

Leveraging Single-Cell RNA Sequencing Experiments to Model Intratumor Heterogeneity

Meghan C. Ferrall-Fairbanks, PhD¹; Markus Ball, PhD¹; Eric Padron, MD¹; and Philipp M. Altrock, PhD¹

PURPOSE Many cancers can be treated with targeted therapy. Almost inevitably, tumors develop resistance to targeted therapy, either from pre-existence or by evolving new genotypes and traits. Intratumor heterogeneity serves as a reservoir for resistance, which often occurs as a result of the selection of minor cellular subclones. On the level of gene expression, clonal heterogeneity can only be revealed using high-dimensional single-cell methods. We propose using a general diversity index (GDI) to quantify heterogeneity on multiple scales and relate it to disease evolution.

MATERIALS AND METHODS We focused on individual patient samples that were probed with single-cell RNA (scRNA) sequencing to describe heterogeneity. We developed a pipeline to analyze single-cell data via sample normalization, clustering, and mathematical interpretation using a generalized diversity measure, as well as to exemplify the utility of this platform using single-cell data.

RESULTS We focused on three sources of patient scRNA sequencing data: two healthy bone marrow (BM) donors, two patients with acute myeloid leukemia—each sampled before and after BM transplantation, four samples of presorted lineages—and six patients with lung carcinoma with multiregion sampling. While healthy/normal samples scored low in diversity overall, GDI further quantified the ways in which these samples differed. Whereas a widely used Shannon diversity index sometimes reveals fewer differences, GDI exhibits differences in the number of potential key drivers or clonal richness. Comparison of pre- and post-BM transplantation acute myeloid leukemia samples did not reveal differences in heterogeneity, although biological differences can exist.

CONCLUSION GDI can quantify cellular heterogeneity changes across a wide spectrum, even when standard measures, such as the Shannon index, do not. Our approach can be widely applied to quantify heterogeneity across samples and conditions.

Clin Cancer Inform. © 2019 by American Society of Clinical Oncology

Licensed under the Creative Commons Attribution 4.0 License 

INTRODUCTION

In many cancers, there still exists a critical need to understand the mechanisms of the evolution of therapy resistance. For example, acute myeloid leukemia (AML) is an aggressive hematologic malignancy the hallmark of which is the proliferation of immature myeloid cells in the bone marrow and life-threatening ineffective hematopoiesis.¹ AML is the most common adult leukemia, with an incidence of approximately 20,000 cases per year and a 5-year survival of only 26%.^{2,3} Diagnosis of AML requires greater than 20% of myeloid immature cells (myeloblasts) in peripheral blood or bone marrow. Median survival of untreated AML is measured in weeks.⁴ Several AML targeted therapies have been recently approved—for example, midostaurin for patients with FLT3 mutated disease and enasidenib for those with mutations in IDH2.^{5,6} These mutations occur at rates of 25% (FLT3) and 5%

(IDH2) of all patients with AML and their targeted therapies are generally well tolerated compared with their chemotherapeutic counterparts.⁷ However, midostaurin—and even more potent FLT3 inhibitors in clinical trial⁸—does not fully eradicate disease, which leads to refractory or relapsed AML in most patients.⁹ Complete response rate for enasidenib in relapse/refractory IDH2 mutated AML is less than 20%. Additional refinements in patient selection are required to realize mutationally directed therapy.⁵ Little is known regarding the emerging resistance mechanism and whether targeted therapies—single or combination—against AML alone can ever be successful.

Conventional dogma postulates that therapeutic resistance occurs via the acquisition of mutations that result in clonal evolution. Emerging data suggest that these mutations are either subclonally present or present at frequencies detectable using digital

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on October 23, 2018 and published at ascopubs.org/journal/cci on April 17, 2019; DOI <https://doi.org/10.1200/CCI.18.00074>

polymerase chain reaction or ultradeep sequencing technologies at diagnosis or before progression. Low-level somatic mutations are also detected in preleukemic states.¹⁰⁻¹³ Somatic mutations are often present years before the diagnosis of therapy-related myeloid neoplasms.^{14,15} Of interest, these mutations are commonly associated with disease progression and transformation.¹⁶ The presence of such low-frequency genetic markers suggests that high levels of intratumor heterogeneity (ITH) persist over long periods of time and that preexisting ITH is a primary driver of future therapy resistance, whereas variation in transcription over time shapes the disease phenotype. A clinically relevant summary metric by which to describe ITH on the transcriptional level has not been developed.

Single-cell RNA (scRNA) sequencing technologies can present a cost-effective method with which to identify transcriptomic heterogeneity and directly measure ITH. Proof-of-concept studies have been performed in AML using DROP sequencing that yields potentially cost-effective single-cell annotations of thousands of transcripts per cell.¹⁷ In triple-negative breast cancer, intercellular heterogeneity of gene expression programs within tumors is variable and correlates with genomic clonality.¹⁸ A study in chronic myeloid leukemia demonstrated that scRNA sequencing was capable of segregating patients with discordant responses to targeted tyrosine kinase inhibitor therapy.¹⁹ These data provide a rationale by which to explore ITH in scRNA sequencing data and to determine whether defined measures of ITH can be predictive of progression, eventually leveraging this process to mitigate progression and relapse.

The goal of the current study was to quantify ITH in cancer such that it has maximal predictive value, in particular in hematologic malignancies. To this end, we present a platform that uses a generalized diversity index that characterizes cell population heterogeneity across a spectrum of scales (orders of diversity).²⁰ These scales range from clonal richness (low order of diversity reveals the number of distinct subpopulations), to more classic measures, such as Shannon or Simpson indices (intermediate order of diversity), to the number of most abundant cell types that can possibly act as key drivers of heterogeneity before transformation or perturbation by therapy (high order of diversity).

MATERIALS AND METHODS

We created a computational and modeling approach to develop a robust statistical picture of the persistent and emerging variability in scRNA sequencing data chiefly on the basis of on DROP sequencing technologies; the 10x Genomics platform offered a variety of data sets linked to disease and treatment dynamics.²¹ We specifically used the data sets of two healthy/control bone marrow mononuclear cell samples (BMMCs) and two individuals with AML BMMCs sampled pre- and post-bone marrow transplantation (BMT) to develop and test our ITH pipeline.

First, we ran publicly available FASTQ-format files—a typical output from a DROP sequencing experiment—through the cellranger count pipeline and then through the Cell Ranger aggr pipeline to pool the samples together for comparison during cluster analysis, interrogated through the 10x Genomics Loupe Cell Browser (Data Supplement). To test the robustness and validity of our diversity metrics and the ITH pipeline, we extended our analysis to include additional publicly available data sets for other hematopoietic cell types (CD34⁺, CD14⁺, CD19⁺, and CD4⁺),²¹ as well as matched normal-tumor lung cancer samples from six patients,²² for which we used the same approaches and pipelines. To calculate summary metrics—outlined in Figure 1—first the transcript expression data were clustered into groups of cells with similar transcript expressions (Cell Ranger aggr). We next quantified the distance between each of the clusters to determine if clusters separated on the basis of healthy or disease status (healthy *v* AML). A Euclidean distance was calculated between mean expression values for each gene of each cluster to establish a distance metric (Fig 2).

Second, we sought to characterize across-sample differences by calculating the Kolmogorov-Smirnov (KS) distance²³ of the cell count distributions in each cluster. This was done to compare samples or pooled samples of the same condition—for example, disease versus healthy—in terms of the cellular distribution over the identified clusters (Figs 3A-3G).

Third, we calculated an ecological diversity index²⁴ using the cellular frequencies over clusters across a range of order of diversity (Figs 3H and 3I). To assess the robustness of our diversity metric, we performed downsampling of the original data sets and found the relative change in diversity index across a range of order of diversity to determine the sensitivity of our diversity metric (Fig 4).

Last, we applied our ITH pipeline and diversity metric to two additional data sets—a hematopoietic cell-type data set that compared CD34⁺ cells with CD4⁺, CD14⁺, and CD19⁺ cell populations,²¹ as well as a lung cancer data set with matched normal-tumor tissue sites taken from six different patients with lung cancer²² (Fig 5 and Data Supplement). Additional specific details of our methods, such as cells per sample, are described in the Data Supplement and are available online (including all code used to generate our results).²⁵

RESULTS AND DISCUSSION

We established a proof of concept that we can generate clinically relevant summary metrics of ITH by analyzing publicly available scRNA sequencing data.^{21,22} Within BMMC samples from diagnosed AML and healthy control groups, we sought to establish how to summarize both inter- and intrasample ITH. First, we clustered the transcript expression of two healthy individuals and two patients with AML, each sampled twice, once before and once after allogeneic BMT.

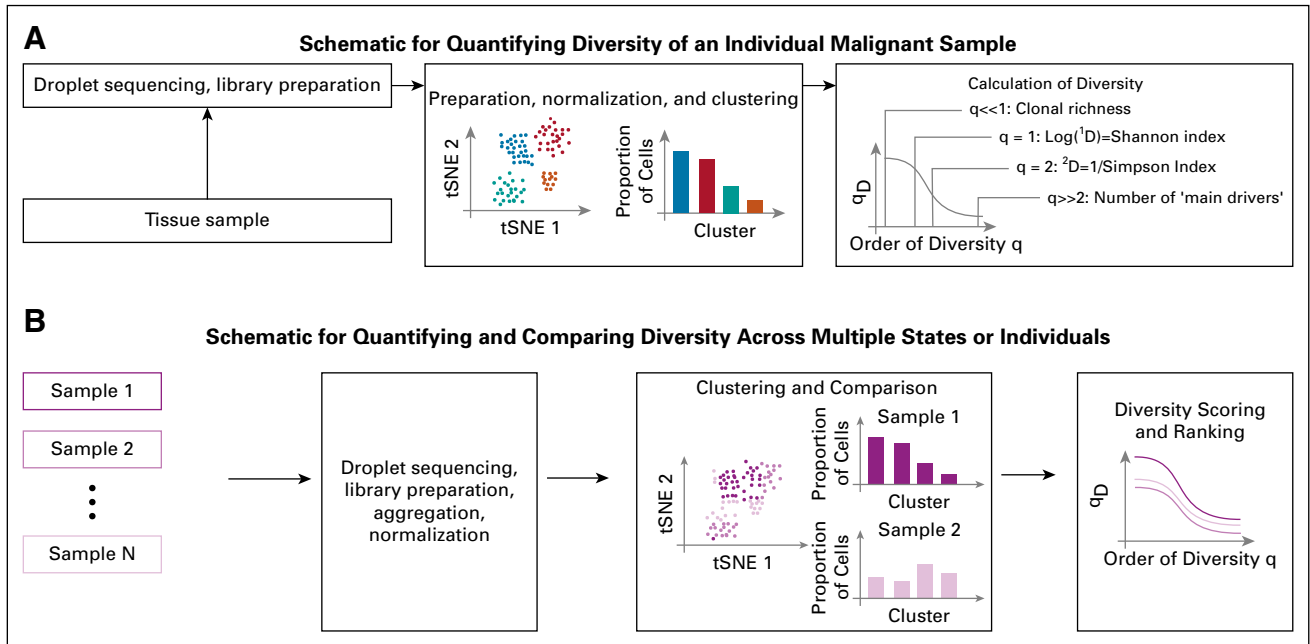


FIG 1. Schematic of our single-cell RNA sequencing–based approach. (A) Workflow for calculating a generalized diversity index for a single sample. After sequencing and library preparation, normalization to reduce the number of false negatives or false positives is applied—for example, using the 10x Genomics platform. Clustering can then be applied (Loupe Cell browser or other platforms; Data Supplement), from which we can calculate diversity. (B) A similar approach can be used when multiple samples are compared. Data normalization and clustering now have to be implemented considering all samples (Data Supplement), and diversity scoring can inform a ranking of intratumor heterogeneity across samples. Single dots in the t-distributed stochastic neighbor embedding (tSNE) plots represent single cells, which might either be marked according to their cluster classification or according to their sample of origin.

As verification, we sought to distinguish between healthy and AML samples on the basis of the mean expression values across cells and across clusters (Fig 2A). With 23 clusters identified (Figs 2B and 2C), a network of clusters emerged, displayed as an undirected graph where the distance between mean unique molecular identifier counts determines the thickness and length of the edges (Fig 2D). The size of the node was chosen to indicate the total number of cells in the cluster. We colored each node according to the condition—health, pre-BMT AML, or post-BMT AML—that was in the majority (breakdown of actual proportions per cluster shown in Fig 2E). This indicated that the large clusters with mostly healthy cells are most similar in average gene expression, whereas the large clusters with mostly AML cells cluster separately (to the right). Post-BMT cells clustered more closely to the healthy samples than to pre-BMT AML samples. This result supports the idea that these patients were potentially still transitioning but closer to a healthy phenotype; however, some AML-dominant clusters still grouped near the healthy/post-BMT super cluster. On the basis of this bulk measure alone, one may not be able to easily distinguish between healthy and diseased cells. Therefore, other quantifications and metrics to describe gene expression differences may better discriminate between patients with different clinical presentations and staging.

To determine metrics that are better at discriminating between healthy and disease AML, we analyzed and summarized inter- and intraheterogeneity in two different ways. First, we considered the grouping of cells into clusters of similar gene expression in each sample. To this end, we used KS distance, which compares two discrete probability mass functions (the fraction of cells per cluster; Fig 3A). We identified similar distributions within the same condition and different distributions between conditions, with post-BMT being a notable exception (Figs 3B–3D). KS distance between the two healthy samples was 0.139 and 0.174 between the two pre-BMT AML samples, but between the two post-BMT samples it was 0.551. As we had clustered all six samples together, we could also compare them pooled by condition, which revealed that conditions distribute differently across the identified cellular subpopulations in high-dimensional gene expression space (Figs 3E–3G).

Second, we calculated a general diversity index (GDI) for each condition (healthy, pre-BMT AML, and post-BMT AML). The mathematical definition of GDI, q_D , is shown in Figure 3H. We established segregation of the different clinical conditions according to this ecology-based diversity index.²⁴ Pre-BMT AML samples had a consistently higher diversity index compared with the healthy sample. This held true across the entire order of diversity range, q (Fig 3I). Of interest, on this level, post-BMT samples also

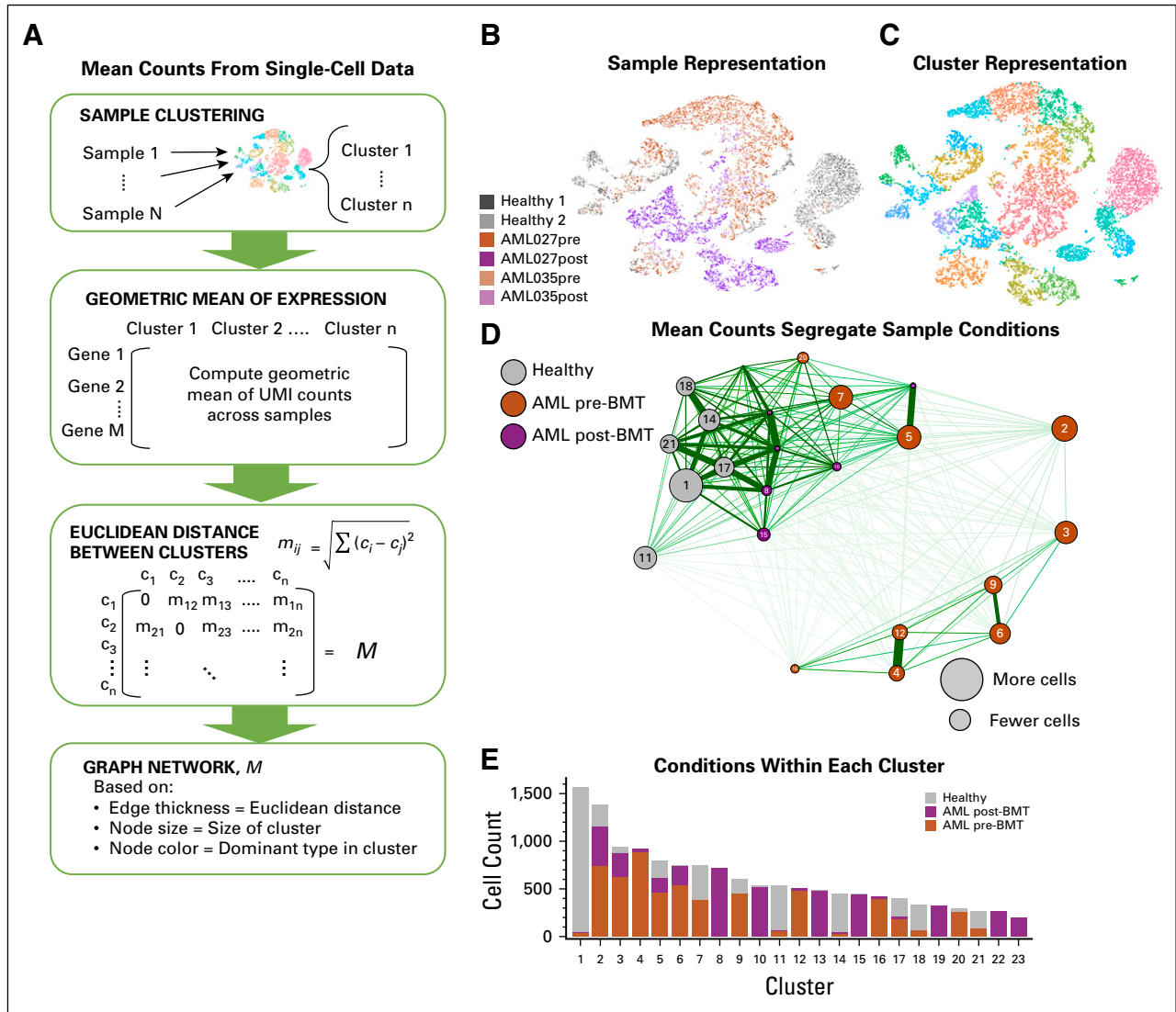
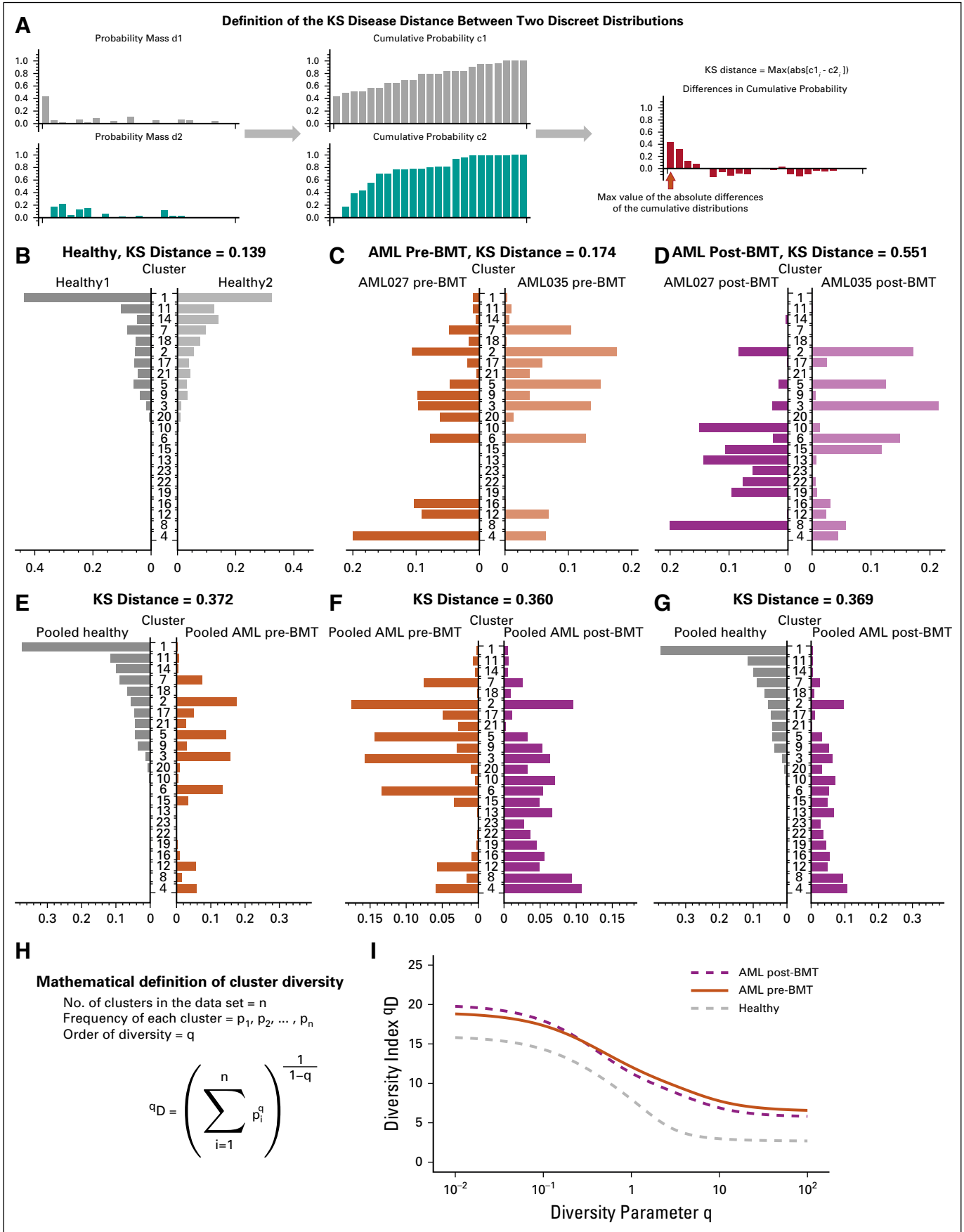


FIG 2. Mean cellular gene expression across clusters within patients can separate disease conditions to some degree. Here, we built a network on the basis of mean differences in overall expression. (A) We calculated the geometric mean of unique molecular identifier (UMI) counts across samples and genes for each cluster. Then, a Euclidean distance was calculated between clusters. Here, we used publicly available single-cell RNA sequencing data²¹: two healthy donor bone marrow mononuclear cell samples (BMMCs) and BMMCs from two patients with acute myeloid leukemia (AML) pre–bone marrow transplant (BMT) and post-BMT. These six samples were then clustered using 10x Genomics Loupe Cell browser (for alternative clustering methods see the Data Supplement), for which we show the (B) sample-based and (C) cluster-based t-distributed stochastic neighbor embedding (tSNE) plots from the Loupe browser. Each dot represents a single cell, which is colored either according to sample of origin or its assigned cluster. (D) Cluster-based differences in mean gene expression over UMI counts gave rise to a clustering of the clusters. Nodes in the resulting graph were colored on the basis of the dominant cell type from each condition present in each cluster (gray for healthy, orange for AML pre-BMT, and purple for AML post-BMT), and the distance between nodes was chosen inversely proportional to the difference in mean gene expression level. (E) Individual distributions of cells from a specific condition in each cluster are shown.

scored unanimously higher in GDI. This could indicate that post-BMT settings may require a certain amount of time after transplantation to evolve toward a healthy spectrum of intraleukemic diversity. In addition, in a comparison of the individual samples within each condition (Figs 4A-4C), post-BMT samples were most different from each other.

To interrogate the robustness of GDI further and to establish confidence in the metric, we downsampled the data set, then reclustered and calculated the ^qD spectrum (Figs 4D-4I).

During downsampling, we analyzed each sample individually by randomly removing 50% of the cells, then calculating the number of clusters identified for that individual's transcript expression, and finally calculating the diversity index for specific q values of interest, including $q = 10^{-2}$; $q = 10^{-1}$; $q = 1$, which relates to the Shannon index; $q = 2$, which defines the inverse of the Simpson index; $q = 10$; and $q = 10^2$. Distributions shown were obtained from 1,000 runs of independent downsampling.



Intriguingly, these distributions showed that with removing one half of the cells, diversity scores did not change more than one or two units in either direction. Compared with the diversity spectrum shown in [Figure 3I](#), this suggests that if healthy diversity spectra were shifted up by two units (10% of the maximum) and AML samples diversity spectra were shifted down by two units, there would still be visible separation between the healthy and AML conditions.

To further validate our metric, we implemented on our approach with two other data sets. One data set described different hematopoietic cellular subtypes—CD34⁺, CD4⁺, CD14⁺, and CD19⁺.²¹ CD34⁺ is a hematopoietic progenitor cell marker and represents a polyclonal population that includes many different subtypes—hematopoietic stem cells, multipotent progenitor cells, common myeloid progenitor cells, common lymphoid progenitor cells, megakaryocytes erythroid progenitor cells, and granulocytes macrophage progenitor cells—all of which express CD34.²⁶⁻²⁸ The CD34⁺ polyclonal population contrasts with the CD4⁺, CD14⁺, and CD19⁺ populations, which represent more homogenous cellular populations (helper T cells, monocytes, and B cells, respectively). This clonality pattern was recovered by GDI ([Fig 5A](#)), where the CD34⁺ population had a considerably higher diversity score across the spectrum. Of interest, lower q values seem to separate differentiated cells more robustly.

Finally, we quantified ITH using lung cancer scRNA sequencing samples.²² We analyzed six different patients, each with up to three different tumor sites—core, middle, and edge—and a patient-matched adjacent normal lung tissue sample. Using our GDI metric, we see that the diversity spectrum of the normal lung tissue was much lower than any of the tumor site diversity scores (pooled conditions; [Fig 5B](#)). Of interest, more clear separation of conditions was achieved at high orders of diversity q , which indicated differences in the number of driver clones at different sites within the tumors. These additional data sets further support the ability of a quantified diversity metric to discriminate between healthy and diseases states, which can be applied in a clinical setting.

In conclusion, scRNA sequencing efforts have helped greatly to uncover population structures and mapping to specific cellular population patterns.²⁹ Although these methods can also elucidate tumorigenesis³⁰ and immune profiles,³¹ as well as detect and track genomic profiles of clones,^{32,33} the overall utility of scRNA sequencing for cancer progression survival metrics has been elusive.³⁴

Here, we demonstrate the potential utility of two scRNA sequencing-based scores of cellular heterogeneities using a GDI that may be elevated in disease. Remarkably, using previously published data, without additional processing, our quantification of intratumor heterogeneity was able to accurately distinguish AML from healthy individuals as well as from post-transplantation conditions. These data suggest that ITH can be estimated using diversity-based summary statistics and that these summary statistics can be leveraged to predict clinical outcome.

The aim of the current study was to optimize and identify a clinically relevant summary index for ITH in the context of AML, for which targeted single-cell genome sequencing was also able to sensitively uncover complex clonal evolution.³² We anticipate that our intraleukemic heterogeneity (ILH) metric will be prognostic for leukemia-free survival and potentially overall survival, even after correcting for known clinical prognostic variables. We have also demonstrated how this metric can also be used to effectively describe heterogeneity in other malignancies, including such solid tumors as lung carcinomas.

From a clinical perspective, in terms of tumor heterogeneity and the emergence of resistance clones during targeted therapy,^{35,36} we expect our metric can discriminate patients clinically. We hypothesize that these heterogeneity metrics would be elevated independently—at least a priori—in potentially highly resistant patients. The advantage of the more general metric used here is that it allows us to look across many orders of diversity and potentially pick a desired range of heterogeneity quantification. For example, one might be interested especially in lower q values, where a higher diversity score may indicate an individual (sample) more at risk for resistance evolution as it shows high standing variation. In contrast, differences at high q values point to key differences in the number of important driver clones, which might uncover distinct vulnerabilities that can be targeted in combination or adaptively.

Diversity measures have long received attention in ecology and evolution.^{20,37} Here, we measured diversity—and thus tumor heterogeneity—using a general definition of non-spatial diversity³⁸ in the form of qD quantity ([Figs 3H-3I](#)). This approach considers all possible orders of diversity q , but also allows for the comparison of disease stages according to a specific diversity index (fixed choice of q), which emerge as special cases of qD . Species (clonal) richness of a sample is given by $q = 0$. The Shannon index (log scale) can be found when q approaches 1. The

FIG 3. Cluster-based diversity scoring reveals strong differences between healthy individuals and patients with cancer. In our analysis, using data from Zheng et al,²¹ we evaluated our ability to score significant differences in cluster diversity across healthy and acute myeloid leukemia (AML) samples pre- and post-bone marrow transplant (BMT). (A) As a first indicator of between-sample or between-condition differences, we used the Kolmogorov-Smirnov (KS) distance for discrete probability mass functions. (B and C) Little difference was found in the KS distance within condition differences, except in (D) post-BMT samples. (E-G) Between-condition differences were larger when comparing pooled samples across conditions. (H) We calculated a general diversity index (GDI) qD to quantify diversity across orders of diversity q . (I) For all orders of diversity measure, patients with AML (pre- and post-BMT) had a higher diversity index compared with healthy individuals (two samples per condition), which suggests that GDI can be used as a metric for stratification.

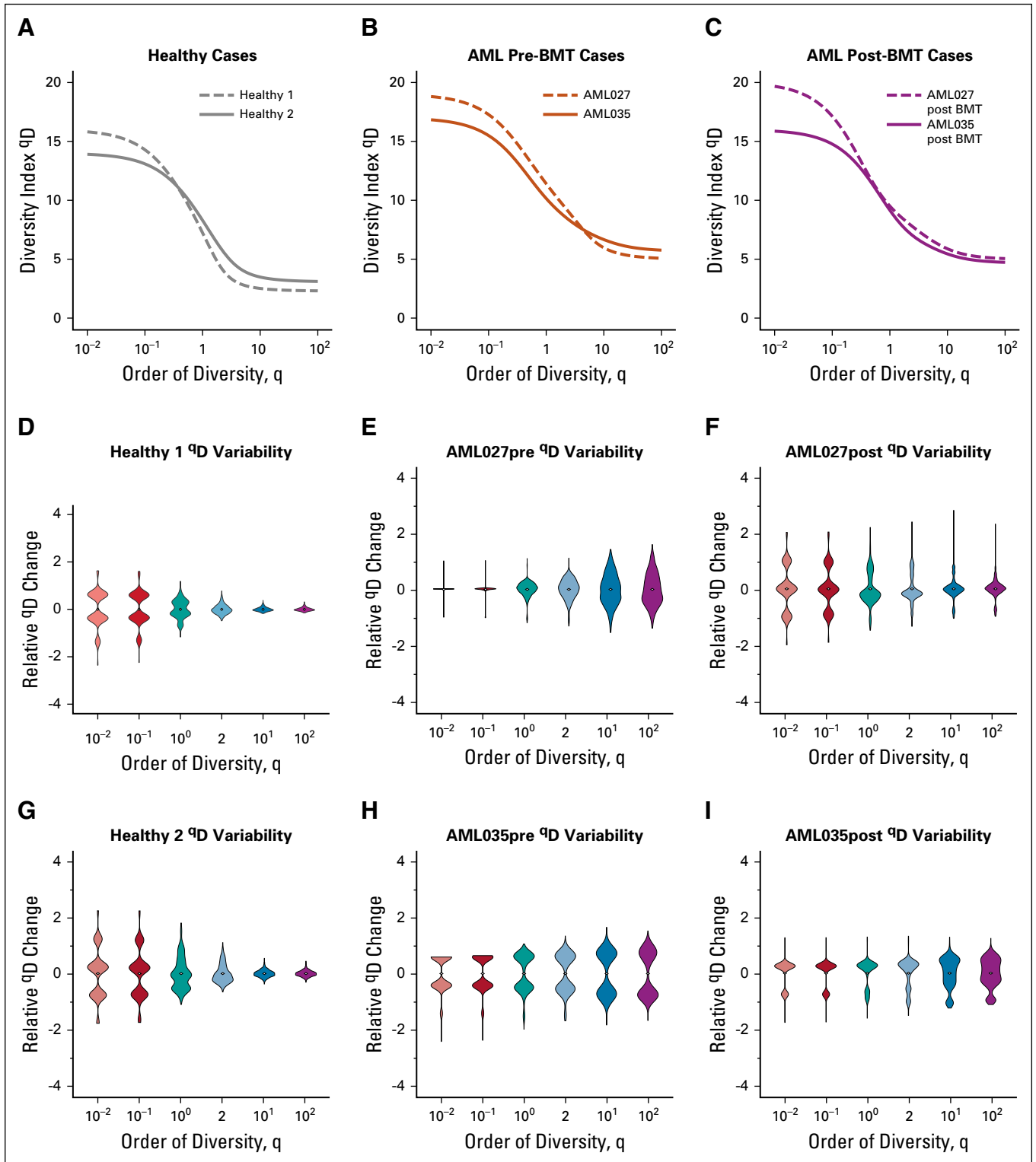


FIG 4. Patients with acute myeloid leukemia (AML) have consistently higher diversity compared with healthy individuals. Individual diversity spectrums were reported for (A) healthy, (B) AML pre-bone marrow transplantation (BMT), and (C) AML post-BMT samples (each line is from one sample). Cell-gene matrices were downsampled to 50% of the cells 1,000 times, and qD scores were calculated using the full pipeline (see Data Supplement) for specific values of $q = 0.01, 0.1, 1, 2, 10,$ and 100 . Distributions of relative qD changes for (D and G) healthy, (E and H) AML, and (F and I) post-BMT AML samples showed that, in general, lower q values lead to less change in measured diversity. Across all cases, the diversity score did not change by more than two units (relative change is measured by dividing the entire distribution by the distribution mean). For sample sizes, see the Data Supplement.

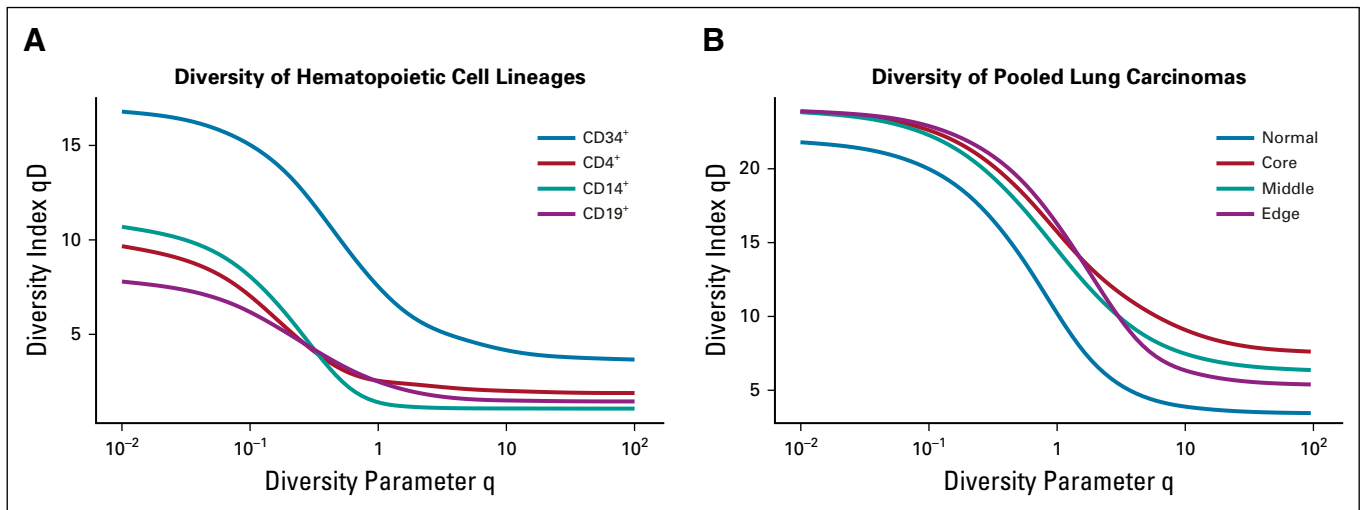


FIG 5. Higher diversity indicates higher clonality in normal tissues and solid tumors. (A) Additional available data from Zheng et al²¹ for CD34⁺ cells, CD4⁺ helper T cells, CD14⁺ monocytes, and CD19⁺ B cells were run through pipeline S (Data Supplement), and the continuum of diversity was calculated for each population. The naturally polyclonal population (CD34⁺) shows the highest diversity score. Each of the other differentiated immune cell compartments are more homogeneous across orders of diversity. (B) In solid tumors, location matters. Normal tumor matched lung carcinoma samples were obtained from publicly available data for six patients with lung cancer²² (individual patients; Data Supplement). The diversity metric across q demonstrates an increase in diversity within tumors across different tumor locations.

Simpson index, which approximates the probability that any two cells are identical, emerges from the case of $q = 2$. Both the Shannon and Simpson indexes have been used in mathematical and statistical models of cancer evolutionary dynamics to quantify tumor heterogeneity as it potentially changes during tumor growth, with disease progression, or during treatment.³⁹⁻⁴¹ Shannon entropy-based statistics have also been used to quantify single-cell heterogeneity to deliver insights into emerging or disappearing clones during transitions between clinical conditions.⁴²

scRNA sequencing experiments provide a snapshot of the cell population on the level of gene expression and can characterize how transcriptomes of individual cells compare with the bulk. In contrast to mass cytometry, DROP sequencing is fast and extremely high throughput. Other single-cell technologies, such as flow cytometry, can be used to generate single-cell data for a relatively small subset of potential markers that distinguish between normal and disease. This requires that the researcher or clinician know what these markers are in advance. Among a variety of outcomes that may be distinguished by our metric, one can be the study for segregating samples on the basis of on disease severity, for which additional follow-up knowledge will be needed.

Further extending our results to the potential impact in a clinical setting in leukemias, ILH is a known reservoir for tumor resistance and clinical refractoriness to targeted therapies.⁹ Clinical responses have been modest to current targeted therapies for the treatment of AML. Specific means to change ILH can be of particular appeal in such cases, as they might help render tumors less aggressive and hinder their ability to rapidly evolve resistance. In particular,

hypomethylating agents—cytosine analogs that irreversibly bind to DNA methyltransferase, an enzyme that is required for methylation of CpG-rich DNA—have the potential to diminish ILH.⁴³ Transcriptome changes upon treatment with hypomethylating agent therapy have not been analyzed at single-cell resolution. Our analyses in the current study provide a quantitative basis by which to understand and reliably track these changes.

Clear separation of diversity metrics by condition, as we show it, might not be expected in general. A weakness of our approach is that it does not consider any meaning of the associated phenotypes or genotypes; therefore, as it stands, our method cannot be transferred to improve the predictive power of existing bulk signatures. Hence, existing survival data are unlikely to be useful to prove that GDI is predictive of survival, and novel databases that uniquely connect high-throughput single-cell experiments with clinical outcomes are needed.

However, once the appropriate cohorts are established, changes in an individual's diversity score could indicate unique features of disease progression. In the context of adaptive therapy,^{44,45} which aims at tumor burden control rather than difficult tumor eradication, it might be critical to identify the appropriate scale of diversity that best predicts outcomes. One could speculate that there is an optimal window of diversity that should be maintained—low diversity could indicate fast disease progression and high diversity could mean that the tumor could adapt to the treatment schedule too quickly. The concept we have introduced here is sufficiently flexible in its ability to quantify optimally predictive windows of diversity that should be maintained during adaptive therapy.

AFFILIATION

¹All authors: Moffitt Cancer Center and Research Institute, Tampa, FL.

Preprint version available on [bioRxiv](https://doi.org/10.1101/2020.03.11.200311).

CORRESPONDING AUTHOR

Philipp M. Altrock, PhD, H. Lee Moffitt Cancer Center and Research Institute, 12902 USF Magnolia Dr, SRB 24007, Tampa, FL 33612; e-mail: philipp.altrock@moffitt.org.

SUPPORT

Supported in part by the Biostatistics and Bioinformatics Shared Resource at the H. Lee Moffitt Cancer Center & Research Institute, a National Cancer Institute–designated Comprehensive Cancer Center (P30-CA076292).

AUTHOR CONTRIBUTIONS

Conception and design: Eric Padron, Philipp M. Altrock

Financial support: Philipp M. Altrock

Administrative support: Philipp M. Altrock

Collection and assembly of data: Meghan C. Ferrall-Fairbanks, Markus Ball

Data analysis and interpretation: All authors

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated.

Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Eric Padron

Honoraria: Incyte, Karyopharm Therapeutics

Research Funding: Incyte (Inst), Cell Therapeutics (Inst)

No other potential conflicts of interest were reported.

REFERENCES

1. Perl AE: The role of targeted therapy in the management of patients with AML. *Hematology (Am Soc Hematol Educ Program)* 2017:54-65, 2017
2. Percival ME, Tao L, Medeiros BC, et al: Improvements in the early death rate among 9380 patients with acute myeloid leukemia after initial therapy: A SEER database analysis. *Cancer* 121:2004-2012, 2015
3. Medeiros BC, Satram-Hoang S, Hurst D, et al: Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States. *Ann Hematol* 94:1127-1138, 2015
4. Arber DA, Orazi A, Hasserjian R, et al: The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127:2391-2405, 2016
5. Stein EM, DiNardo CD, Pollyea DA, et al: Enasidenib in mutant *IDH2* relapsed or refractory acute myeloid leukemia. *Blood* 130:722-731, 2017
6. Stone RM, Mandrekar SJ, Sanford BL, et al: Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *N Engl J Med* 377:454-464, 2017
7. Papaemmanuil E, Gerstung M, Bullinger L, et al: Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* 374:2209-2221, 2016
8. Cooper TM, Cassar J, Eckroth E, et al: A phase I study of quizartinib combined with chemotherapy in relapsed childhood leukemia: A Therapeutic Advances in Childhood Leukemia & Lymphoma (TACL) study. *Clin Cancer Res* 22:4014-4022, 2016
9. Smith CC, Paguirigan A, Jeschke GR, et al: Heterogeneous resistance to quizartinib in acute myeloid leukemia revealed by single-cell analysis. *Blood* 130:48-58, 2017
10. Paguirigan AL, Smith J, Meshinchi S, et al: Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia. *Sci Transl Med* 7:281re2, 2015
11. Corces-Zimmerman MR, Hong WJ, Weissman IL, et al: Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc Natl Acad Sci USA* 111:2548-2553, 2014
12. Ding L, Ley TJ, Larson DE, et al: Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481:506-510, 2012
13. Padron E, Yoder S, Kunigal S, et al: ETV6 and signaling gene mutations are associated with secondary transformation of myelodysplastic syndromes to chronic myelomonocytic leukemia. *Blood* 123:3675-3677, 2014
14. Wong TN, Ramsingh G, Young AL, et al: Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* 518:552-555, 2015
15. Gillis NK, Ball M, Zhang Q, et al: Clonal haemopoiesis and therapy-related myeloid malignancies in elderly patients: A proof-of-concept, case-control study. *Lancet Oncol* 18:112-121, 2017
16. Zuffa E, Franchini E, Papayannidis C, et al: Revealing very small FLT3 ITD mutated clones by ultra-deep sequencing analysis has important clinical implications in AML patients. *Oncotarget* 6:31284-31294, 2015
17. Macosko EZ, Basu A, Satija R, et al: Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202-1204, 2015
18. Karaayvaz M, Cristea S, Gillespie SM, et al: Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat Commun* 9:3588, 2018
19. Giustacchini A, Thongjuea S, Barkas N, et al: Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* 23:692-702, 2017
20. Hill MO: Diversity and evenness: A unifying notation and its consequences. *Ecology* 54:427-432, 1973
21. Zheng GX, Terry JM, Belgrader P, et al: Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8:14049, 2017
22. Lambrechts D, Wauters E, Boeckx B, et al: Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* 24:1277-1289, 2018
23. Carruth J, Tygert M, Ward R: A comparison of the discrete Kolmogorov-Smirnov statistic and the Euclidean distance. *arXiv1206.6367:1-15*, 2012
24. Lou J: Entropy and diversity. *Oikos* 113:363-375, 2006
25. GitHub: <https://github.com/MathOnco/scRNAseqIHT>
26. Greaves MF, Brown J, Molgaard HV, et al: Molecular features of CD34: A hemopoietic progenitor cell-associated molecule. *Leukemia* 6:31-36, 1992 (suppl 1)
27. Krause DS, Fackler MJ, Civin CI, et al: CD34: Structure, biology, and clinical utility. *Blood* 87:1-13, 1996

28. Zhang Y, Gao S, Xia J, et al: Hematopoietic hierarchy: An updated roadmap. *Trends Cell Biol*[epub ahead of print on June 20, 2018]
29. Trapnell C, Cacchiarelli D, Grimsby J, et al: The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32:381-386, 2014
30. Ellsworth DL, Blackburn HL, Shriver CD, et al: Single-cell sequencing and tumorigenesis: Improved understanding of tumor evolution and metastasis. *Clin Transl Med* 6:15, 2017
31. Schelker M, Feau S, Du J, et al: Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat Commun* 8:2032, 2017
32. Pellegrino M, Sciambi A, Treusch S, et al: High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res* 28:1345-1352, 2018
33. Navin NE: The first five years of single-cell cancer genomics and beyond. *Genome Res* 25:1499-1507, 2015
34. Wang L, Livak KJ, Wu CJ: High-dimension single-cell analysis applied to cancer. *Mol Aspects Med* 59:70-84, 2018
35. Gatenby RA, Silva AS, Gillies RJ, et al: Adaptive therapy. *Cancer Res* 69:4894-4903, 2009
36. Stanková K, Brown JS, Dalton WS, et al: Optimizing cancer treatment using game theory: A review. *JAMA Oncol*[epub ahead of print on August 9, 2018]
37. MacArthur RH: Patterns of species diversity. *Biol Rev Camb Philos Soc* 40:510-533, 1965
38. Tuomisto H: A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecologia* 164:853-860, 2010
39. Almendro V, Cheng YK, Randles A, et al: Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Reports* 6:514-527, 2014
40. Marusyk A, Tabassum DP, Altmann PM, et al: Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* 514:54-58, 2014
41. Altmann PM, Liu LL, Michor F: The mathematics of cancer: Integrating quantitative models. *Nat Rev Cancer* 15:730-745, 2015
42. Li J, Smalley I, Schell MJ, et al: SinCHet: A MATLAB toolbox for single cell heterogeneity analysis in cancer. *Bioinformatics* 33:2951-2953, 2017
43. Kaminskas E, Farrell AT, Wang YC, et al: FDA drug approval summary: Azacitidine (5-azacytidine, Vidaza) for injectable suspension. *Oncologist* 10:176-182, 2005
44. Zhang J, Cunningham JJ, Brown JS, et al: Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nat Commun* 8:1816, 2017
45. Gallaher JA, Enriquez-Navas PM, Luddy KA, et al: Spatial heterogeneity and evolutionary dynamics modulate time to recurrence in continuous and adaptive cancer therapies. *Cancer Res* 78:2127-2139, 2018

