class II in the tumor microenvironment: functions of nonprofessional antigen-presenting cells. Curr. Opin. Immunol. *83*, 102330.

73. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782.

74. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172.

75. Slyper, M., Porter, C.B.M., Ashenberg, O., Waldman, J., Drokhlyansky, E., Wakiro, I., Smillie, C., Smith-Rosario, G., Wu, J., Dionne, D., et al. (2020). A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat. Med. *26*, 792–802.

76. Lim, B., Lin, Y., and Navin, N. (2020). Advancing Cancer Research and Medicine with Single-Cell Genomics. Cancer Cell *37*, 456–470.

77. Vallejo, A.F., Harvey, K., Wang, T., Wise, K., Butler, L.M., Polo, J., Plummer, J., Swarbrick, A., and Martelotto, L.G. (2022). snPATHO-seq: unlocking the FFPE archives for single nucleus RNA profiling. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.23.505054.

78. Chung, H., Melnikov, A., McCabe, C., Drokhlyansky, E., Van Wittenberghe, N., Magee, E.M., Waldman, J., Spira, A., Chen, F., Mazzilli, S., et al. (2022). SnFFPE-Seq: towards scalable single nucleus RNA-Seq of formalin-fixed paraffin-embedded (FFPE) tissue. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.25.505257.

79. Xu, Z., Zhang, T., Chen, H., Zhu, Y., Lv, Y., Zhang, S., Chen, J., Chen, H., Yang, L., Jiang, W., et al. (2023). High-throughput single nucleus total RNA sequencing of formalin-fixed paraffin-embedded tissues by snRandom-seq. Nat. Commun. *14*, 2734.

80. Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat. Methods *12*, 519–522.

81. Nam, A.S., Kim, K.T., Chaligne, R., Izzo, F., Ang, C., Taylor, J., Myers, R.M., Abu-Zeinah, G., Brand, R., Omans, N.D., et al. (2019). Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature *571*, 355–360.

82. Rodriguez-Meira, A., Buck, G., Clark, S.A., Povinelli, B.J., Alcolea, V., Louka, E., McGowan, S., Hamblin, A., Sousos, N., Barkas, N., et al. (2019). Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. Mol. Cell. *73*, 1292–1305.e8.

83. Gu, H., Raman, A.T., Wang, X., Gaiti, F., Chaligne, R., Mohammad, A.W., Arczewska, A., Smith, Z.D., Landau, D.A., Aryee, M.J., et al. (2021). Smart-RRBS for single-cell methylome and transcriptome analysis. Nat. Protoc. *16*, 4004–4030.

84. Chaligne, R., Gaiti, F., Silverbush, D., Schiffman, J.S., Weisman, H.R., Kluegel, L., Gritsch, S., Deochand, S.D., Gonzalez Castro, L.N., Richman, A.R., et al. (2021). Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. Nat. Genet. *53*, 1469–1479.

85. Cortes-Lopez, M., Chamely, P., Hawkins, A.G., Stanley, R.F., Swett, A.D., Ganesan, S., Mouhieddine, T.H., Dai, X., Kluegel, L., Chen, C., et al. (2023). Single-cell multi-omics defines the cell-type-specific impact of splicing aberrations in human hematopoietic clonal outgrowths. Cell Stem Cell *30*, 1262–1281.

86. Izzo, F., Myers, R.M., Ganesan, S., Mekerishvili, L., Kottapalli, S., Prieto, T., Eton, E.O., Botella, T., Dunbar, A.J., Bowman, R.L., et al. (2024). Mapping genotypes to chromatin accessibility profiles in single cells. Nature *629*, 1149–1157.

87. Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.R., Kurian, A., Van Valen, D., West, R., et al. (2018). A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell *174*, 1373–1387.e19.

88. Vanhersecke, L., Brunet, M., Guégan, J.P., Rey, C., Bougouin, A., Cousin, S., Moulec, S.L., Besse, B., Loriot, Y., Larroquette, M., et al. (2021). Mature tertiary lymphoid structures predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1 expression. Nat. Can. (Ott.) *2*, 794–802.

89. Stahl, P.L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

90. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

91. Russell, A.J.C., Weir, J.A., Nadaf, N.M., Shabet, M., Kumar, V., Kambhampati, S., Raichur, R., Marrero, G.J., Liu, S., Balderrama, K.S., et al. (2024). Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. Nature *625*, 101–109.

92. Carpenter, E.S., Elhossiny, A.M., Kadiyala, P., Li, J., McGue, J., Griffith, B.D., Zhang, Y., Edwards, J., Nelson, S., Lima, F., et al. (2023). Analysis of Donor Pancreata Defines the Transcriptomic Signature and Microenvironment of Early Neoplastic Lesions. Cancer Discov. *13*, 1324–1345.