Please cite this article in press as: Jones and McGranahan, Deciphering the landscape of transcriptional heterogeneity across cancer, Cancer Cell (2023), https://doi.org/10.1016/j.ccell.2023.07.008

Cancer Cell

CellPress

Spotlight

Deciphering the landscape of transcriptional heterogeneity across cancer

Thomas P. Jones[1,2] and Nicholas McGranahan[1,2,*]
[1]Cancer Genome Evolution Research Group, University College London Cancer Institute, London, UK
[2]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK
*Correspondence: nicholas.mcgranahan.10@ucl.ac.uk
https://doi.org/10.1016/j.ccell.2023.07.008

By integrating scRNA-seq datasets across 77 studies and 24 cancer types, in *Nature*, Gavish et al. uncover recurrent patterns of gene expression that explain a significant proportion of transcriptomic heterogeneity observed in cancer and explore their functional significance.

Cancer development is an evolutionary process where malignant cells accumulate genetic and non-genetic changes throughout the disease course.[1] These changes generate diverse cancer cell phenotypes and intra-tumor heterogeneity (ITH), which serves as the substrate for Darwinian evolution. Certain phenotypes may confer a selective advantage to subpopulations of cells.

Research into cancer evolution has predominantly focused on cancer as a disease of the genome, uncovering key driver alterations and mutational processes that shape the evolving cancer genome.[2] Importantly, however, selection takes place at the level of the phenotype, not genotype. More recently, the advent of single-cell RNA sequencing (scRNA-seq) has facilitated exploration of transcriptional ITH with unprecedented granularity and revealed the functional significance of phenotypic diversity. Indeed, non-genetic events are increasingly acknowledged to be key drivers of tumor evolution, with transcriptional cell states able to influence disease progression and the likelihood of metastasis.[3,4]

A key challenge when investigating transcriptional ITH during cancer evolution is distinguishing between functional and non-functional variation (including transcriptional noise). Indeed, not all observed differences between cancer cells are biologically meaningful. To infer the functional relevance of a given cell state, recurrent patterns across multiple tumors are of particular interest. Recurrent patterns of expression, found across multiple tumors, can be considered as potential "hallmarks" of tumor biology, representing evolutionarily conserved cell states that are replicated across diverse tumor environments. Identifying bona fide hallmarks requires large and diverse scRNA-seq catalogs. Writing in *Nature*, Gavish et al.[5] curate an impressive single-cell expression atlas across diverse cancer types to thoroughly investigate recurrent patterns of expression ITH and their functional importance.

This Curated Cancer Cell Atlas (3CA) consists of scRNA-seq data from more than 1.8 million cells compiled from 77 different studies across 24 cancer types (Figure 1). Distinct cell types were annotated using canonical cell-type marker expression, and cells were assigned to either malignant or non-malignant populations. Commendably, this comprehensively annotated dataset has been made available to the wider research community (weizmann.ac.il/sites/3CA/).

In order to detect recurrent expression programs, denoted by the authors as meta-programs, Gavish et al. first cataloged expression signatures within individual tumors. Expression signatures were inferred using non-negative matrix factorization (NMF) to identify patterns and underlying structures within single-cell expression matrices of tumors. The >5,000 signatures discovered in individual tumors could then be compared across the cohort. This analysis was performed for both malignant and non-malignant cells, allowing the authors to investigate tumor-specific meta-programs in the context of the tumor microenvironment (TME).

Clustering of signatures across the cohort identified 41 meta-programs, with 83% occurring over several cancer types. 20 of these meta-programs had been previously described[6,7], whereas 21 were novel, highlighting the added sensitivity provided by this expansive 3CA cohort. Additionally, 66% of all tumor-specific expression signatures were represented within these meta-programs, suggesting that they encapsulate a significant proportion of the functional expression ITH observed during tumor evolution. Building on these findings, the authors refined these meta-programs into 11 broader hallmarks, reflecting global biological processes or cell states (Figure 1).

Common pan-cancer hallmarks included those linked with core cellular functions such as progression through the cell cycle. Conversely, other features were found at consistently high levels in single cancer types. These likely reflect differences between non-genetic factors across tumor types such as the cellular differentiation state within the tissue of origin (and thus were generally part of a broader "lineage-related" hallmark).

Cancer hallmarks may reflect functionally relevant aspects of tumor biology, and thus provide valuable insights into malignant cell populations present within individual tumors that drive clinical features of disease. Indeed, Gavish et al. demonstrated associations between meta-programs and overall survival, disease grade and stage, lymph node metastasis, and therapeutic resistance using average meta-program abundances derived from bulk RNA-seq data from the TCGA cohort.

Please cite this article in press as: Jones and McGranahan, Deciphering the landscape of transcriptional heterogeneity across cancer, Cancer Cell (2023), https://doi.org/10.1016/j.ccell.2023.07.008

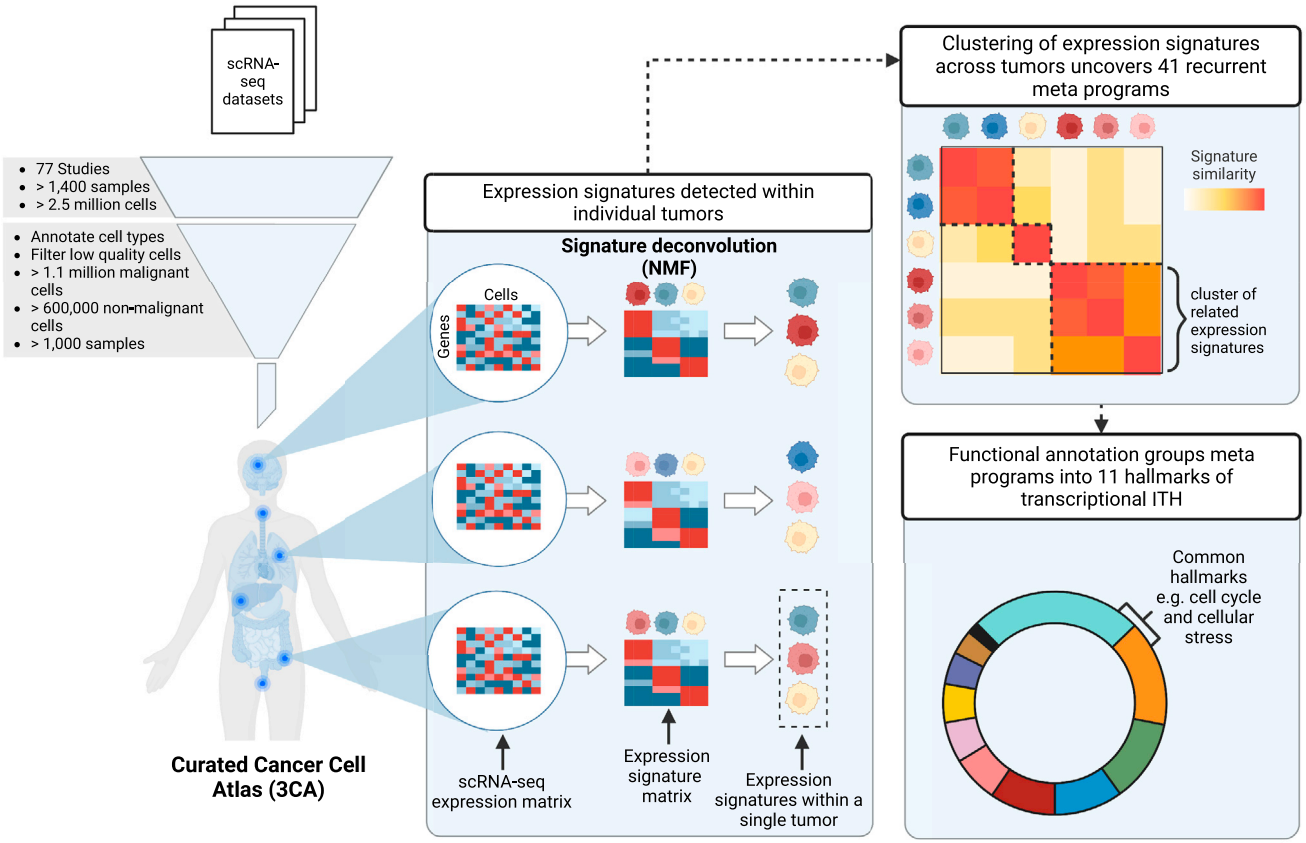

Figure 1. Uncovering the landscape of transcriptional ITH across cancer

Gavish et al. integrated scRNA-seq data from 77 studies across 24 cancer types. NMF was performed on individual tumor expression matrices to infer patterns of gene expression within single tumors. These signatures were then clustered across the cohort into 41 meta-programs of expression. Functional annotation of meta-programs grouped them into the 11 hallmarks of transcriptional ITH, which can be found at variable frequencies across the cohort.
Created with BioRender.com.

Given their clinical relevance, Gavish et al. went on to investigate genetic regulation of meta-programs. This is of great interest as heritable determinants of cellular phenotypes might offer novel therapeutic targets. Strikingly, only 24% of genomically defined subclones were associated with significant up/downregulation of a single program, suggesting meta-programs reflect, in part, the phenotypes available to a set of related cells and that a focus solely on genomic clones may miss considerable functionally relevant transcriptional diversity. Given that phenotypic plasticity has been associated with treatment resistance and metastasis, identifying tumors or subclones with high meta-program diversity might be of particular importance.

Lastly, Gavish et al. extended their research to non-malignant cell types. Meta-programs derived from normal epithelial cells exhibited similarities to those observed in the malignant population. This suggests much of the transcriptomic ITH seen within malignant cells is already present within the healthy tissue of origin, and that differences observed in related meta-programs likely reflect the oncogenic switch from healthy to malignant states. This transition often involved aberrant coupling or decoupling of related biological pathways, which may highlight specific, non-genetically encoded dependencies and/or vulnerabilities within an evolving tumor.

Another significant form of variation that sculpts cellular phenotypes is the TME. Within this complex environment, Gavish et al. identify significant co-correlations between meta-programs. Positive correlations suggest the presence of common responses to shared regulatory or environmental pressures, such as those that control angiogenesis or influence responses to interferon alpha. Understanding these cellular networks is of vital importance. For example, combination therapies that target both tumor-specific drivers of angiogenesis as well as normal cellular responses within the TME might enhance treatment efficacy.

In conclusion, Gavish et al. have generated a vast compendium of scRNA-seq data and outlined recurrent patterns of expression that highlight key cellular pathways active within malignant and non-malignant cells. However, as the meta-programs are summarized by 50 representative genes, more granular expression signatures consisting of fewer genes are likely to have been overlooked. Additionally, protein diversity, which ultimately defines cellular phenotypes, may not be concordant with that observed at the level of mRNA.[8]

This work also opens up several avenues for future research. First, continuous integration of published datasets into 3CA will improve sensitivity to discover rarer meta-programs, perhaps containing fewer genes, that might further explain some of

the 34% of tumor-specific expression programs that were not attributable to the hallmarks of transcriptional ITH. Second, the determinants of expression ITH outlined in this study could be further refined. Multiomic technologies that are able to capture both genetic and epigenetic alterations alongside cellular states in single cells might illuminate key targetable drivers of these clinically relevant programs. Third, recent work has revealed patterns of cell-state transitions and heritable gene-expression programs utilizing highly detailed, phenotypically annotated tumor phylogenies.[9] Defining the heritability of meta-programs and understanding their transition rates could provide valuable insights for future treatments. Heritable meta-programs might be potential targets for therapies directed at specific cell states, and combination therapies that utilize modulation of transition rates toward a more differentiated and stable cell state could increase treatment efficacy. Finally, integration of more scRNA-seq datasets with well-curated clinical data will likely reveal the clinical importance of the meta-programs over the course of tumor evolution.

Overall, findings from this study inform our understanding of the extensive transcriptomic heterogeneity present within primary tumors, and provide insight into several common, tumor-specific patterns of transcriptional ITH.

ACKNOWLEDGMENTS

T.P.J receives funding from Cancer Research UK. N.M. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (grant number 211179/Z/18/Z), and also receives funding from Cancer Research UK Lung Cancer Centre of Excellence, Rosetrees, and the NIHR BRC at University College London Hospitals.

DECLARATION OF INTERESTS

N.M. has received consultancy fees and has stock options in Achilles Therapeutics. N.M. holds European patents relating to targeting neoantigens (PCT/EP2016/059,401), identifying patient response to immune checkpoint blockade (PCT/EP2016/071,471), determining HLA LOH (PCT/GB2018/052,004), and predicting survival rates of patients with cancer (PCT/GB2020/050,221).

REFERENCES

1. Black, J.R.M., and McGranahan, N. (2021). Genetic and non-genetic clonal diversity in cancer evolution. Nat. Rev. Cancer *21*, 379–392.

2. Frankell, A.M., Dietzen, M., Al Bakir, M., Lim, E.L., Karasaki, T., Ward, S., Veeriah, S., Colliver, E., Huebner, A., Bunkum, A., et al. (2023). The evolution of lung cancer and impact of subclonal selection in TRACERx. Nature *616*, 525–533.

3. Marjanovic, N.D., Hofree, M., Chan, J.E., Canner, D., Wu, K., Trakala, M., Hartmann, G.G., Smith, O.C., Kim, J.Y., Evans, K.V., et al. (2020). Emergence of a high-plasticity cell state during lung cancer evolution. Cancer Cell *38*, 229–246.e13.

4. Quinn, J.J., Jones, M.G., Okimoto, R.A., Nanjo, S., Chan, M.M., Yosef, N., Bivona, T.G., and Weissman, J.S. (2021). Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science *371*, eabc1944.

5. 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.

6. 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.

7. 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.

8. Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

9. Schiffman, J.S., D'Avino, A.R., Prieto, T., Potenski, C., Fan, Y., Hara, T., Suvà, M.L., and Landau, D.A. (2023). Defining Ancestry, Heritability and Plasticity of Cellular Phenotypes in Somatic Evolution. Preprint at bioRxiv.