

Dissecting tumor cell programs through group biology estimation in clinical single-cell transcriptomics

Received: 8 December 2021

Accepted: 20 February 2025

Published online: 01 March 2025

 Check for updates

Shreya Johri^{1,2}, Kevin Bi^{1,2}, Breanna M. Titchen^{1,2,3}, Jingxin Fu^{1,2}, Jake Conway^{1,2,4}, Jett P. Crowdis^{1,2}, Natalie I. Vokes^{5,6}, Zenghua Fan^{7,8,9}, Lawrence Fong^{7,8,9}, Jihye Park^{1,2}, David Liu^{1,2}, Meng Xiao He^{1,2} & Eliezer M. Van Allen^{1,2}✉

With the growth of clinical cancer single-cell RNA sequencing studies, robust differential expression methods for case/control analyses (e.g., treatment responders vs. non-responders) using gene signatures are pivotal to nominate hypotheses for further investigation. However, many commonly used methods produce a large number of false positives, do not adequately represent the patient-specific hierarchical structure of clinical single-cell RNA sequencing data, or account for sample-driven confounders. Here, we present a non-parametric statistical method, BEANIE, for differential expression of gene signatures between clinically relevant groups that addresses these issues. We demonstrate its use in simulated and real-world clinical datasets in breast cancer, lung cancer and melanoma. BEANIE outperforms existing methods in specificity while maintaining sensitivity, as demonstrated in simulations. Overall, BEANIE provides a methodological strategy to inform biological insights into unique and shared differentially expressed gene signatures across different tumor states, with utility in single-study, meta-analysis, and cross-validation across cell types.

Single-cell transcriptomic profiling of patient tumors has enabled the high-resolution study of cell type specific transcriptional programs dictating disease progression and treatment response. Building on seminal cellular atlases for specific cancer types, single-cell RNA sequencing (scRNA-seq) studies are increasingly focused on deriving hypotheses by evaluating groups of patients (e.g., treated vs. untreated, responders vs. non-responders, and early- vs. late-stage). Numerous early studies focused on deriving these hypotheses through differential expression of individual genes, inspiring the development of custom bioinformatic methods for scRNA-seq datasets^{1–4}. However,

individual genes alone offer limited insight into the complex cellular programs underlying distinct biological processes. Therefore, studies are increasingly focused on examining differences in gene signatures (experimentally and/or computationally derived aggregation of related genes or pathways) between patient groups. While many methods to extract gene signatures^{5,6} and score cells for gene signatures^{7,8} have been developed, differential expression of these signatures continues to be performed with methods developed for bulk RNA-sequencing (bulk RNA-seq) datasets. These include statistical tests such as the Mann-Whitney U test followed by BH correction (MWU-BH) and

¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ²Broad Institute of Harvard and MIT, Cambridge, MA, USA. ³Harvard Graduate Program in Biological and Biomedical Sciences, Boston, MA, USA. ⁴Harvard Graduate Program in Bioinformatics and Integrative Genomics, Boston, MA, USA. ⁵Department of Thoracic and Head and Neck Oncology, MD Anderson Cancer Center, Houston, TX, USA. ⁶Department of Genomic Medicine, MD Anderson Cancer Center, Houston, TX, USA. ⁷Division of Hematology/Oncology, University of California, San Francisco, CA, USA. ⁸Department of Medicine, University of California, San Francisco, CA, USA. ⁹Parker Institute for Cancer Immunotherapy, University of California, San Francisco, CA, USA.

✉ e-mail: eliezerm_vanallen@dfci.harvard.edu

regression-based methods such as Generalized Linear Models (GLMs). Numerous studies have also applied generic bulk RNA-seq methods after pseudobulking scRNA-seq datasets to the patient level^{9–12}.

These conventional bulk RNA-seq differential expression methods have a number of limitations in the analysis of scRNA-seq data, with particular challenges presented in their application to tumor cells profiled from patient biopsies. First, these methods do not encapsulate the hierarchical structure of tumor scRNA-seq data. This is because tumor cells tend to exhibit more intra-patient similarity as opposed to inter-patient similarity due to the expression of patient-specific transcriptional programs, driven by DNA-level alterations and epigenetics^{13–16}. This challenge, in turn, may lead to differential expression results being skewed by patient-specific biology, instead of representing genuine biological differences collectively representative of clinically stratified patient groups. Furthermore, the number of cells (and hence, data points) sequenced in scRNA-seq datasets are typically orders of magnitude large compared to bulk RNA-seq datasets, thereby potentially increasing the power of statistical tests to detect minute differences (by rejecting the null hypothesis) between the groups under consideration, which may not reflect biologically or clinically relevant observations. This issue may lead to high false positives rates, as echoed in prior works within the context of differential expression using individual genes¹⁷. Notably, methods developed for differential expression of individual genes are not directly applicable to gene signatures either. This is primarily because these methods model specific statistical distributions (e.g., the zero-inflated negative binomial distribution³, hurdle models that combine logistic regression for zero vs. non-zero counts with gaussian linear models for non-zero counts¹). While these parametric distributions are well-suited for modeling the expression of individual genes, the distribution of gene signatures heavily depends on the constituent genes. This dependency can lead to deviations from the assumed parametric distributions, making it challenging to apply standard gene-level differential expression methods directly to gene signatures. As a result of these significant methodological challenges, case/control analyses of scRNA-seq data derived from patient tumor samples often do not involve detailed assessments of the tumor compartments, which has restricted the capability to derive hypotheses around tumor cellular programs in increasingly complex clinical contexts.

In this work, we developed a nonparametric statistical group biology estimation method (group Biology EstimAtion iN sIngle cEll; “BEANIE”) (Fig. 1, “Methods”) for clinical scRNA-seq data, building upon work presented in He et al.¹⁸ Our method automatically incorporates differences arising from patient-specific biology within clinically stratified patient groups, adjusting for disparities in cell counts per patient, and accounting for transcriptomic differences between the two patient groups using a background distribution. Our method is therefore able to identify whether gene signatures characteristic of cell states or biological pathways of interest are differentially expressed between patient groups in a statistically significant and robust way. We applied BEANIE to simulated and patient-derived scRNA-seq datasets across cancer types, and evaluated its performance against six existing methods: MWU-BH, GLM, pseudobulk, and their respective leave-one-out cross-validation (LOOCV) variants (MWU-BH-LOOCV, GLM-LOOCV, and pseudobulk-LOOCV). In simulated datasets, BEANIE showed superior trade-off between specificity and sensitivity in identifying differentially expressed gene signatures compared to the existing methods. Furthermore, when applied to clinical scRNA-seq datasets, BEANIE enabled identification of biologically meaningful gene signatures with high specificity as well as could inform hypotheses in a meta-analysis setting using samples from multiple datasets. Finally, we demonstrated BEANIE’s findings also corroborate across cell type compartments within heterogeneous patient samples derived from different biopsy sites. In summary, BEANIE identifies hypotheses with high specificity that represent transcriptional differences

representative of group biology in both single-study and meta-analysis settings of scRNA-seq data.

Results

BEANIE outperforms conventional methods in identifying differentially expressed gene signatures in simulated datasets

BEANIE is a non-parametric method designed to identify differentially expressed gene signatures between two patient groups using a counts matrix (genes x cells) and a predefined set of gene signatures. The method begins by generating Monte Carlo simulations, ensuring uniform sample representation through subsampling an equal number of cells per sample. For each simulation, *p* values for each gene signature (t_signature) are calculated using the Mann-Whitney U test to compare the two groups. To account for non-biological variability between the patient groups, BEANIE uses a background distribution of gene signatures (b_signatures) composed of randomly sampled genes for correction. This correction is intended to enrich for observed differences that are biologically meaningful. Results for each gene signature are then aggregated across Monte Carlo simulations to derive an empirical *p* value and a robustness metric. The robustness metric assesses how representative the differentially expressed gene signature is of the underlying biology of the patient groups. It is calculated by determining whether the exclusion of a patient sample from the differential testing affects the statistical significance of the empirical *p* values. Gene signatures that are both statistically significant and robust to sample exclusion are identified as differentially expressed gene signatures (Fig. 1; “Methods”). BEANIE is particularly intended for identifying differentially expressed gene signatures within heterogeneous tumor compartments and integrates with scRNA-seq analysis workflows. After cell type identification and tumor state delineation, BEANIE can be used to identify gene signatures that are differentially expressed between the two groups comprising the predefined tumor state. While BEANIE can also be applied across the entire tumor compartment without predefined tumor states, using it in combination with tumor state delineation is intended to enable a finer granularity of results and thus yield more nuanced biological insights.

We benchmarked BEANIE’s ability to detect differentially expressed gene signatures against six conventional methods in the field: MWU-BH, MWU-BH-LOOCV, GLM, GLM-LOOCV, pseudobulk, and pseudobulk-LOOCV (see Methods). For this comparison, we used a simulated tumor scRNA-seq dataset and selected 50 oncology-related gene signatures from MSigDB (see Data Availability). To mimic real-world clinical data, we generated a dataset comprising two comparison groups, each with 10 patients and 100 cells per patient. Both groups were derived from a common seed population of tumor cells from multiple patients, ensuring an identical initial distribution (see Methods). We then introduced controlled perturbations along two axes: magnitude (1, 2, and 3 standard deviations, with 3 std representing outliers) and fraction (ranging from 0.1 to 1). These perturbations represent various scenarios that can lead to differences between patient groups. For each perturbation condition, we conducted 1000 trials, enabling precise estimation of method specificity and sensitivity across different levels of inter- and intra-patient variation typical in clinical datasets (Fig. 2a; Methods).

The evaluated methods demonstrated a wide range of specificity (Fig. 2b, Supplementary Table 1); however, BEANIE generally demonstrated superior performance, with a high mean specificity of 0.999 that significantly outperformed other methods. Specifically, it exhibited 6.60% higher specificity than MWU-BH (mean specificity = 0.933), 5.81% higher than MWU-BH-LOOCV (mean specificity = 0.941), 13.71% higher than GLM (mean specificity = 0.862), and 12.11% higher than GLM-LOOCV (mean specificity = 0.878), with all comparisons yielding adjusted *p* values less than 0.0001 (see “Methods”). BEANIE’s specificity was comparable to that of pseudobulk methods. Furthermore, when 50% or more of simulated patients were perturbed with varying

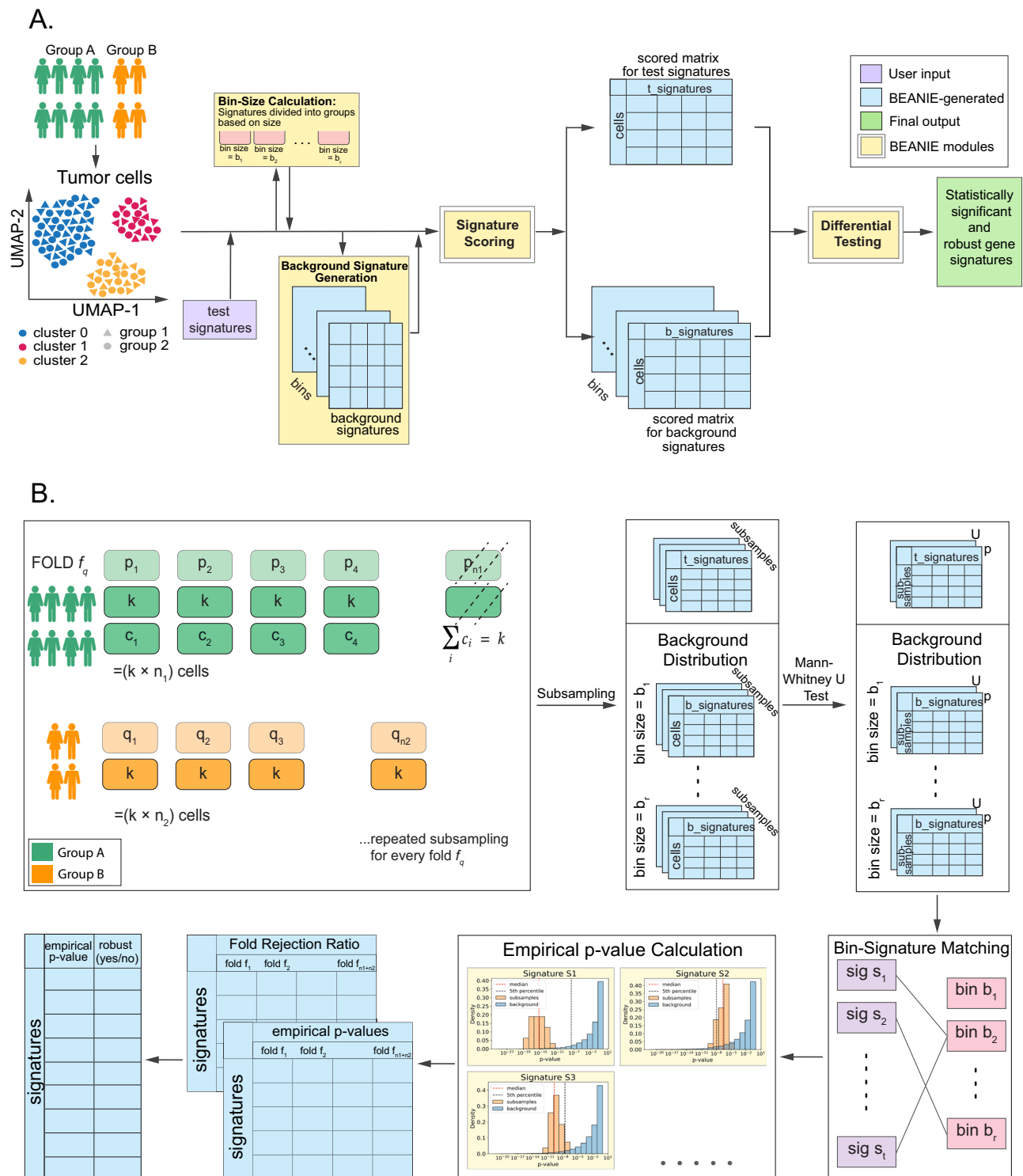


Fig. 1 | BEANIE (group Biology EstAtionN in single cell). A Method Overview: Tumor cells from multiple patient samples are clustered to identify tumor states. BEANIE then focuses on shared states between the two patient groups. For each corresponding tumor state, user inputs include a gene-by-cell count matrix, sample and group IDs, and a list of gene signatures to test (referred to as test signatures or $t_signatures$). Test signatures are first organized into bins based on their size (i.e., number of genes per signature). For each bin, a list of background signatures ($b_signatures$) of equivalent size is created by random gene sampling to serve as a control distribution in subsequent p value calculations. Signature scoring is performed per cell for both test and random gene signatures, followed by differential expression analysis to identify statistically significant and robust gene signatures. **B** Differential Expression Workflow: Differential expression testing relies on a Monte Carlo approximation of empirical p values through subsampling, combined

with leave-one-out cross-validation by excluding individual patient samples. Initially, the data (counts matrix) is divided into folds, with each fold f_q representing the exclusion of one sample from either comparison group. In the subsampling step (Monte Carlo simulation), an equal number of cells are sampled from each patient to ensure balanced representation. A Mann-Whitney U test is then performed per subsample for each fold, for both the test gene signatures and the background distribution (derived from random gene signatures). Each test gene signature is matched with a corresponding background distribution based on bin size, and an empirical p value is computed (reflecting the test distribution's median percentile relative to the background). Additionally, a Fold Rejection Ratio (FRR) (see “Methods”) is calculated per test gene signature for each fold, providing a measure of the robustness of each gene signature to patient sample exclusion.

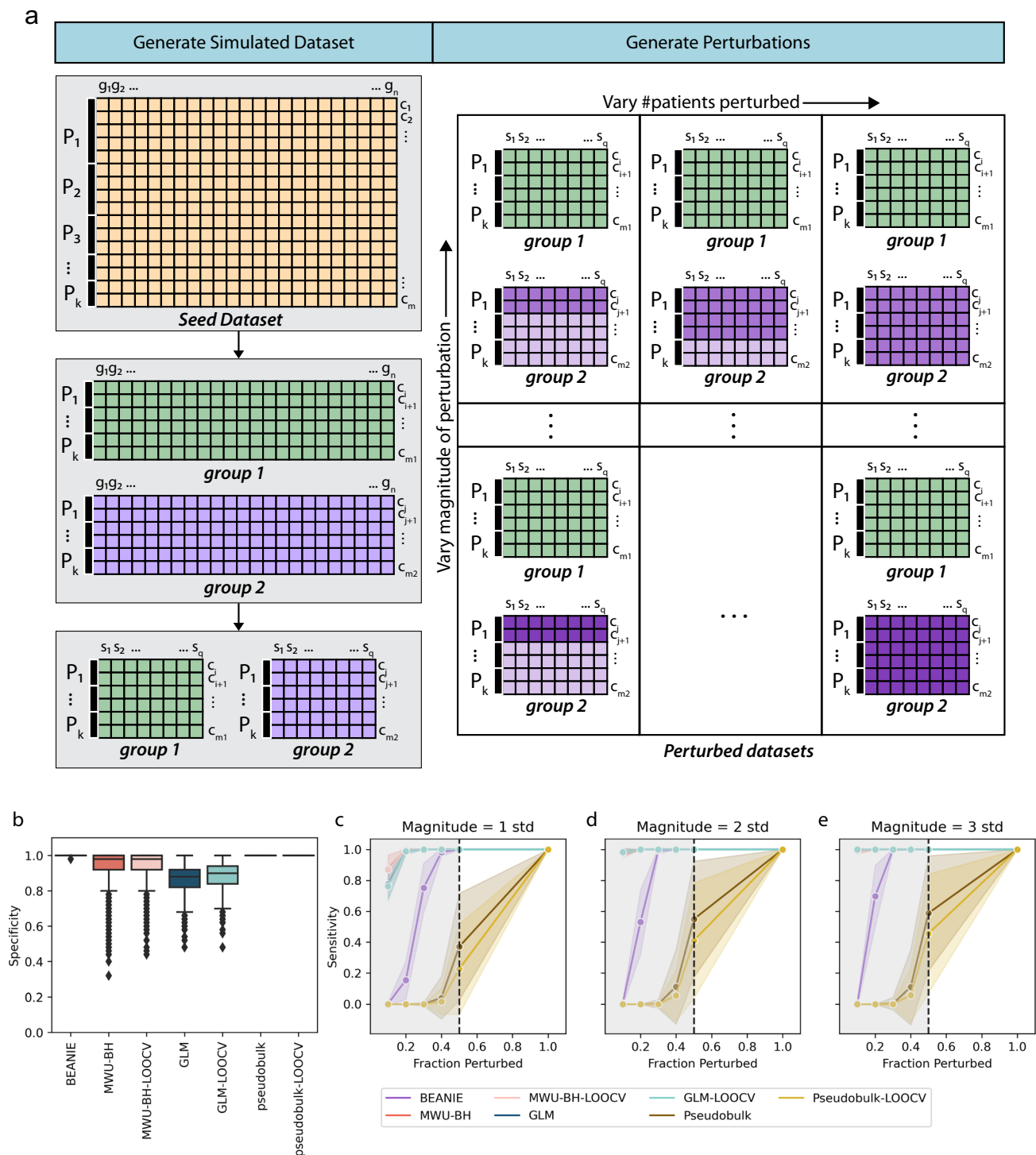


Fig. 2 | Sensitivity and specificity analysis of methods for differential expression of gene signatures. **a** Schematic for generating the simulated dataset and introducing controlled perturbations to model different levels of inter- and inpatient variation (see Methods). **b** Boxplot depicting the specificity of BEANIE in identifying the differential expression of gene signatures ($n = 50$) in the simulated dataset ($n = 1000$ trials) benchmarked against six conventional methods (MWU-BH, MWU-BH-LOOCV, GLM, GLM-LOOCV, pseudobulk and pseudobulk-LOOCV). The box represents the interquartile range (IQR), the line within the box indicates the median, whiskers represent the smallest and largest data points within 1.5 times the IQR and outliers are represented as individual points. **c–e** Lineplots

showing the sensitivity of BEANIE and conventional methods (MWU-BH, MWU-BH-LOOCV, GLM, GLM-LOOCV, pseudobulk and pseudobulk-LOOCV) in identifying the differential expression of gene signatures ($n = 50$) in the simulated dataset ($n = 1000$ trials) as a function of varying fraction of perturbed samples and magnitude of perturbation (1–3 std). Solid lines represent the mean sensitivity, and the shaded regions around each line indicate the 95% confidence intervals derived from the variability across 1000 trials for each perturbation condition. Perturbation of less than 50% samples in the group are shaded in gray.

magnitudes (1–3 std), BEANIE achieved a perfect sensitivity of 1.0 across all perturbation magnitudes (Supplementary Table 2). This performance was matched by MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV (Fig. 2c–e, Supplementary Tables 3–6), though these methods had significantly lower specificity compared to BEANIE. Conversely, while pseudobulk and pseudobulk-LOOCV exhibited similar specificity to BEANIE, they showed considerably lower sensitivity at the 50% perturbation mark across all perturbation magnitudes (Supplementary Tables 7, 8). Notably, pseudobulk-LOOCV consistently demonstrated lower sensitivity than pseudobulk across all perturbation levels.

These findings underscore the trade-offs between sensitivity and specificity: MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV excel in sensitivity at low perturbation levels but lack specificity, while pseudobulk methods offer high specificity but diminished sensitivity in this setting. In contrast, BEANIE effectively balances both specificity and sensitivity at biologically relevant perturbation magnitudes. We also calculated the time taken by BEANIE's differential testing module for a dataset of different sizes, achieved by varying the number of patients per group. We found that BEANIE scales almost linearly with an increase in both number of cells and number of patients (Supplementary Fig. 1).

BEANIE reveals biologically meaningful insights in immunotherapy treatment for scRNA-seq datasets

Building upon simulated benchmarks for systematic evaluation across controlled conditions, we extended our evaluation to real-world clinical datasets to test BEANIE's ability to extract biologically meaningful signals related to the tumor compartment. We used publicly available breast cancer dataset from Bassez et al.¹⁹ (Supplementary Fig. 2, Supplementary Table 9; see “Data Availability”) to identify robust differentially expressed gene signatures related to anti-PD-1 treatment. To account for the mixed efficacy of anti-PD-1 treatment on breast cancer, as noted in prior studies^{20–23}, we first stratified our analysis by breast cancer subtype—hormone-sensitive ER+ and triple-negative breast cancer (TNBC). We then used BEANIE, along with MWU-BH, MWU-BH-LOOCV, GLM, GLM-LOOCV, pseudobulk, and pseudobulk-LOOCV, to assess the differential expression of cancer-related gene signatures curated from MSigDB and published literature (see “Data Availability”). This analysis was conducted for both anti-PD-1 naive and anti-PD-1 treated tumors across each breast cancer subtype (see “Methods”). We focused on shared tumor states between the groups, defined as states containing more than one sample from each group. While not the focus of this analysis, the group-specific tumor states may also reveal unique biological features pertinent to each group.

ER+ Breast Cancers. We first compared anti-PD-1 naive vs. anti-PD-1 treated ER+ breast cancers from matched patient biopsies (27,118 cells, 30 samples; Fig. 3a). Tumor cells were clustered, with batch correction for patient IDs, to identify three shared tumor states between the naive and treated ER+ breast cancers, each represented by more than one sample per group (Supplementary Fig. 2; “Methods”). For each of the shared tumor states, BEANIE identified differentially expressed gene signatures (based on the empirical p value and robustness ratio) with higher specificity compared to MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV, and higher sensitivity compared to pseudobulk and pseudobulk-LOOCV. For example, in tumor state 1, BEANIE identified five differentially expressed gene signatures, as opposed to 196 identified by MWU-BH, 200 by MWU-BH-LOOCV, 194 by GLM, and 199 by GLM-LOOCV, while no candidate signatures were nominated by pseudobulk and pseudobulk-LOOCV methods (Fig. 3b, Supplementary Tables 10–12). These gene signatures also yielded insights into the biological pathways associated with the distinct tumor states (Fig. 3c, Supplementary Table 11). The five candidate signatures differentially expressed in tumor state 1 identified via BEANIE included those

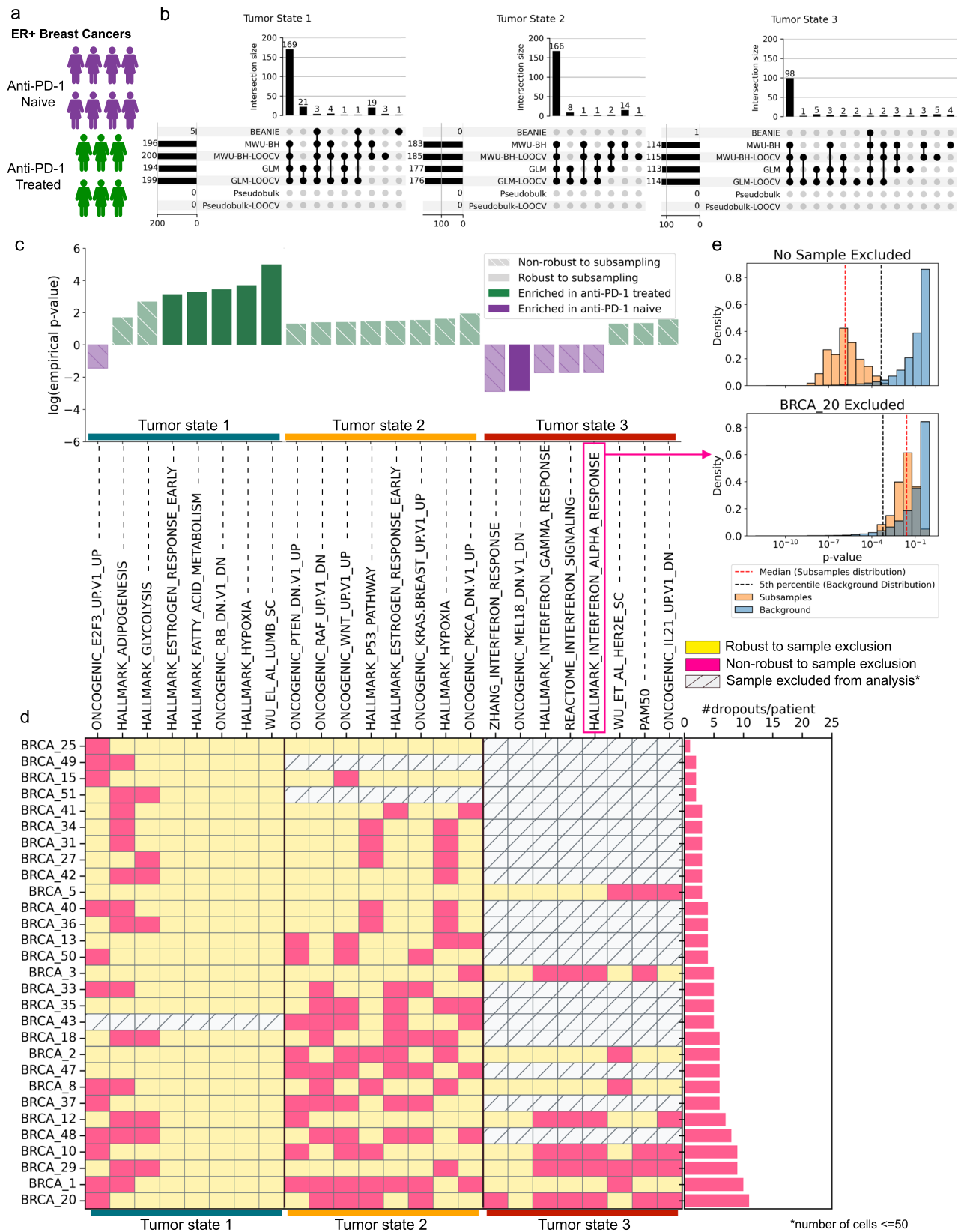
involving hypoxia, metabolic reprogramming, and estrogen response, which have been implicated in ER+ tumor progression and immunotherapy resistance^{24–26}. Interestingly, the gene signature for luminal B breast cancer (more aggressive, faster proliferating subtype) was also differentially expressed in anti-PD-1 treated tumors. In tumor state 3, a gene signature for epigenetic alterations due to knockdown of Me18 was found to be differentially expressed in the anti-PD-1 naive tumors, similar to observations in other cancers²⁷. In contrast, no robust differentially expressed gene signatures for tumor state 2. Overall, BEANIE identified robust gene signatures associated with anti-PD-1 exposure, some of which may be potentially related to immunotherapy resistance in the ER+ tumors, in line with prior studies²³.

We also examined gene signatures that were identified as statistically significant (adjusted p value ≤ 0.05), yet exhibited non-robustness to sample exclusion by BEANIE (Fig. 3d), such as the HALLMARK_INTERFERON_ALPHA_RESPONSE gene signature from MSigDB. We found that exclusion of one or more samples led to the test distribution median becoming greater than the 5th percentile of the background distribution, compared to it being less than the 5th percentile when the sample was included (median p value with sample = 1.465×10^{-6} , median p value without sample = 0.026; Fig. 3e), leading to statistically significant but non-robust results. Instead, these particular gene signatures were driven by sample-specific biology and were not representative of the group level biology. Notably, these non-robust signatures were incorrectly identified as differentially expressed by conventional MWU-BH and GLM and their LOOCV variants.

Triple negative breast cancers. We next compared anti-PD-1 naive vs. anti-PD-1 treated triple negative breast cancers from matched patient biopsies (25,389 cells, 13 patients) and identified four shared tumor states with greater than one patient sample per group (Supplementary Fig. 3). Similar to the ER+ breast cancers, BEANIE had a higher specificity in detecting robust differentially expressed gene signatures compared to MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV, and displayed higher sensitivity than pseudobulk and pseudobulk-LOOCV, in the TNBC setting (Supplementary Table 13–15). All evaluated gene signatures were enriched in the anti-PD-1 naive group relative to the anti-PD-1 treated group, (Supplementary Table 14), including signatures for epithelial-to-mesenchymal transition (tumor state 1), and the PI3K/AKT/mTOR signaling pathway (tumor state 2). These signatures, indicative of tumor aggressiveness and progression in certain settings, have been implicated in TNBC and immunotherapy contexts in prior works²⁸. Surprisingly, we found immune response-related gene signatures downregulated in the treatment-naive group, suggesting potential resistance mechanisms to therapy that could be clarified with further information on treatment responses of patients.

Relative performance on false positives. We additionally evaluated gene signatures consisting of immune cell surface markers as negative control for the tumor compartment analyses performed. When anti-PD1 naive and anti-PD1 treated groups were compared for differential expression of immune cell surface markers, T cell, B cell, NK cell, mast cell, and other myeloid cell surface markers were identified as differentially expressed by MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV in both ER+ and TNBC tumor cells (Supplementary Table 16–17). On the other hand, BEANIE did not detect any of the immune cell gene signatures as differentially expressed, similar to pseudobulk methods. These results further corroborated that the low specificity of MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV lead these conventional methods to identify a high number of false positives.

Therefore, BEANIE enabled identification of biologically meaningful and robust gene signatures associated with anti-PD-1 treatment in ER+ and triple-negative breast cancer subtypes in clinical scRNA-seq datasets, and displayed substantially lower susceptibility to detecting false positives relative to conventional methods.



BEANIE identifies robust gene signatures specific to tumor subtype biology

We next assessed BEANIE's utility for identifying gene signatures related to subtype specific biology in breast cancer. We compared the ER+ and TNBC subtypes using the scRNA-seq breast cancer dataset from Bassez et al.¹⁹ as described before (Supplementary Fig. 2,

Supplementary Table 3; see "Data Availability"), stratified by anti-PD-1 treatment exposure to account for therapy induced tumor micro-environment remodeling.

Anti-PD-1 treated breast cancer. We compared the biological differences in tumor states between ER+ and TNBC tumors from anti-PD-1-

Fig. 3 | Differential expression of gene signatures between anti-PD-1 naive vs. anti-PD-1 treated ER+ breast cancers. **a** Schematic showing the comparison groups of anti-PD-1 naive vs. anti-PD-1 treated ER+ breast cancers. **b** UpSetPlot depicting the overlap between differentially expressed gene signatures identified by BEANIE, MWU-BH, GLM, pseudobulk and their LOOCV variants, across tumor states (c) Barplot depicting examples of biologically relevant gene signatures identified as differentially expressed by BEANIE for each of the tumor states. **d** Sample Dropout Plot depicting the samples resulting in non-robustness of various statistically significant gene signatures across tumor states using BEANIE. The heatmap highlights the samples with a Fold Rejection Ratio (FRR) less than

threshold (0.9), and the horizontal bar plot depicts the number of statistically significant but non-robust gene signatures ('dropout' signatures) when the specific sample is excluded. **e** Histogram illustrating the effect of sample exclusion on the robustness of evaluated gene signatures during differential expression testing with BEANIE. During the testing of differential expression of HALLMARK_INTERFERON_ALPHA_RESPONSE, when BRCA_20 is present, the distribution of subsamples is less than the 5th percentile of the background distribution (i.e., statistically significant). However, when BRCA_20 is excluded, the distribution shifts such that the median is not in the top 5th percentile of the background distribution. Therefore, the empirical *p* value is greater than 0.05 for the fold when BRCA_20 is excluded.

treated patients (28,290 cells, 28 samples; Fig. 4a, Supplementary Fig. 4). BEANIE identified gene signatures enriched across these tumor states, as well as uniquely enriched within individual tumor cell states with high specificity. On the other hand, MWU-BH, MWU-BH-LOOCV, GLM, and GLM-LOOCV had low specificity. Pseudobulk and pseudobulk-LOOCV nominated a lower number of candidate differentially expressed gene signatures compared to BEANIE (Fig. 4b, Supplementary Fig. 4, Supplementary Tables 18–20). BEANIE identified gene signatures enriched across two or more tumor states (Fig. 4c), and amongst them were signatures related to oncogenic signaling, immune response, metabolic response, and cell structure regulation (Fig. 4d). Furthermore, biologically supported gene signatures that were specific to tumor states were also nominated by BEANIE. For instance, in tumor state 1, gene signatures related to cell growth and proliferation (including RAF and MEK signaling) were enriched in the TNBC group, while a gene signature for early response to estrogen was enriched in the ER+ group. Moreover, consistent with prior reports, previously defined gene signatures²⁹ that represent basal (aggressive) and luminal (indolent) phenotypes were identified as upregulated in the TNBC and ER+ tumor cells, respectively. In tumor state 2, gene signatures for dysregulated metabolism pathways were enriched in the TNBC group, including PIGF pathways. Prior studies have established the role of PIGF signaling in cell motility and cytoskeletal remodeling necessary for cellular migration^{30,31}, as well as an association with poor clinical outcomes in breast cancer^{32–35}. In tumor state 3, a gene signature consisting of genes downregulated in response to HOXA9 knockdown was differentially expressed in the TNBC compared to ER+ group, therefore indicating that genes controlled by HOXA9 were upregulated in the TNBC group. This finding corroborates previous observations related to the role of HOXA9 in de-differentiation and morphogenesis of tumor cells³⁶, which is also associated with the more aggressive and undifferentiated phenotype of TNBC.

Anti-PD-1 naive breast cancer. We next identified differentially expressed gene signatures between ER+ and TNBC tumors from anti-PD-1 naive patients (24,217 cells, 28 samples), with three tumor states (Supplementary Fig. 5, Supplementary Tables 21–23). BEANIE identified gene signatures enriched across two or more of these tumor states, including hallmark interferon gamma response and hallmark TNF-alpha signaling via NFK-B. These gene signatures are implicated in various cancers, and prior studies have found interferon gamma response to be an indicator of tumor aggressiveness³⁷. BEANIE also identified gene signatures uniquely enriched within individual tumor cell states, including mTOR signaling, interferon alpha, and ERK signaling, among others (Supplementary Fig. 5, Supplementary Table 21). Gene signatures enriched in individual tumor states included PAM50³⁸ in tumor state 1. In tumor state 2, gene signatures related to adipogenesis, fatty acid metabolism, and cholesterol homeostasis were differentially expressed in the TNBC subtype, suggesting potentially heightened disruption of energy storage and metabolism-related characteristics in this subpopulation. Differentially enriched gene signatures in tumor state 3 showed upregulated interferon gamma related signatures in the ER+ group, including interferon gamma response and interferon alpha response signatures. This observation is contrary

to the differential expression of cancer-related gene signatures in the TNBC subtype for all other tumor states identified in this analysis.

Again, for both anti-PD1 and treatment-naive tumors, many of the gene signatures consisting of the immune cell surface markers were found to be differentially expressed by MWU-BH, MWU-BH-LOOCV, GLM and GLM-LOOCV but not BEANIE or pseudobulk methods (Supplementary Table 24, 25). Notably, both shared and state-specific gene signatures were observed, identifying hypotheses for the different potential ways in which ER+ and TNBC tumor cells reprogram, and aligning with results from prior studies on clinical outcomes^{39–41}.

BEANIE identifies robust gene signatures in meta-analyses across datasets

Building upon single-cohort study performance, we next assessed the relative applicability of BEANIE compared to established approaches for a meta-analysis composed of multiple independent scRNA-seq datasets. We curated lung adenocarcinoma (LUAD) patient samples from three datasets—Kim et al.⁴², Maynard et al.⁴³, Qian et al.⁴⁴ (see “Methods, Data Availability”), resulting in 17,759 tumor cells from 36 patients (Supplementary Table 9). We assessed the intrinsic biological differences between early-stage and late-stage tumor cells across these studies. Recognizing the importance of prior treatment in potentially altering the tumor cells and introducing varying levels of heterogeneity, we focused on treatment-naive patient samples (Fig. 5a, b). As above, for each analysis, we compared the performance of BEANIE with respect to conventionally used methods, MWU-BH, MWU-BH-LOOCV, GLM, GLM-LOOCV, pseudobulk and pseudobulk-LOOCV.

We first identified three shared tumor states between all lung cancer datasets and then identified differentially expressed gene signatures with all of these strategies. BEANIE identified differentially expressed gene signatures with more specificity compared to MWU-BH, MWU-BH-LOOCV, GLM and GLM-LOOCV (Fig. 5c, Supplementary Table 26–28). Once again, as negative control, many of the gene signatures consisting of cell surface markers for different immune cell types (T cells, B cells, NK cells, mast cells and other myeloid cells) were identified as differentially enriched by MWU-BH, MWU-BH-LOOCV, GLM and GLM-LOOCV, but not BEANIE or pseudobulk methods (Supplementary Table 29). In tumor states 1 and 4, BEANIE identified gene signatures differentially expressed in the early-stage LUAD tumors. These included a gene signature for NRF2 (NFE2L2) activation in tumor state 1, and the glycolysis pathway and a hormone independent tumor growth gene signature in tumor state 4 (Fig. 5d). Prior studies observed that permanent activation of Nrf2 promotes various cancer properties, including malignancy progression, chemotherapy resistance, and poor patient prognosis⁴⁵. Furthermore, glycolysis dysregulation has been indicated in metabolic reprogramming phenotypes relevant to metastasis^{46–48}. Additionally, in tumor state 2, we observed differential expression of genes in the late-stage LUAD tumors related to downregulation due to AKT1 gene overexpression.

The meta-analysis findings demonstrated that BEANIE can identify robust differentially enriched gene signatures within and across tumor states that were not identified by either MWU-BH or GLM in such settings.

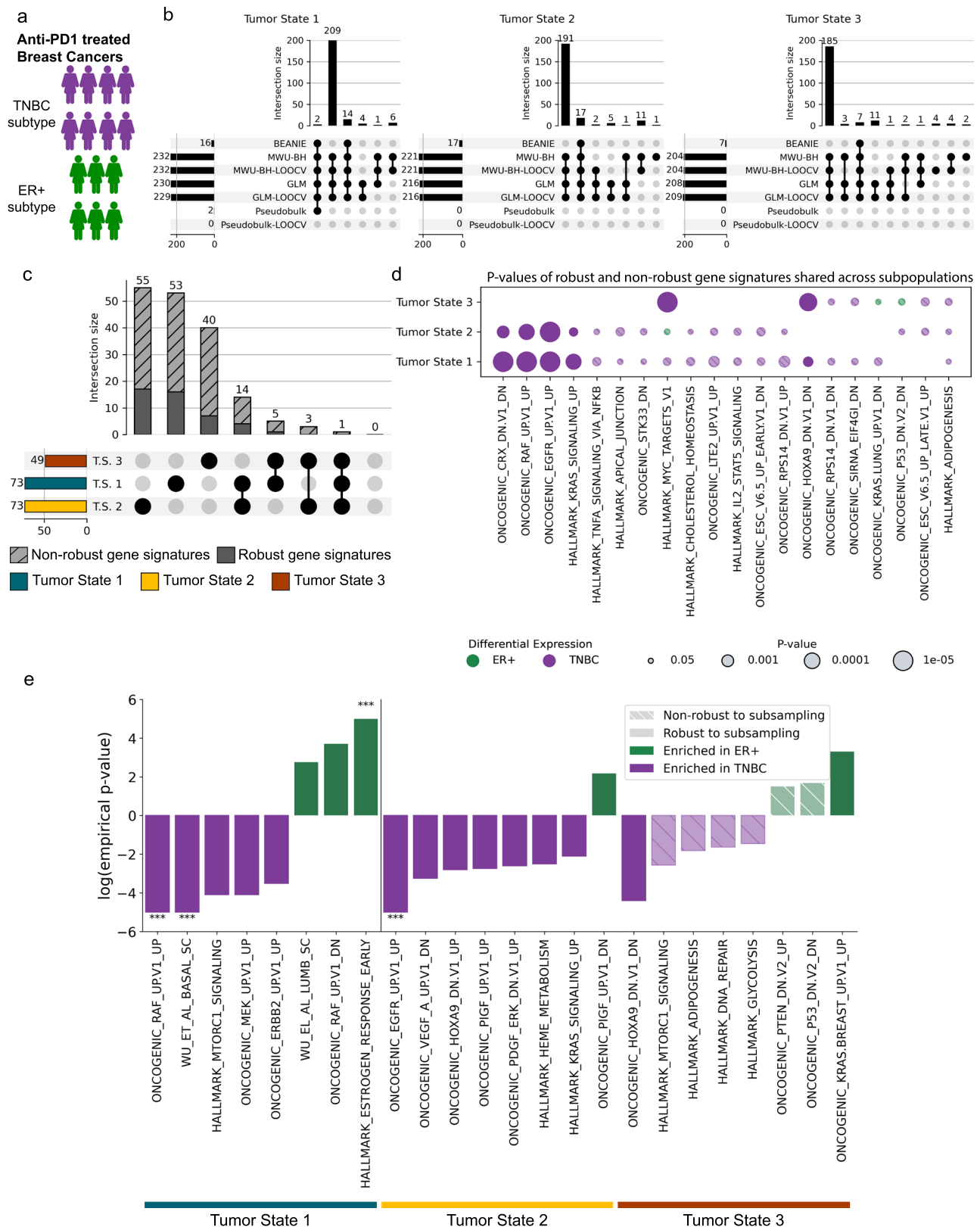


Fig. 4 | Differential expression of gene signatures between TNBC vs. ER+ anti-PD-1 treated breast cancers. **a** Schematic showing the comparison groups—TNBC vs. ER+ anti-PD-1 treated breast cancers. **b** UpsetPlot depicting the overlap between differentially expressed gene signatures identified by BEANIE, MWU-BH, GLM, pseudobulk and their LOOCV variants, across tumor states. **c** Upsetplot showing

the gene signatures differentially expressed across tumor cell states or within an individual tumor state. **d** Dotplot showing the gene signatures enriched across tumor cell states (shared among two or more tumor states). **e** Barplot depicting examples of differentially expressed gene signatures with biological relevance identified by BEANIE for each of the tumor states (***) indicates p values < 0.0001).

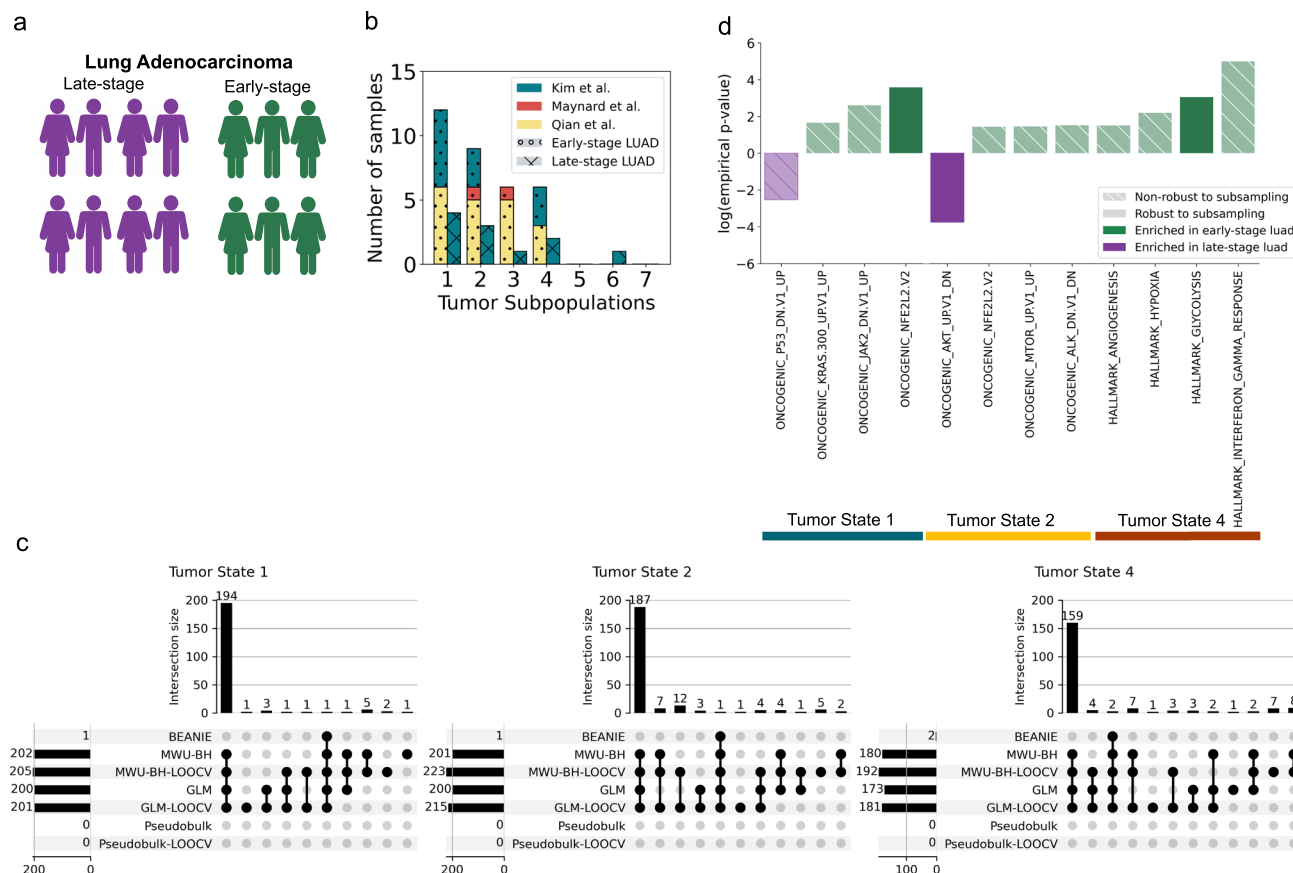


Fig. 5 | Differentially expressed gene signatures in early-stage vs. late-stage treatment-naïve LUAD. **a** Schematic showing the comparison groups—early-stage vs. late stage treatment-naïve LUAD. **b** Number of samples per tumor state across the two clinical groups, colored by datasets. **c** UpSetPlot depicting the overlap

between differentially expressed gene signatures identified by BEANIE, MWU-BH, GLM, pseudobulk and their LOOCV variants, across tumor states. **d** Barplot depicting examples of biologically meaningful differentially expressed gene signatures identified by BEANIE for each of the tumor states.

BEANIE yields corroborating insights across cell types

Lastly, we investigated the consistency of biological findings across cell types. To study this, we assessed whether gene signatures identified as differentially expressed by BEANIE in the tumor cells were consistently supported by biological findings in the corresponding immune cells from the same tumors. We used the dataset from Jerby-Arnon et al.⁹ (see Data Availability) for our analysis, which included data for both immune checkpoint blockade (ICB)-naïve and ICB-treated melanoma patients from various biopsy sites, totalling 2018 tumor cells and 5168 immune cells across 14 patients, with 7 patients per group (Supplementary Fig. 6, Supplementary Table 9).

We first conducted a global analysis on the entire tumor cell compartment for identifying differentially expressed gene signatures across all tumor cells. Similar to prior observations, BEANIE identified a small number of differentially expressed gene signatures between tumor cells from ICB-naïve and ICB-treated groups, compared to MWU-BH, MWU-LOOCV, GLM and GLM-LOOCV, whereas no gene signatures were identified by the pseudobulk methods (Supplementary Fig. 5, Supplementary Table 30–33). Among those identified as differentially expressed by BEANIE (Supplementary Table 31), gene signatures for immune inflammatory responses were upregulated in tumor cells from the ICB-naïve group, specifically the gene signatures for *STAT5* upregulation in response to *IL2* stimulation and genes upregulated by *IL6* via *STAT3*. Consistent with these findings in the tumor cells, we observed that *IL2* was differentially expressed (p value = 0.0099) in the T cells of the ICB-naïve group. Furthermore, we found a differential *STAT3* expression in the T cells (p value = 0.0041). These biologically consistent observations across cell type specific

compartments support the findings for *STAT5* upregulation in response to *IL2*, and potential *IL6* induced downstream signaling via *STAT3* in the tumor cells of the ICB-naïve group^{49,50}. Together, these results describe the tumor microenvironment of the ICB-treated group (consisting of treatment-resistant patients) as one depleted of T cells, with reduced *IL2-STAT5* signaling, *TNFA-NFKB* signaling, and inflammatory response relative to the ICB-naïve group.

Discussion

Validating experimentally and computationally derived gene signatures^{5–8} through differential expression analysis between clinical groups is a key step toward maximizing learning from scRNA-seq studies. Most methods for differential expression of individual genes^{1–4} are parametric and not well-suited for gene signatures, which have varying distribution due dependent on the collection of genes that constitutes the gene signature. As a result, many current approaches for analyzing differential expression of gene signatures rely on techniques originally developed for bulk sequencing. These include statistical approaches such as the Mann-Whitney U test, regression based techniques like Generalized Linear Models, and pseudobulk methods that aggregate single-cell data to mimic bulk sequencing formats^{9–12}. While some studies incorporate methods like leave-one-out cross-validation alongside these approaches, they typically optimize either sensitivity or specificity, but not both simultaneously. This limitation can lead to problematic outcomes - an overwhelming number of candidate differentially expressed genes, including many false positives, or the inadvertent discarding of potentially important differentially expressed genes. These issues significantly complicate the

interpretation of results and experimental validation in subsequent studies.

The limitations of conventional methods such as MWU-BH and GLMs for analyzing clinical scRNA-seq datasets for differential expression of gene signatures stem from two main issues. First, while the large number of cells in scRNA-seq datasets increases the statistical power to detect group differences, these methods often fail to distinguish biologically meaningful variations from patient-specific differences that may be unrelated to the clinical group. Second, the unequal contribution of cells from each patient, combined with the diverse tumor biology within the same clinical group, makes these methods susceptible to non-robust results driven by sample-specific variability. Pseudobulk methods avoid these issues by averaging all cells from a patient into one data point, however this leads to loss of granularity in analysis as well as reduces the number of data points to below the threshold for samples required for statistical significance in many clinical scRNA-seq datasets. This issue is particularly concerning in clinical settings, where group comparisons—such as those related to treatment effects or disease progression—can strongly influence downstream functional studies and clinical trials.

To overcome these challenges, we developed BEANIE, a non-parametric statistical approach tailored to accurately estimate group-level biology in clinical scRNA-seq datasets. BEANIE employs Monte Carlo subsampling techniques to address the variability in cell numbers per patient—a common issue in scRNA-seq data that can skew differential expression analyses. BEANIE is particularly well-suited for clinical datasets with multiple patients, as it minimizes the influence of patient-specific biological differences by utilizing leave-one-out cross-validation and correction against a background distribution. In benchmarking tests with simulated datasets, BEANIE outperformed conventional methods in balancing the sensitivity-specificity trade-off. Further validation on publicly available datasets from five clinical single-cell transcriptomic studies demonstrated BEANIE's ability to identify statistically significant, robust gene signatures while effectively distinguishing between sample-driven and group-driven signatures. BEANIE successfully extracted biologically meaningful signals from tumor cells in datasets spanning breast cancer, lung cancer, and melanoma, covering a range of treatments, cancer stages, and subtypes. BEANIE's findings in the tumor compartment were also consistent when tested in immune cells from the same patients, highlighting the promises of using this approach.

While BEANIE excels in detecting statistically significant and robust gene signatures with high sensitivity and specificity, its effectiveness partly depends on the quality of the gene signatures being analyzed. Larger databases containing cell-type specific gene signatures would enable better insights. Furthermore, heterogeneity within the clinical group may limit BEANIE's ability to extract meaningful biological signals (e.g., in analyzing treatment-naïve vs treated patients, the treated group may include both responders and non-responders). Incorporating additional clinical and molecular information, such as mutational status, can enhance the analysis and provide deeper insights. Ultimately, larger datasets, particularly those generated through consortium efforts, are essential to capture patient variability and improve the robustness of the conclusions drawn. In summary, BEANIE represents a significant advancement in the analysis of clinical single-cell transcriptomic datasets, offering enhanced sensitivity and specificity in detecting biologically meaningful gene signatures. Its effectiveness can be further amplified with high-quality data, additional clinical insights, and larger, more comprehensive datasets.

Methods

Data preprocessing

Breast cancer dataset. Raw data was downloaded (.rds objects and.csv metadata files) from <https://lambrechtslab.sites.vib.be/en/>

single-cell (Bassez et al. 2021). HER2+ and chemotherapy + anti-PD1 groups were excluded due to low sample size. The dataset in four ways –(i) treatment-naïve ER+ vs. treatment-naïve TNBC, (ii) anti-PD1 treated ER+ vs. anti-PD1 treated TNBC, (iii) treatment-naïve ER+ vs. anti-PD1 treated ER+, and (iv) treatment-naïve TNBC vs anti-PD1 treated TNBC. Original author annotations for tumor cells were used.

Meta-analysis of lung adenocarcinoma

Kim et al.⁴²: Raw counts and cell annotations were downloaded from GSE131907 (GSE131907_Lung_Cancer_raw_UMI_matrix.txt.gz, GSE131907_Lung_Cancer_cell_annotation.txt.gz). Additional metadata about patient history was downloaded from supplementary table S1 presented in the paper. Only patient samples from the primary tumor site were retained for further analysis, and samples from pleural effusions and metastasis sites were removed. Original author annotations for tumor cells were used.

Maynard et al.⁴³: Raw counts were downloaded from https://github.com/czbiohub-sf/scell_lung_adenocarcinoma (NI01_Nonimmune_Seurat_object_annotated.RData contained all epithelial cells, and NI05_all_epithelial_annotated_normal_and_tumor.RData contained annotations for tumor vs. non-tumor cells). Additional metadata was downloaded from Supplementary Table S1 of the paper. Only samples acquired from primary tumor sites, and those from patients diagnosed with lung adenocarcinoma were retained. Other samples were removed from further analysis. Original author annotations for tumor cells were used.

Qian et al.⁴⁴: Raw counters were downloaded from <https://lambrechtslab.sites.vib.be/en/pan-cancer-blueprint-tumour-microenvironment-0>. Additional metadata information (tumor stage) was acquired from Supplementary Table S1 of the paper. Original author annotations for tumor cells were used.

Data from the three lung cancer datasets was jointly analyzed to identify differentially expressed gene signatures between early-stage and late-stage treatment naïve patients. Stages I, II and IIIA were classified as early-stage tumors, and stage IV as late-stage tumors, consistent with Kim et al.⁴². Patients with stage IIIB were excluded from analysis. Original author annotations for tumor cells were used.

Melanoma dataset. Counts matrix and cell type annotations were downloaded from the GEO database (GSE115978, Jerby-Arnon et al.⁹). Other metadata including metastatic status, location of tumor etc. was acquired from the Supplementary Table S1 of the paper.

Identification of tumor states

To perform differential gene signature expression analysis using BEANIE, we first identified tumor states using leiden clustering, in line with analysis methodology used for other scRNA-seq data analysis. The dataset is subsetted to tumor cells, followed by log-normalization of the raw counts and identification of highly variable genes. It is then scaled followed by PCA analysis. Harmony is used for the correction of batch effects (“sample_id” is used as a confounding variable). Finally, a k-nearest neighbors graph is constructed and leiden clustering is performed to identify tumor states (see “code availability”).

Simulations for benchmarking BEANIE with existing methods

We developed a custom pipeline for generation of scRNA-seq dataset containing tumor cells from multiple patients. Seed patient samples were chosen from a clinical scRNA-seq dataset, inspired from methodology adopted by *Splatter*⁵¹. These patient samples were homogeneous with respect to tumor type (all patients were diagnosed with triple negative breast cancer). Cells were then sampled without replacement from each patient sample to create two groups. This means that given a seed dataset containing n patients and k cells per patient, cells are randomly sampled from each patient without replacement to form two groups. Therefore, the two groups ($k/2$ cells per patient and n patients per group) have the same overall distribution

and hence theoretically have no biological difference. At the same time, since each group consists of multiple patients, the inter-patient heterogeneity is maintained as would be expected in a realistic clinical biological dataset.

We next scored the simulated dataset for a set of cancer-associated gene signatures extracted from MSigDB (see Data Availability). Controlled perturbations were then performed on the cells \times gene signature matrix along two axes - fraction of variation and degree of variation. To model the fraction of variation between the two groups, we modulated the number of patients perturbed (10%, 50%, and 100%), and to model the degree of variation, we modulated the magnitude of perturbation introduced (1 std, 2 std, and 3 std). We note that perturbations of fraction of patients could represent real world scenarios such as a mix group of responders and non-responders to therapy, in a group comparison between treatment-naïve and treated patients. Similarly, the perturbations of different magnitudes could represent levels of biological difference due to the variable of interest (progression, treatment etc.). A combination of these two axes allowed precise estimation of sensitivity and specificity at different kinds and levels of variation possible in clinical datasets. We then calculate sensitivity and specificity as follows for BEANIE, MWU-BH, GLM, pseudobulk, MWU-BH-LOOCV, GLM-LOOCV, pseudobulk-LOOCV.

Sensitivity calculation:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (1)$$

Specificity calculation:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (2)$$

BEANIE

Signature scoring. For each cell c_i in the normalized counts matrix, signature scoring is performed for the set of gene signatures provided as input by the user (test signatures). The default signature scoring method is adapted from AUCCell⁷.

(i) For each gene g_k , the cells are ranked by calculating the percentile of each cell across the gene g_k in terms of normalized expression of the gene, i.e., cells with higher expression values of that particular gene will have a higher percentile. The ties are randomly broken (i.e., if two cells have the exact same expression of the gene, which is common in single-cell datasets, those cells are randomly assigned a percentile value).

$$P_{c_i, g_k} = \frac{100 \times n_{c_i, g_k}}{n_c} \quad (3)$$

where n_{c_i, g_k} = ordinal rank of c_i for expression of g_k (sorted from smallest to largest), P_{c_i, g_k} = percentile of c_i for expression of g_k , and n_c is the total number of cells

(ii) Next, for every cell c_i , genes are ranked based on their calculated percentile values across that cell. Genes which have a higher percentile across the cell are given lower ranks. This scoring system takes into account the importance of each gene in a given cell relative to that gene's importance in other cells, i.e., genes which have a lower rank are more important for the cell in question as compared to genes with a higher rank.

$$P'_{g_k, c_i} = \frac{100 \times n'_{g_k, c_i}}{n_g} \quad (4)$$

where n'_{g_k, c_i} = ordinal rank of g_k for P'_{g_k, c_i} values (sorted from largest to smallest), P'_{g_k, c_i} = percentile of g_k for P'_{g_k, c_i} values, and n_g is the total number of genes for the cell c_i

(iii) For each gene signature S_j , a recovery curve per cell c_i is generated by calculating the enrichment of the top constituent genes ranked from S_j , followed by a calculation of the Area Under the Curve (AUC), which measures the expression of c_i 's top constituent genes ranked from S_j . The AUC is therefore the score of the cell for S_j .

$$AUC_{j,i} = \int f_i(x) dx, f_i(x) = n \left(S_j \cap R_{x,i} \right) \quad (5)$$

where S_j = set of genes comprising a gene signature and $R_{x,i}$ = set of top constituent x genes based on P'_{g_k, c_i} .

Other signature scoring methods available in BEANIE include weighted mean and z-scoring.

Background distribution generation. A background distribution is generated for the biological interpretability of the results as follows: (i) Bins are created based on the gene set size of each signature S_j (default bin size = 10, tunable parameter). (ii) Random signatures (r signatures) ($R_k, k = 1, 2, \dots, n_b$, where n_b = total number of bins) for each of the bin sizes are generated such that they are representative of both lowly expressed and highly expressed genes. For this step, the normalized matrix is used, and the genes are sorted based on their expression values across all samples. Equal numbers of genes from every 20th percentile are then randomly subsampled such that the sum of all genes equals the bin size. This random sampling is repeated multiple times to generate different random signatures ($R_{kl}, l = 1, 2, \dots, n_r$, where n_r = the total number of times subsampling is repeated). The rationale for generating the random signatures is that they should not represent any biologically meaningful gene signature, and as a consequence, their differential expression can be used as a null distribution (background distribution) for interpretation of the results in a biological context. (iii) Each cell c_i is scored for R_{kl} 's using the aforementioned signature scoring method.

Folds and subsampling. To accomplish BEANIE's two-fold aim of having equal sample representation and quantifying robustness for S_j , two statistical techniques, Monte Carlo approximations (subsampling) and leave-one-out cross-validation (sample exclusion), are coupled. First, the data is divided into folds ($f_q, q = 1, 2, \dots, n_p$, where n_p = number of samples), with each fold f_q representing the exclusion of one sample from either group. For each fold f_q , cells are subsampled such that each sample is represented by an approximately equal number of cells. This is done by first subsampling an equal number of cells from all samples, followed by additional subsampling in the sample-excluded group to compensate for the cells that would have otherwise been subsampled from the excluded sample. The additional subsampling ensures that the total number of cells subsampled from the two groups being tested always remains constant regardless of which group the excluded sample belongs to, which is necessary to ensure that the folds are comparable with each other. The subsampling is then repeated multiple times to establish adequate representation of each patient sample.

Identification of differentially expressed signatures. A multi-step strategy is adopted to identify differentially expressed signatures. First, for each subsample belonging to the fold f_q , a MWU test is performed between the two groups for every S_j . Additionally, for each fold f_q , a null p value distribution is generated by a MWU test between the two groups for every R_{kl} . The null distribution generated is fold-specific to ensure that the sample excluded from the fold is also excluded for the generation of the null distribution. The percentile of the subsample's p value against the null p value distribution is then calculated, hereafter referred to as the empirical p value. A median empirical p value is calculated for these subsamples to represent the p value for a given fold, followed by a median across all folds to

represent the cell's p value. To quantify the robustness of S_j to sample exclusion, a ratio (henceforth referred to as the Fold Rejection Ratio (FRR)) is defined and calculated for every fold f_q .

$$FRR_q = \begin{cases} \frac{n(F_{q,p \leq 0.05})}{n(F_{0,p \leq 0.05})} & \text{if } n(F_{0,p \leq 0.05}) \neq 0 \\ 0 & \text{if } n(F_{0,p \leq 0.05}) = 0 \end{cases} \quad (6)$$

where F_q = set of subsamples for the fold f_q which have an empirical p value ≤ 0.05 and F_0 = set of subsamples for the fold f_0 (when no sample is excluded) which have an empirical p value ≤ 0.05 .

A FRR value closer to 1 indicates that exclusion of the sample has no effect on the empirical significance of the gene signature S_j , and a lower value indicates the opposite. We use a threshold of 0.9 (hyper-parameter) to call signatures as robust or not, i.e., if the FRR for a particular S_j is greater than 0.9 for *all* folds, then the gene signature is considered robust to sample exclusion.

Gene ranking. For every gene signature S_{ji} , the genes are then ranked for the robustness of their \log_2 fold change between the two groups. This is particularly useful for larger gene sets. For every sample, a mean gene expression (MGE) is calculated for every gene using the normalized counts. A similar strategy of subsampling coupled with sample exclusion is used for ranking. The MGE matrix is then divided into folds, with each fold representing the exclusion of one sample. A \log_2 fold change is then calculated for each fold, and the standard deviation, along with the mean across folds, is also calculated. Genes with both outlier MGE values and outlier \log_2 fold changes (i.e., MGE values and \log_2 fold changes more than 1.5 times the interquartile range above the third quartile or below the first quartile) are classified as non-robust to sample exclusion. The final ranking of genes is performed based on decreasing \log_2 fold change, increasing standard deviation, and robustness status.

Benchmarking with existing methods

MWU-BH. Mann-Whitney U (MWU) tests followed by Benjamini-Hochberg (BH) correction are performed for the calculation of p values. The Python package *scipy* is used for the MWU p value calculation and the function *multipletests* from the Python package *statsmodels* is used for the BH correction (see “Code Availability”).

GLMs. GLM with a binomial distribution link function were used for calculation of p values. The Python package *statsmodels* is used to implement this method. The signature scores are used as covariates (exog variable), and the group labels (e.g., treatment-naive or -treated and early-stage or late-stage) as the response variable to be modeled (endog variable) (see “Code Availability”).

Pseudobulk. Pseudobulk analysis was performed by averaging the normalized counts of cells within each sample to generate representative patient-level datasets. These pseudobulk representations were subsequently analyzed using the Mann-Whitney U test, coupled with Benjamini-Hochberg correction (MWU-BH), to robustly compute p values, ensuring the identification of statistically significant differences (see “Code Availability”).

MWU-BH-LOOCV. The Mann-Whitney U test with Benjamini-Hochberg correction (MWU-BH) was performed using a leave-one-out cross-validation (LOOCV) approach. For each fold of the cross-validation, a p value was calculated, and the final p value was determined as the median across all folds (see “Code Availability”).

GLM-LOOCV. The generalized linear model (GLM) test was conducted with leave-one-out cross-validation (LOOCV). For each cross-validation

fold, a p value was derived, and the overall p value was computed as the median of the individual fold p values (see “Code Availability”).

Pseudobulk-LOOCV. Pseudobulk analysis with the Mann-Whitney U test and Benjamini-Hochberg correction (MWU-BH) was carried out using a leave-one-out cross-validation (LOOCV) framework. The final p value was calculated as the median of p values obtained from each fold of the cross-validation (see “Code Availability”).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All datasets used in the study are publicly available. The Bassez et al.¹⁹ dataset was downloaded from <https://lambrechtslab.sites.vib.be/en/single-cell>. The Kim et al.⁴² dataset was downloaded from GEO database under the accession number [GSE131907](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131907), and additional metadata about patient history was downloaded from supplementary table S1 of the published manuscript. The Maynard et al.⁴³ dataset was downloaded from https://github.com/czbiohub-sf/scell_lung_adenocarcinoma and additional metadata was downloaded from Supplementary Table S1 of the paper. The Qian et al.⁴⁴ dataset was downloaded from <https://lambrechtslab.sites.vib.be/en/pan-cancer-blueprint-tumour-microenvironment-0> and additional metadata information was acquired from Supplementary Table S1 of the paper. The Jerby-Arnon et al.⁹ dataset was downloaded from GEO database under the accession number [GSE115978](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115978) and other metadata was acquired from the Supplementary Table S1 of the paper. Hallmark and Oncogenic gene sets are publicly available for download from MSigDb (<https://www.gsea-msigdb.org/gsea/msigdb>).

Code availability

BEANIE can be run as a python package, the code for which is available at <https://github.com/vanallenlab/beanie>⁵² under the GNU General Public License (version 2). Code for reproducing all analysis performed in this study can be found at <https://github.com/vanallenlab/beanie-analysis>⁵³ under the GNU General Public License (version 2).

References

- Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 1–13 (2015).
- Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S. & Vert, J.-P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.* **9**, 1–17 (2018).
- He, L. et al. NEBULA is a fast negative binomial mixed model for differential or co-expression analysis of large-scale multi-subject single-cell data. *Commun. Biol.* **4**, 1–17 (2021).
- Zhang, M. et al. IDEAS: individual level differential expression analysis for single-cell RNA-seq data. *Genome Biol.* **23**, 1–17 (2022).
- Cortal, A., Martignetti, L., Six, E. & Rausell, A. Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. *Nat. Biotechnol.* **39**, 1095–1102 (2021).
- Marquez-Galera, A., de la Prida, L. M. & Lopez-Atalaya, J. P. A protocol to extract cell-type-specific signatures from differentially expressed genes in bulk-tissue RNA-seq. *STAR Protoc.* **3**, 101121 (2022).
- Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
- DeTomaso, D. et al. Functional interpretation of single cell similarity maps. *Nat. Commun.* **10**, 1–11 (2019).
- Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997.e24 (2018).

10. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
11. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
12. Habib, N. et al. Disease-associated astrocytes in Alzheimer's disease and aging. *Nat. Neurosci.* **23**, 701–706 (2020).
13. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
14. Lütge, A. et al. CellMixS: quantifying and visualizing batch effects in single-cell RNA-seq data. *Life Sci. Alliance* **4**, e202001004 (2021).
15. Chen, W. et al. A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. *Nat. Biotechnol.* **39**, 1103–1114 (2020).
16. Richards, L. M. et al. A comparison of data integration methods for single-cell RNA sequencing of cancer samples. *bioRxiv* 2021.08.04.453579 <https://doi.org/10.1101/2021.08.04.453579> (2021).
17. Squair, J. W. et al. Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 1–15 (2021).
18. He, M. X. et al. Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat. Med.* **27**, 426–433 (2021).
19. Bassez, A. et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **27**, 820–832 (2021).
20. Debien, V. et al. Immunotherapy in breast cancer: an overview of current strategies and perspectives. *npj Breast Cancer* **9**, 1–10 (2023).
21. Jerusalem, G. et al. Neoadjuvant nivolumab + palbociclib + anastrozole for oestrogen receptor-positive/human epidermal growth factor receptor 2-negative primary breast cancer: results from CheckMate 7A8. *Breast* **72**, 103580 (2023).
22. Schmid, P. et al. Atezolizumab and Nab-Paclitaxel in advanced triple-negative breast cancer. *N. Engl. J. Med.* **379**, 2108–2121 (2018).
23. Yao, J., Li, S. & Wang, X. Identification of breast cancer immune subtypes by analyzing bulk tumor and single cell transcriptomes. *Front. Cell Dev. Biol.* **9**, 781848 (2021).
24. Lipid profile in breast cancer. From signaling pathways to treatment strategies. *Biochimie* **219**, 118–129 (2024).
25. Zipinotti Dos Santos, D. et al. The impact of lipid metabolism on breast cancer: a review about its role in tumorigenesis and immune escape. *Cell Commun. Signal.* **21**, 161 (2023).
26. Jehanno, C. et al. Hypoxia and ER α transcriptional crosstalk is associated with endocrine resistance in breast cancer. *Cancers* **14**, 4934 (2022).
27. Wiederschain, D. et al. Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. *Mol. Cell. Biol.* **27**, 4968–4979 (2007).
28. Keenan, T. E. et al. Molecular correlates of response to eribulin and pembrolizumab in hormone receptor-positive metastatic breast cancer. *Nat. Commun.* **12**, 1–13 (2021).
29. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2021).
30. Taylor, A. P. & Goldenberg, D. M. Role of placenta growth factor in malignancy and evidence that an antagonistic PlGF/Flt-1 peptide inhibits the growth and metastasis of human breast cancer xenografts. *Mol. Cancer Ther.* **6**, 524–531 (2007).
31. Taylor, A. P., Leon, E. & Goldenberg, D. M. Placental growth factor (PlGF) enhances breast cancer cell motility by mobilising ERK1/2 phosphorylation and cytoskeletal rearrangement. *Br. J. Cancer* **103**, 82–89 (2010).
32. Chen C.N. et al. The significance of placenta growth factor in angiogenesis and clinical outcome of human gastric cancer. *Cancer Lett.* **213**, 73–82 (2004).
33. Parr, C., Watkins, G., Boulton, M., Cai, J. & Jiang, W. G. Placenta growth factor is over-expressed and has prognostic value in human breast cancer. *Eur. J. Cancer* **41**, 2819–2827 (2005).
34. Wei, S.-C. et al. Placenta growth factor expression is correlated with survival of patients with colorectal cancer. *Gut* **54**, 666–672 (2005).
35. Escudero-Esparza, A., Martin, T. A., Douglas-Jones, A., Mansel, R. E. & Jiang, W. G. PGF isoforms, PLGF-1 and PGF-2 and the PGF receptor, neuropilin, in human breast cancer: prognostic significance. *Oncol. Rep.* **23**, 537–544 (2010).
36. Gilbert, P. M. et al. HOXA9 regulates BRCA1 expression to modulate human breast tumor phenotype. *J. Clin. Investig.* **120**, 1535–1550 (2010).
37. Gocher, A. M., Workman, C. J. & Vignali, D. A. A. Interferon- γ : teammate or opponent in the tumour microenvironment? *Nat. Rev. Immunol.* **22**, 158–172 (2021).
38. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
39. Oshi, M. et al. Adipogenesis in triple-negative breast cancer is associated with unfavorable tumor immune microenvironment and with worse survival. *Sci. Rep.* **11**, 12541 (2021).
40. Ossovskaya, V. et al. Exploring molecular pathways of triple-negative breast cancer. *Genes Cancer* **2**, 870–879 (2011).
41. Clusan, L., Ferrière, F., Flouriot, G. & Pakdel, F. A basic review on estrogen receptor signaling pathways in breast cancer. *Int. J. Mol. Sci.* **24**, 6834 (2023).
42. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 1–15 (2020).
43. Maynard, A. et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* **182**, 1232–1251.e22 (2020).
44. Qian, J. et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res.* **30**, 745–762 (2020).
45. Pouremamali, F., Pouremamali, A., Dadashpour, M., Soozangar, N. & Jeddi, F. An update of Nrf2 activators and inhibitors in cancer prevention/promotion. *Cell Commun. Signal.* **20**, 1–16 (2022).
46. Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *J. Gen. Physiol.* **8**, 519–530 (1927).
47. Potter, M., Newport, E. & Morten, K. J. The Warburg effect: 80 years on. *Biochem. Soc. Trans.* **44**, 1499–1505 (2016).
48. Li, X. et al. Identifying metabolic reprogramming phenotypes with glycolysis-lipid metabolism discoordination and intercellular communication for lung adenocarcinoma metastasis. *Commun. Biol.* **5**, 1–13 (2022).
49. Sansone, P. & Bromberg, J. Targeting the interleukin-6/Jak/stat pathway in human malignancies. *J. Clin. Oncol.* **30**, 1005–1014 (2012).
50. Block, K. M., Hanke, N. T., Maine, E. A. & Baker, A. F. IL-6 stimulates STAT3 and Pim-1 kinase in pancreatic cancer cell lines. *Pancreas* **41**, 773–781 (2012).
51. Zappia, L., Hipson, B. & Oshlack, A. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol.* **18**, 1–15 (2017).
52. Johri, S. *Vanallenlab/beanie: v1.0.0*. (Zenodo). <https://doi.org/10.5281/ZENODO.14617303> (2025).
53. Johri, S. *Vanallenlab/beanie-Analysis: v1.0.0*. (Zenodo). <https://doi.org/10.5281/ZENODO.14616985> (2025).

Acknowledgements

This project was funded by NIH U01 CA233100 (E.M.V.A. and L.F.), R01CA227388 (E.M.V.A.), U2CCA233195 (E.M.V.A.), DOD W81XWH-21-PCR-PDSA (E.M.V.A.), PCF-Movember Challenge Award (E.M.V.A.). S.J. is supported by the 2023 Quad Fellowship. J.P. is supported by R50CA265182.

Author contributions

E.M.V.A. and M.X.H. conceived the original idea. S.J. developed the idea further, designed experiments, performed the analyses, and developed the Python package for the presented method. K.B., B.M.T., J.F., and M.X.H. helped in development of the method from a single-cell perspective. M.X.H., J.F., J.C. and D.L., provided input from a statistical perspective. J.P.C. provided input for the development of visualization modules in the package. N.I.V. and D.L. helped in clinical interpretation of the results. Z.F., J.P., L.F., D.L., M.X.H., and E.M.V.A. contributed to the overall analyses. S.J., B.M.T., K.B., and E.M.V.A. wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

E.M.V.A. has received research support (to institution) from Novartis and BMS. E.M.V.A. serves as a consultant or on scientific advisory boards of Tango Therapeutics, Genome Medical, Invitae, Enara Bio, Janssen, Manifold Bio, Monte Rosa. E.M.V.A. has equity in Tango Therapeutics, Genome Medical, Syapse, Enara Bio, Manifold Bio, Microsoft, Monte Rosa. E.M.V.A. receives travel reimbursements from Roche/Genentech. E.M.V.A. has filed institutional patents on chromatin mutations and immunotherapy response, and methods for clinical interpretation, and has intermittent legal consulting on patents for Foale & Hoag. L.F. has received research support (to institution) from Roche/Genentech, Abbvie, Bavarian Nordic, Bristol Myers Squibb, Dendreon, Janssen, Merck, and Partner Therapeutics. L.F. has served on the scientific advisory boards of Actym, Allector, Astra Zeneca, Atreca, Bioalta, Bolt, Bristol Myer Squibb, Immunogenesis, Merck, Merck KGA, Nutcracker, RAPT, Scribe, Senti, Soteria, TeneoBio, and Roche/Genentech. M.X.H. is currently an employee of Genentech/Roche and has been a consultant to Amplify Medicines, Ikena Oncology, and Janssen. Work in this manuscript was performed when M.X.H. was a PhD student in the Biophysics program at Harvard Medical School. J.C. has been a consultant to Tango Therapeutics. He is also currently an employee of PathAI. N.I.V. serves on advisory boards of Sanofi and Oncocyte. The remaining Authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-57377-6>.

Correspondence and requests for materials should be addressed to Eliezer M. Van Allen.

Peer review information *Nature Communications* thanks Alessandro Laganà, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025