PLOS ONE

RESEARCH ARTICLE

Transcriptome analysis reveals high tumor heterogeneity with respect to re-activation of stemness and proliferation programs

Artem Baranovsky[1,2], **Timofei Ivanov**[1], **Marina Granovskaya**[3], **Dmitri Papatsenko**[1†], **Dmitri D. Pervouchine**[1*]

1 Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia, **2** Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany, **3** Research and Development, Jenguro Ingal, Moscow, Russia

† Deceased.
* d.pervouchine@skoltech.ru

Abstract

Significant alterations in signaling pathways and transcriptional regulatory programs together represent major hallmarks of many cancers. These, among all, include the reactivation of stemness, which is registered by the expression of pathways that are active in the embryonic stem cells (ESCs). Here, we assembled gene sets that reflect the stemness and proliferation signatures and used them to analyze a large panel of RNA-seq data from The Cancer Genome Atlas (TCGA) Consortium in order to specifically assess the expression of stemness-related and proliferation-related genes across a collection of different tumor types. We introduced a metric that captures the collective similarity of the expression profile of a tumor to that of ESCs, which showed that stemness and proliferation signatures vary greatly between different tumor types. We also observed a high degree of intertumoral heterogeneity in the expression of stemness- and proliferation-related genes, which was associated with increased hazard ratios in a fraction of tumors and mirrored by high intratumoral heterogeneity and a remarkable stemness capacity in metastatic lesions across cancer cells in single cell RNA-seq datasets. Taken together, these results indicate that the expression of stemness signatures is highly heterogeneous and cannot be used as a universal determinant of cancer. This calls into question the universal validity of diagnostic tests that are based on stem cell markers.

OPEN ACCESS

Citation: Baranovsky A, Ivanov T, Granovskaya M, Papatsenko D, Pervouchine DD (2022) Transcriptome analysis reveals high tumor heterogeneity with respect to re-activation of stemness and proliferation programs. PLoS ONE 17(5): e0268626. https://doi.org/10.1371/journal.pone.0268626

Editor: Sumitra Deb, Virginia Commonwealth University, UNITED STATES

Received: October 5, 2021

Accepted: May 3, 2022

Published: May 19, 2022

Copyright: © 2022 Baranovsky et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets analyzed during the current study are available in the GDC Data Portal repository, https://portal.gdc.cancer.gov/repository. The datasets generated during the current study are available online at https://doi.org/10.5281/zenodo.5101514. The source code used for the analysis is available at https://github.com/melonheader/bulk-cancer-stemness.

Introduction

Cancer is one of the major causes of death worldwide [1]. A significant progress in cancer treatment has been achieved with the development of a large therapeutic arsenal including chemotherapy, surgery, radiation therapy, and immunotherapy [2–4]. An important direction in clinical research focuses on the so-called Cancer Stem Cells (CSC), a subpopulation of tumorous cells with high tumor-initiating potential [5]. CSCs are increasingly regarded as

Funding: DPe acknowledges the Russian Science Foundation grant 21-64-00006. TI acknowledges the Russian Foundation for Basic Research grant 19-34-90174. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

prominent targets for anti-cancer therapy, although the degree of expression of the stem cell-like phenotype, referred to as stemness, may vary between different tumors [6].

Several hypotheses relate stemness with the origin of cancer. It is acknowledged that cancers arise either from a malignant transformation of a progenitor cell, or from a non-stem cell, which reacquired the stemness potential [7–9]. This paradigm is sustained by significant convergence of stem cells (SC) and CSCs in the activated signaling cascades, moreover in their overlapping expression of a set of biomarkers encompassing the classical self-renewal-associated pathways Wnt/β-catenin, Bmi-1, sonic hedgehog, Notch, and PTEN [10]. Additionally, both SCs and CSCs express tissue-specific stem cell markers [11–14]. Such concordant molecular profile stipulates key aspects of SC and CSC phenotype including longevity, dormancy, niche dependence, and the potential for asymmetric cell division [15–18].

Tumors display frequent inter- and intratumor heterogeneity in alterations of the global gene expression programs and, in particular, in the expression of stemness markers [19–22]. Commonly used CSC-associated markers have a variable expression in glioblastoma [23]. The expression of breast cancer stem cell markers, ALDH1A1 and CD133, shows spatial heterogeneity across patients [19]. Stemness phenotype is also heterogeneous among many normal adult SC populations in the human body, where the SCs uphold tissue regenerative capacity [24, 25]. Moreover, the degree of activation of stemness programs is associated with increased expression of multiple immunosuppressive pathways and decreased anticancer immunity [26]. This suggests that the stemness state may itself be highly heterogeneous within and between tumor types, which may play an important role in cancer pathogenesis.

A number of metrics have been developed to quantify stemness [20, 26, 27]. These metrics correlate with intratumoral heterogeneity, antitumor immune response, and clinical prognosis [20, 26, 27]. However, the heterogeneity of the stemness signature has not been assessed in detail [19]. Here, we reanalyzed the transcriptomes of 19 tumor types from The Cancer Genome Atlas (TCGA) Consortium and juxtaposed them with the transcriptomes of human embryonic stem cells [28], adult stem cells [29, 30], and induced pluripotent stem cells (iPSC) within 4 days of differentiation [31]. The comparison of these transcriptomes from the perspective of reactivation of the stemness program revealed a pronounced heterogeneity both within and between tumor types, which in many cases was associated with tumor-specific patient survival. To interrogate intratumoral heterogeneity, we additionally demonstrated an increased variability of stemness signature in cancer cells compared to non-cancer cells and in metastatic outgrowth compared to primary tumors, and convergence of cancer cells to stem cells using single cell transcriptomes.

Materials and methods

RNA-seq data and related clinical data

Gene expression values for tumors and normal tissues from the TCGA dataset and all relevant metadata were downloaded in the form of read counts using R package `TCGAbiolinks` [32] (S1 File). The fastq files for stem cell (SC) datasets comprising iPSC (fibroblasts purified from skin-punch biopsies from six males and six females that were reprogrammed using transfection with an episomal plasmids containing OCT3/4, SOX2, KLF4, L-MYC, LIN28, and an shRNA against p53 [33]), ESC, other types of pluripotent stem cells (PSC), and adult stem cells (ASC) were either downloaded from Sequence Read Archive (SRA) or obtained by direct download from arrayexpress ftp server (S1 Table). All fastq files were processed by a uniform pipeline, in which the reads were first checked for quality, trimmed using `TrimGalore`, and mapped with STAR version 2.7.1a [34] to the December 2013 assembly of the human genome (hg38, GRCh38). Gene expression levels were quantified using the Subread tool [35] implemented in Rsubread package version 1.34.7 [36]. The read count matrices from TCGA and stem cell datasets were merged, filtered by gene expression, and normalized using edgeR [37]. The resulting combined matrix was corrected for tumor purity as explained below.

The raw UMI counts for the scRNA-seq dataset (63,689 cells from 23 primary colorectal cancer and 10 matched normal mucosa samples) were downloaded from GEO under the accession GSE132465 [38]. The raw UMI counts for mESC and iPSC scRNA-seq datasets were downloaded from GEO and ArrayExpress under the accession numbers GSE135509 and E-MTAB-6687, respectively [39]. All counts were normalized for sequencing depth per cell and transformed to $\log_2$-scale using a pseudocount of one. The subsequent analysis and visualization of single cell RNA-seq data, including integration using a reciprocal PCA workflow, were done using R package Seurat version 4.0.2. Only sample 4 of iPSC dataset was used as it possessed the highest average stemness signature score. The lung adenocarcinoma scRNA-seq dataset with metastatic samples was accessed and processed following the guidelines provided by the authors of the original publication [40].

Correction for tumor purity

To mitigate the influence of normal cells on gene expression in tumor samples, we regressed out the effect of tumor purity defined by consensus tumor purity estimate (CPE). First, we downloaded a table of precomputed CPE values which is available as a Supplementary Data 1 in [41]. Next, we filtered out tumor samples where CPE was not available and built a collection of linear models of the form $Y_{i,s} \sim \beta_{i,0} + \beta_{i,1}{}^{*}P_s + \varepsilon_{i,s}$ individually for each gene i, where $Y_{i,s}$ is the TMM-normalized $\log_2(CPM)$ of the gene i in the sample s, and P_s is the value of the CPE of the sample s, and $\varepsilon_{i,s}$ is the error term. The purity-corrected TMM-normalized $\log_2(CPM)$ of gene i in the sample s were obtained from $Y_{i,s}$ by subtracting the linear part, i.e., $Y_{i,s}^{cor} = Y_{i,s} - \beta_{i,1} * P_s$. Since the CPE metric provides an estimate for the proportion of tumor cells, which are presumably absent in normal tissues, we assigned a random value close to zero from normal distribution with the mean 0.08 (0.05 percentile of CPE distribution) and standard deviation 0.03 ($\sim \frac{1}{3}$ of the 0.05 percentile of CPE distribution) to CPE of all normal tissues and iPSCs. The resulting values of $Y_{i,s}^{cor}$ showed no statistically significant dependence on CPE (mean $R^2 \simeq 0$).

Differential gene expression analysis

The edgeR package was used for differential expression analysis. Raw read counts were normalized by edgeR using the TMM (Trimmed Mean of M-values) method [37]. The model was fitted to the data using glmQLFit function. Differential gene expression was estimated with glmTreat function. For GO enrichment analysis, differentially expressed genes were defined by $|\log_2(FC)| \geq 1$ and adjusted p-value cutoff of 0.05.

Gene ontology analysis

Gene Ontology analysis of differentially expressed genes was done using R package clusterProfiler [42] with the default parameters.

Signature gene sets

The set of stemness gene signatures ($n = 271$) was originally designed by DPa by a combination of literature search and analysis of transcriptomic datasets. However, his efforts were interrupted by force majeure circumstances, and the rest of the team had to reassemble a similar set using another procedure outlined below.

We considered datasets spanning various experimental protocols, e.g. tissue-gene expression atlases, ESC differentiation time series, and knockdowns of pluripotency-associated transcription factors (ppTF-KDs) [43–46]. The datasets were analyzed one at a time, with individually selected cutoffs (S2 Table). The microarray gene expression data were downloaded using the R package `GEOquery` [47], normalized using the `rma` function available in the R package `affy` [48], $\log_2$-transformed and processed following the guidelines from the `affy` package vignette.

Our collection included two gene expression atlases, GSE1133 and GSE10246, both containing gene expression data from various mouse tissues and cell lines. We identified a group of samples, in which the stemness program could be active (S2 Table), and selected only one group in each atlas that was closest to ESCs (Blastocysts in GSE1133 and mouse ESCs in GSE10246). In each atlas, we computed $\log_2 FC$ between the selected group and the rest of tissues or cell lines and computed E, the log-sum of gene expression values across samples. Then, we selected genes that satisfied the following conditions in at least $n_{comp.}$ of the comparisons in each atlas: $\log_2 FC > 0.05$, $E > 3$, $n_{comp.} = 34$ for GSE1133, and $\log_2 FC > 0.1$, $E > 4.5$, $n_{comp.} = 49$ for GSE10246. This resulted in two lists of putative stemness genes with 5764 and 2670 elements for GSE1133 and GSE10246, respectively.

Next, we analyzed a collection of ESC-line differentiation datasets, which included time series analysis of 14 days of differentiation in three mouse ESC-lines: J1, R1 and v6.5. As before, we computed E and $\log_2 FC$ between the first (0 hours) and last (14 days) time points, separately for each cell line. The cutoffs $\log_2 FC > 0.05$ and $E > 3$ resulted in three lists of 3286, 3567, and 2609 putative stemness genes corresponding to J1, R1 and v6.5 lines, respectively. A similar analysis of shRNA knockdowns of five core pluripotency transcription factors, *POU5F1*, *NANOG*, *SOX2*, *ESRRB* and *SALL4*, in mouse ESCs resulted in 8588 genes with $\log_2 FC > 0.05$ and $E > 3$ in at least one of the experiments.

The intersection of these lists consisted of $n = 454$ genes, to which we added a set of manually collected and curated putative stemness genes, which were not picked by our analysis ($n = 19$) (S2 File). Then, we removed tissue-specific genes [49] ($n = 7$), proliferation signature genes ($n = 49$, see below), and genes that were functionally associated with the cell cycle or proliferation according to either KEGG or MsigDB databases ($n = 28$). This resulted in a list of 389 unique murine genes, which were mapped to their human orthologs using R package `BiomaRt`. The details of this procedure including cutoffs and group comparisons are also summarized in S2 Table.

To define the proliferation signatures, we used the list of genes that are functionally associated with proliferation as provided by Ben-Porath *et al* ($n = 326$) [50]. To infer activity of epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-epithelial transition (MET), we selected TFs associated with either of the two processes, *SNAI1* and *SNAI2* for EMT, and *GRHL1–3* for MET. Then, we combined the selected TFs with their protein interactors and co-expressed genes according to STRING-DB (v.11.0) [51] resulting in a list of $n = 51$ genes. The resulting gene lists are summarized in S3 File.

Calculation of signature intensity metric

To compare the degree of reactivation of the stemness program between different tumors, we introduced a metric called *signature intensity*, which represents the percentage of samples that belong to the Stem cluster among all samples of the given tumor. Recall that a sample was assigned to the Stem cluster (respectively, Normal cluster) if it was located closer to (respectively, further from) ESC/iPSCs according to PC1 than the global median of the PC1 axis. That is, the signature intensity I_j of the tumor j was computed as $I_j = N_j^{Stem}/(N_j^{Stem} + N_j^{Normal})$, where

N_j^{Stem} is the number of samples of the tumor type j in the Stem cluster and N_j^{Normal} is the number of samples of the tumor type j in the Normal cluster.

Survival analysis

To test whether the reactivation of the stemness program correlates with poorer prognosis, we stratified tumor samples into two clusters relative to the median PC1 value of a given tumor type. For a given tumor we subdivided samples into Stem and Normal clusters based on the proximity to iPSCs on PC1. However, instead of using the global median PC1 over all tumors as a threshold, we used the median PC1 of each tumor type. This approach resulted in tumor-specific clusters of balanced size, which were used for survival analysis. We fitted Cox-regression using Stem and Normal clusters as predictors for each tumor when clinical data were available. Hazard ratios (HR) between two clusters were then computed for all tumors using R package `survminer`.

Statistical analysis

All statistical analyses were done using R software version 3.6.0. Confidence intervals for proportions were computed using a 2-sample z-test without continuity correction. All tests were carried out at the 5% significance level with Benjamini-Hochberg correction for multiple testing.

Results

Stemness genes

Several lists of stemness marker genes currently exist, however there is no universal such list [52]. For instance, a recent meta-analysis of stemness genes from several independent studies revealed that only one gene was common among lists derived from three SC populations [53]. Complementary to this, here we assembled a list of pluripotency markers using the data from a range of different experimental approaches including tissue atlases, ESC differentiation time-series, and knockdowns of pluripotency-associated transcriptional factors [43–46]. To distinguish between stemness and other traits, we removed proliferation-related genes and tissue-specific genes [49] (see Methods for details) and obtained a list of (n = 389) stemness genes including master regulators of stemness *POU5F1*, *SALL4*, *NANOG* as well as many other genes involved in the transcriptional regulation and cellular metabolism (S3 File) [54]. The comparison of this set with the ESC signatures provided in other studies [50, 55, 56] revealed a moderate intersection (Fig 1), however not as large as the intersection of ESC signatures with each other. For instance, the gene set of Bhattacharya *et al* overlapped by more than a half with the ESC signature identified by Ben-Porath *et al* (n = 50), while the largest intersection of our set was with the gene set of Wong *et al* (n = 45).

Since tumors are composed of heterogeneous populations of cell types, an admixture of normal cells into tumor samples can influence global gene expression profiles and dilute stemness signature with the signal from normal cells, in which the stemness program is silent. An estimate for the cellular composition of a tumor sample is the so-called tumor purity, which can be quantitatively measured as the percentage of true tumor cells among multiple other cell types that constitute the tumor [41]. Several approaches have been developed to estimate this metric [41, 57–60]. They were recently applied to the TCGA dataset to compute the consensus tumor purity estimate (CPE), which represents the median of four different tumor purity estimates [41]. The distribution of CPE varies greatly both between and within different tumor types (S1 Fig). To account for the effect of cellular composition bias, we built a linear model

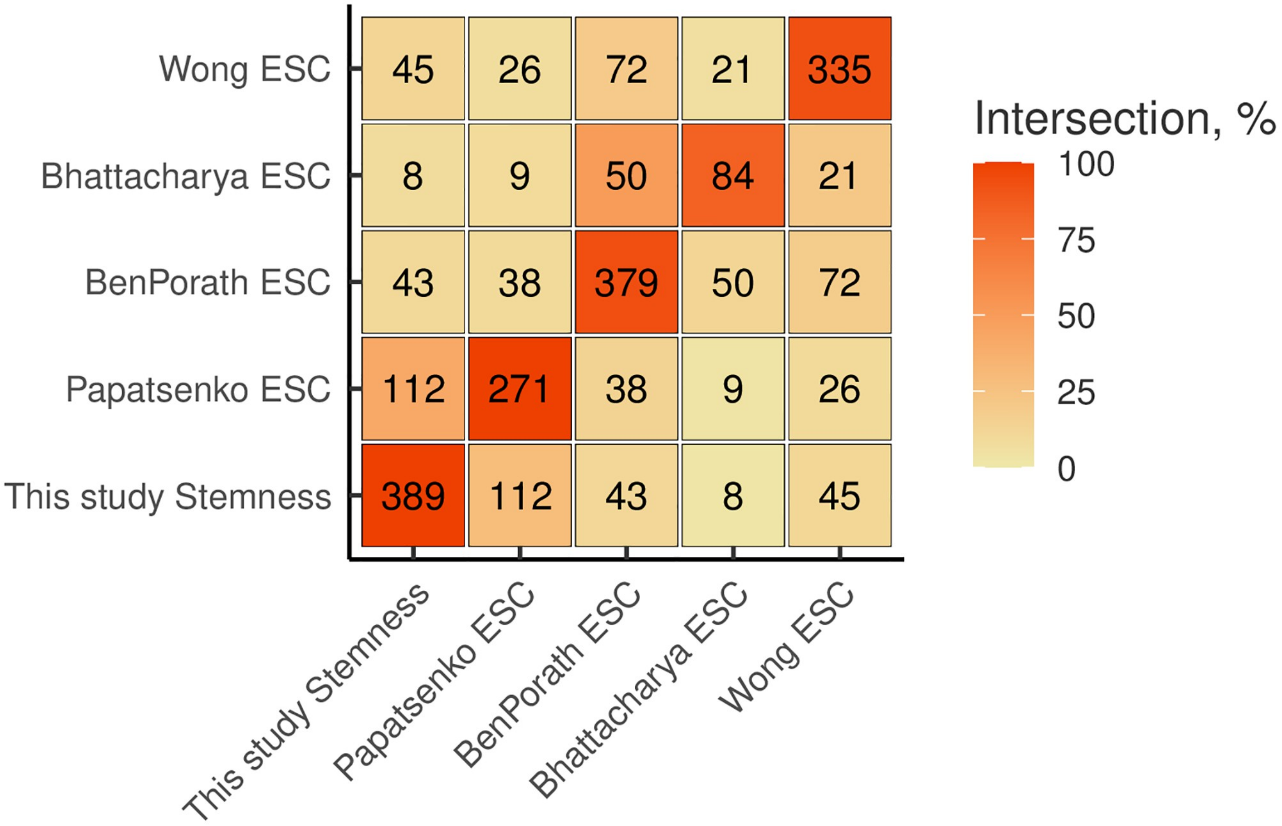


Fig 1. The stemness signature gene set intersects with previously published ESC signature gene sets. The number of genes in the intersection between sets on the X and Y axes is shown in the cells of the matrix. The color scale encodes the size of the intersection relative to the size of the dataset on the Y axis (above the diagonal) and relative to the size of the dataset on the X axis (below the diagonal).

https://doi.org/10.1371/journal.pone.0268626.g001

(see Methods) of gene expression as a function of CPE and used the residuals of this model instead of the raw gene expression values (S2 Fig).

Tumors span continuously from normal tissues to iPSC

To investigate how tumors, normal tissues, ESC, ASC, and iPSCs compare in terms of stemness signature, we used principal component analysis (PCA) of the adjusted gene expression profiles restricted to the set of stemness genes (Fig 2A). Tumor samples distributed over the first principal component (PC1, 24% of the total variance) forming two distinct clusters located between normal tissues and ESCs/iPSCs, while ASCs were located close to the normal tissues on the PC1 axis. At that, the iPSCs differentiation time series also clustered along PC1 so that less differentiated cells were located further from the center. Remarkably, the second principal component (PC2, 8% of the total variance) separated iPSCs and ESCs. The higher-order principal components did not show any clear separation related to stemness genes (S3 Fig), indicating that stemness signature is encoded within PC1. A similar clustering by stemness signatures from previously published gene sets did not provide a clear separation of tumors, normal samples and SCs (S4 Fig).

The expression of particular stemness markers followed the pattern of clustering along PC1 (S5 Fig). *SALL4*, one of the major regulators of pluripotency, was among the genes with the highest loadings in PC1, with a gradual increase in expression along the PC1 axis reaching its maximum at iPSCs. On the other hand, *SALL1*, which is suppressed in breast cancer [61], was

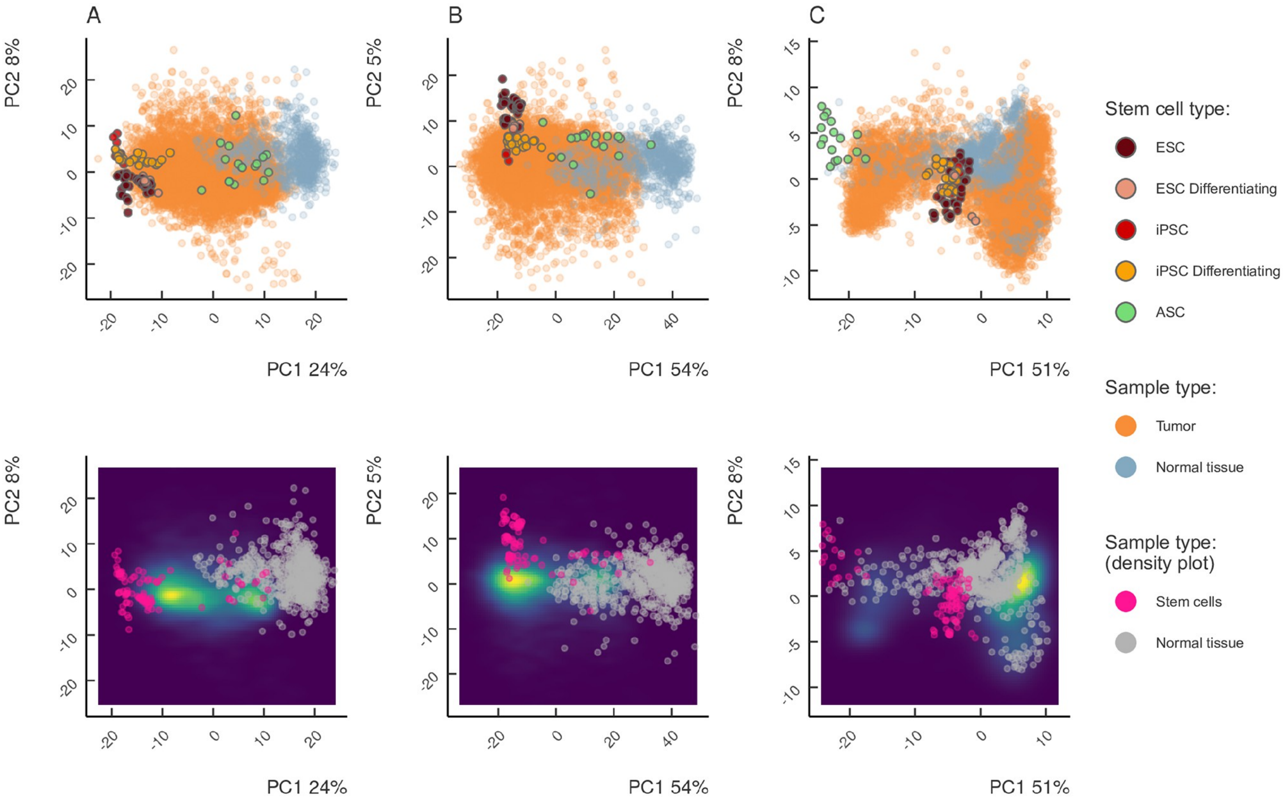


Fig 2. Principal Component Analysis (PCA) of the expression of stemness signatures (A), proliferation signatures (B), and genes representing EMT-MET (C) in differentiating iPSCs, ESCs, ASCs, tumors and normal tissues shown as individual samples (top) and density (bottom). Tumors span continuously between normal tissues and iPSCs in the expression subspaces of stemness, proliferation, but not EMT-MET genes. Tumor samples form two clusters: one closer to ESCs and iPSCs, another closer to the normal tissues and ASCs.

https://doi.org/10.1371/journal.pone.0268626.g002

ubiquitously downregulated in tumors, while being steadily expressed in both stem cells and normal tissues. The expression of *POU5F1*, a key regulator of stem cell pluripotency, substantially increased, while the expression of *CBX7*, a gene that encodes a Polycomb protein which globally regulates cellular lifespan [62] and is a tumor suppressor in both mice and humans [63], decreased towards ESC/iPSCs (S5 Fig).

To check whether the observed clustering along PC1 axis was consistent with patterns reported for particular tumor types, we analyzed the stemness signature in two related, yet different cancers, lung squamous cell carcinoma (LUSC) [64] and lung adenocarcinoma (LUAD) [65]. LUSC has a higher mortality rate compared to LUAD and shows a pronounced upregulation of *sonic-hedgehog*, a major regulator of the developmental pathway linked to stemness and proliferation, which is often active in adult stem cells and is absent in LUAD [66, 67]. Consistent with this, LUSC was located closer to SC along the PC1 axis, while LUAD was closer to normal tissues (S6A Fig). In line with previous observations, we observed an increased expression of the pluripotency-associated transcription factor SOX2 and reduced expression of CBX7 in LUSC (S6B and S6C Fig). Additionally, genes differentially expressed in LUSC were enriched for development-associated GO terms (S7 Fig).

Analysis of proliferation and EMT signatures

The expression of stemness genes revealed clustering of tumor samples in between normal tissues and ESC/iPSCs, which may not be a unique property of stemness genes. However, no clustering was observed when we performed PCA on a random set of genes that were matched by expression levels to stemness genes (S8 Fig). In contrast, when we repeated the same analysis using proliferation markers, tumor samples again scattered across the PC1 axis (55% of variance) forming two separate clusters (Fig 2B), and the differentiation states of iPSCs separated concordantly along this axis. As before, PC2 separated ESCs from iPSCs, while no clear separation was observed in higher order principal components suggesting that PC1 represents the proliferation signature (S3B Fig).

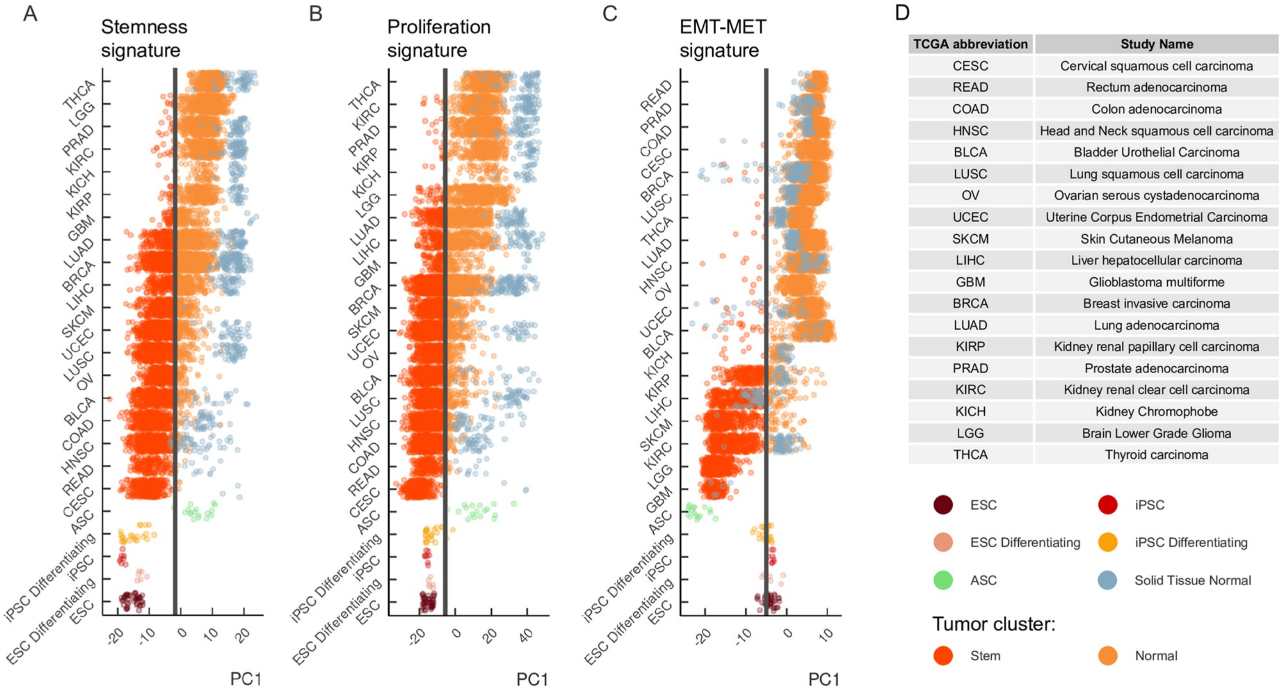
In contrast, the clustering of tumors, normal tissues, and iPSCs with respect to the epithelial to mesenchymal transition (EMT) signature was drastically different (Fig 2C). Except for the marginal standing of ASC, in which EMT must be active, we did not observe any clear separation of tumor samples from normal tissues, nor did we detect a directional trend of tumor samples along any axes. There was no clear separation in higher-order principal components, and most of the tumors showed epithelial phenotype similar to their tissues of origin, except for hepatocellular carcinoma, melanoma and brain tumors (S3C Fig). Instead of the gradient of EMT signature, we observed a switch-like state of master regulators of EMT (S9 Fig). Altogether, this analysis failed to capture EMT signature in tumor tissue, possibly because cells undergoing EMT are located at metastatic outgrowth or even dispersed as a transient circulating population thus preventing their detection in bulk sequencing [68].

Heterogeneity of stemness and proliferation signatures across tumor types

The analysis of stemness and proliferation signatures revealed two clusters of tumor samples, one of which was located closer to iPSCs along PC1, and the other was located closer to normal tissues (Fig 2A and 2B). Since this separation occurred along the PC1 axis in both signatures, we formally defined the clusters by their position relative to the median of the PC1 across all tumor samples (Fig 3, black vertical line). Namely, samples located to the right of the median of PC1 axis towards normal tissues were assigned to the *Normal* cluster, while samples located to the left of the median towards iPSC were assigned to *Stem* cluster.

Next, we investigated whether any tumor type was specifically enriched in one of the two clusters (Fig 3A–3C). Indeed, some tumors were fully localized to one of the clusters, e.g. Rectal Adenocarcinoma and all subtypes of Renal Cell Carcinoma (KIRP, KIRC, KICH), while others showed a significant heterogeneity in terms of both stemness and proliferation signatures, e.g. Lung Adenocarcinoma and Liver Hepatocellular Carcinoma. To quantify the heterogeneity of stemness and proliferation signatures within each tumor type, we computed a metric called *intensity*, which is defined as the proportion of samples of a given tumor type that are located in the Stem cluster (see Methods for details). In other words, this metric captures the degree, to which each tumor type collectively expresses stemness or proliferation signatures.

Despite stemness and proliferation intensities being strongly correlated, they differ significantly for some tumor types (Fig 4A and 4B). For instance, the stemness intensity of glioblastoma (GBM) is 28%±7%, while its proliferation intensity is 50%±8%. Conversely, the stemness intensity of hepatocellular carcinoma (LIHC) is 46%±5%, while its proliferation intensity is 31%±5%. Thus, these two metrics capture consistent, yet different aspects of tumor gene expression. Most tumors with high purity have low stemness and proliferation intensities, consistent with lower aggressiveness of such tumors [41], while tumors of higher stage tend to have higher intensity of both signatures. Remarkably, tumors that arise from tissues with high regenerative capacity have higher intensity of both signatures, presumably reflecting the innate stem cell populations that maintain homeostasis.



TCGA abbreviation	Study Name
CESC	Cervical squamous cell carcinoma
READ	Rectum adenocarcinoma
COAD	Colon adenocarcinoma
HNSC	Head and Neck squamous cell carcinoma
BLCA	Bladder Urothelial Carcinoma
LUSC	Lung squamous cell carcinoma
OV	Ovarian serous cystadenocarcinoma
UCEC	Uterine Corpus Endometrial Carcinoma
SKCM	Skin Cutaneous Melanoma
LIHC	Liver hepatocellular carcinoma
GBM	Glioblastoma multiforme
BRCA	Breast invasive carcinoma
LUAD	Lung adenocarcinoma
KIRP	Kidney renal papillary cell carcinoma
PRAD	Prostate adenocarcinoma
KIRC	Kidney renal clear cell carcinoma
KICH	Kidney Chromophobe
LGG	Brain Lower Grade Glioma
THCA	Thyroid carcinoma

Fig 3. Heterogeneity of stemness and proliferation signatures across tumor types. The first principal component (PC1) encodes the signature of stemness (A), proliferation (B), and EMT-MET signature genes (C). The samples are plotted separately for each tumor in the TCGA project, normal tissues, and iPSCs. Tumor samples are colored according to their position relative to the PC1 median (black vertical line). Tumor samples located to the right (left) of the median are assigned to the Normal (Stem) cluster, respectively. (D) The standard TCGA notation for cancer types.

https://doi.org/10.1371/journal.pone.0268626.g003

Intertumor variability correlates with survival

The intensity of stemness and proliferation signatures separates most tumors into two groups corresponding to Normal and Stem clusters by the median of the PC1 axis. However, this subdivision reflects global heterogeneity of stemness among all analyzed tumor types. In order to assess the intertumor heterogeneity, i.e., the heterogeneity between different tumors of the same tumor type, we used the PC1 median of a given tumor type rather than the global PC1 median across all tumor samples to cluster samples into Stem and Normal groups specific to each tumor type (Fig 5A). This approach provided balanced groups, which we used for survival analysis.

We observed a significant difference in survival between the Stem and Normal clusters only in a fraction of tumors (Fig 5B). Remarkably, the hazard ratio was significant for tumors that did not show the highest stemness and proliferation intensity (the bottom left quadrant in Fig 4), while tumors with the highest stemness and proliferation intensity (the upper right quadrant in Fig 4) did not show a significant difference in survival. For instance, we did not observe a significant difference in hazard ratio for LUSC despite high stemness and proliferation intensity, while skin cutaneous melanoma (SKCM), which had the stemness intensity on the level of LUSC, showed a significant difference in the survival between Stem and Normal clusters (Fig 6). A similar heterogeneity was observed for tumors with low stemness intensity such as kidney renal carcinoma (KIRC) and low grade glioma (LGG).

Intratumor heterogeneity

Unlike intertumor variability, which reflects differences between different tumors of the same type from different patients, intratumor variability refers to genotypic and phenotypic

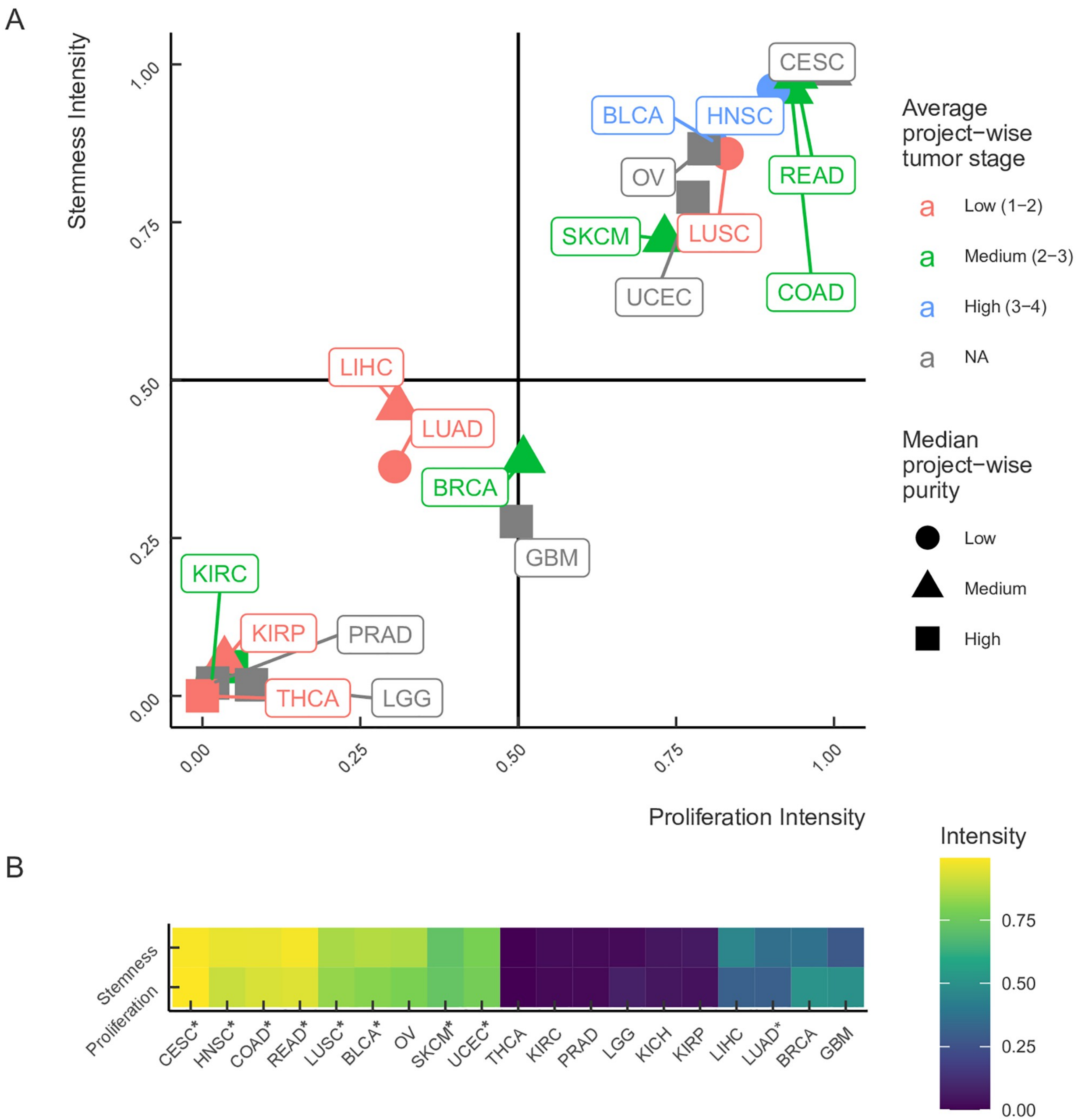


Fig 4. The relationship between stemness and proliferation intensities. (A) Stemness intensity vs. proliferation intensity in 19 solid tumors with different average project-wise tumor stage and median project-wise purity. (B) A heatmap of stemness and proliferation intensities across 19 solid tumors from TCGA. Two groups of tumor samples are identified: top-right quadrant (stemness Intensity>0.5 and proliferation Intensity>0.5) and bottom-left quadrant (stemness Intensity<0.5 and proliferation Intensity<0.5). The notation for cancer types as in Fig 3. Tumors from stem cell rich organs are marked with asterisks.

https://doi.org/10.1371/journal.pone.0268626.g004

differences between clonal populations of cells within a tumor. The quantitative measurement of the transcriptional diversity of cells within a tumor are provided by transcriptomic profiling at a single-cell resolution [69]. To assess the intratumor heterogeneity of stemness signature, we re-analyzed the transcriptomes of 65,362 unsorted single cells from metastatic colorectal

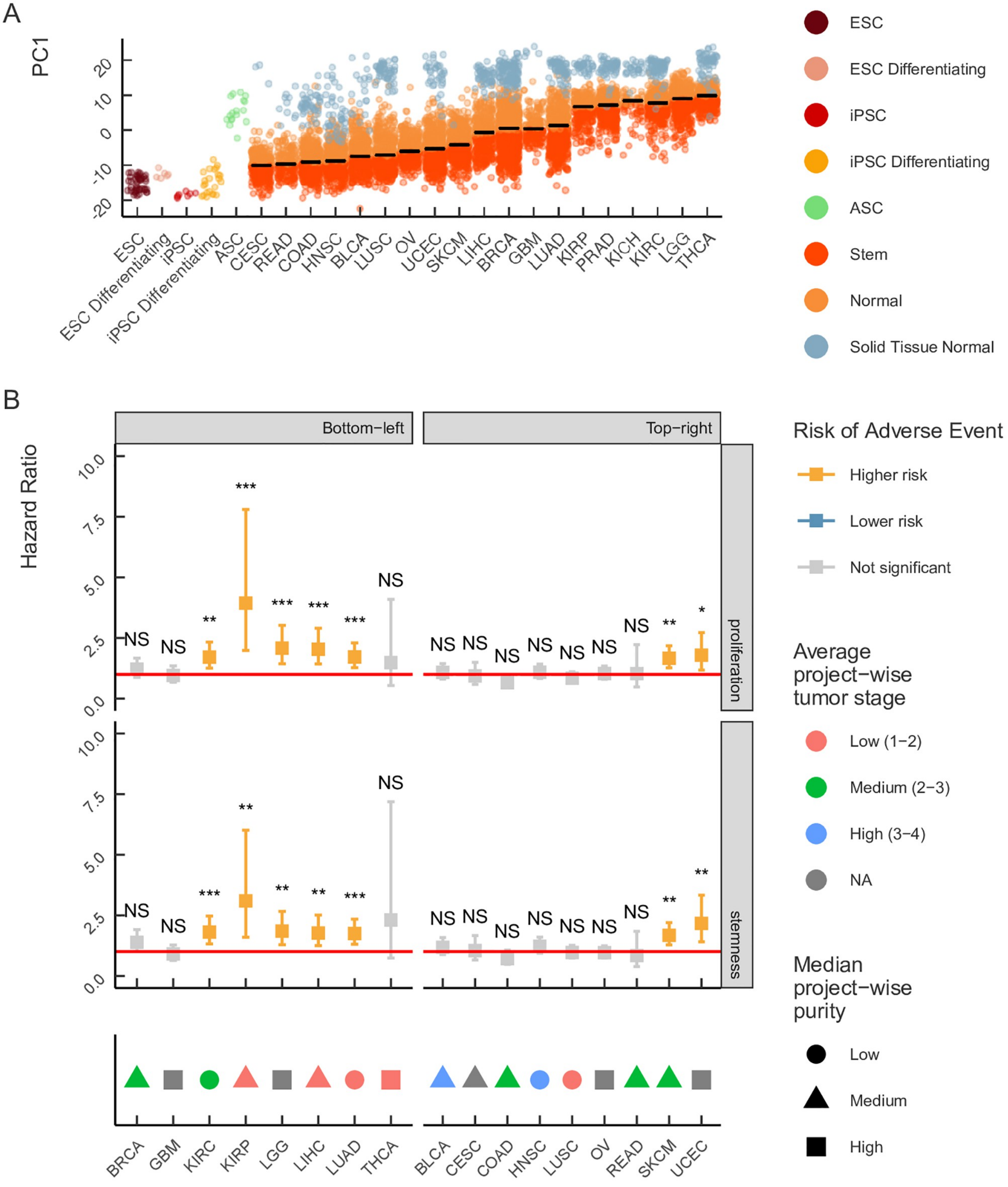


Fig 5. Intertumor heterogeneity of stemness and proliferation signatures. (A) Stem and Normal clusters specific to each tumor type are defined by the location of a sample relative to the median PC1 value across all samples of the given tumor (plotted as horizontal black lines). Samples above (below) the median are assigned to the Normal (Stem) cluster, respectively. (B) Hazard ratios for the comparison of patient survival in Stem and Normal clusters in two sample groups (top-right and bottom-left in Fig 4) with respect to stemness and proliferation intensity. Error bars denote 95% confidence intervals. Tumor types are annotated at the bottom according to the average tumor stage and the median tumor purity. Codes for statistical significance reported are as follows: NS – p-value >0.05, * – 0.01< p-value <0.05, ** – 0.001< p-value <0.01, *** – p-value <0.001. The notation for cancer types as in Fig 3.

https://doi.org/10.1371/journal.pone.0268626.g005

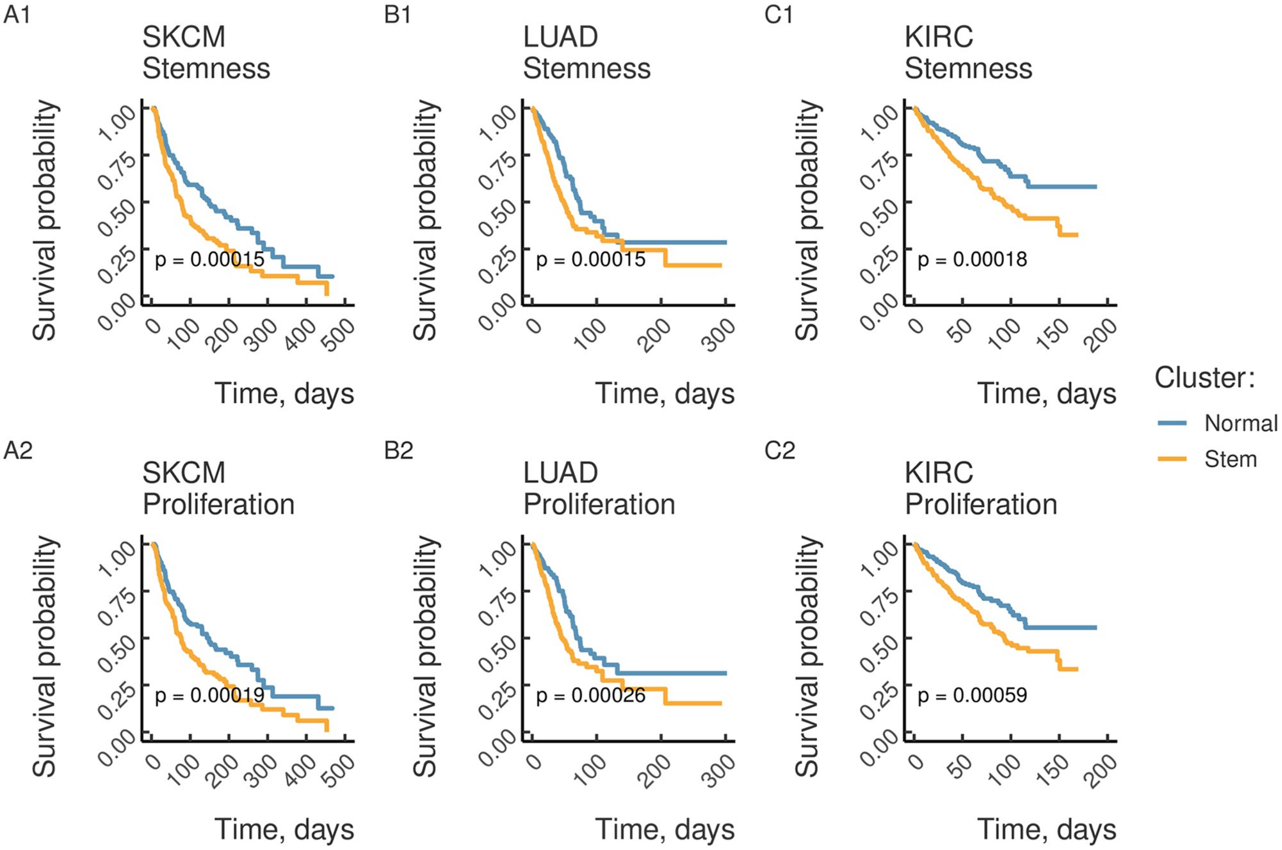


Fig 6. Intertumor heterogeneity of stemness and proliferation is predictive of poor survival. Kaplan-Meier survival curves grouped by clustering according to stemness (proliferation) signature are plotted for skin melanomas (A), lung adenocarcinoma (B) and renal clear cell carcinoma (C). P-values from the logrank test are reported. The notation for cancer types as in Fig 3.

https://doi.org/10.1371/journal.pone.0268626.g006

cancers [38]. We used the loadings, i.e., the coefficients of the linear combination of stemness signature genes, that correspond to the PC1 axis in Fig 2A to project the transcriptional profiles of single cancer and non-cancer cells onto PC1 (Methods). The 2D representation using uniform manifold approximation and projection (UMAP) [70] revealed a clear separation of cancer and non-cancer cells consistent with the annotation from [38] (Fig 7). Remarkably, the cancer cell cluster shows a considerable increase of the stemness signature encoded within PC1 (Fig 7A). However, not only the absolute value of the stemness signature, but also its variability was higher in the cancer cell cluster compared to non-cancer cell cluster indicating that the heterogeneity of stemness signature is observed at all levels of cancer organization, including single-cell level.

To compare the stemness capacity among individual tumor cells and stem cells, we integrated two additional scRNA-seq datasets of mESC [71] and human iPSC [72] cultures along with scRNA-seq dataset of colorectal cancer. We computed the PC1 projection for each cell using PCA loadings from Fig 2 as a quantitative proxy of stemness, proliferation and EMT signatures. As expected, we observed a gradient of stemness signatures from normal to cancer cells and then to stem cells, which encompass both adult stem cells residing in the niches of the normal tissues (labelled ASC) and iPSCs with ESCs (Fig 8A). The proliferation signatures mirrored the stemness signatures with a notable exception of iPSCs, which had stemness signatures below ASC while being proliferative more active. The EMT signatures were distributed

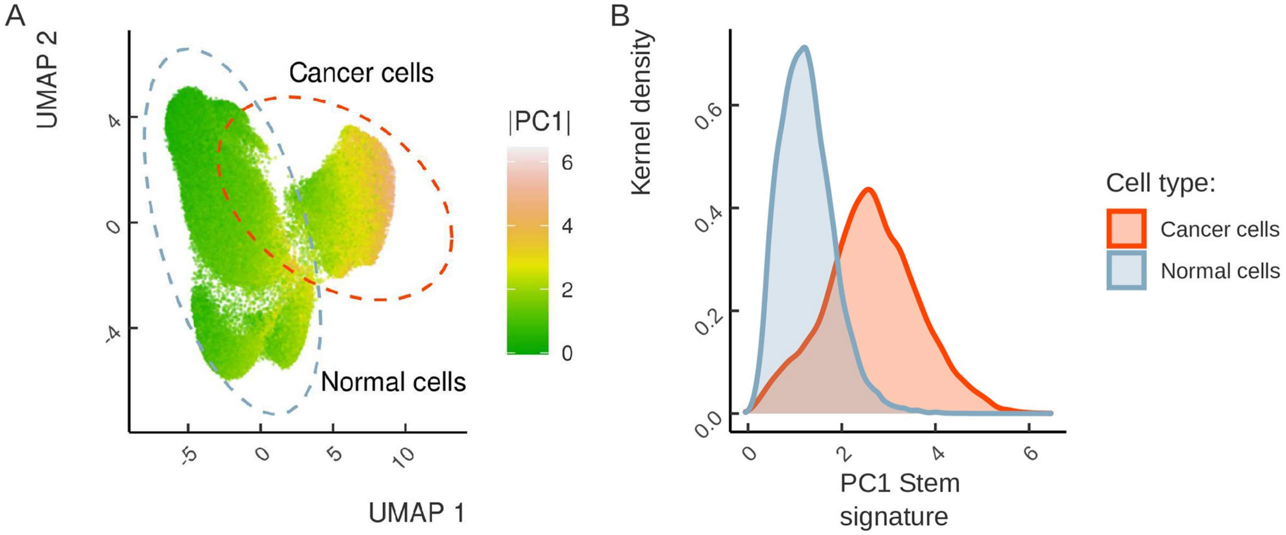


Fig 7. Cancer cells show substantial intratumor heterogeneity of stemness. (A) UMAP dimensionality reduction of the colorectal cancer single-cell RNA-seq dataset. The annotation of normal and cancer cell clusters is from [38]. The cells are colored by the degree of stemness signature reactivation (the absolute value of PC1 projection). (B) The distribution of stemness signature (PC1) (X axis) in normal and cancer cells.

https://doi.org/10.1371/journal.pone.0268626.g007

bimodally with cells of positive scores having more mesenchymal expression characteristics. Remarkably, while the majority of cancer cells had negative scores, presumably reflecting the epithelial origin of colorectal cancer, a large fraction (25%) likely containing CSCs still had mesenchymal signature. The reduction to two dimensions with UMAP revealed two clusters with high overall stemness signatures, one that corresponded to ESC and another to iPSCs, which both converged into a more differentiated state represented by a cluster that was shared between ESCs and iPSCs (Fig 8B).

It was reported that cancer cells comprising metastatic outgrowths possess prominent stem-like characteristics [73]. To investigate this trait of cancer cell stemness, we computed PC1 projections for each cell in a scRNA-seq dataset that contains normal tissues, primary tumors and metastatic lesions of lung adenocarcinoma [40]. Indeed, single cells from these three categories showed significantly increased stemness signatures in metastatic lesions in comparison to both primary tumors and normal tissues (Fig 8C).

Next, we questioned how the heterogeneity of stemness signature compares between individual cells and between tumors that arise in different patients. To address this, we reanalyzed stemness signatures of individual cells in the colorectal cancer cohort taking the patient information into account and found that a substantial proportion of variance of stemness signature (13.5%, 1-way ANOVA, $P\text{-}value < 10^{-16}$) is attributed to intertumoral, as opposed to intratumoral heterogeneity. Remarkably, cancer cells in some patients (SMC10, SMC08) were bimodally distributed along the stemness signature score, potentially suggesting a presence of cancer cell subpopulations of different potency (Fig 8D).

Discussion

Transcriptome comparisons have been a powerful tool for studying gene expression signatures in disease [74–76]. We used principal component analysis (PCA), which represents each sample as a point in an n-dimensional space with coordinates corresponding to gene expression values and applies a dimensionality reduction to identify principal components with the largest variance. The outcome of these transformations, however, critically depends on the set of

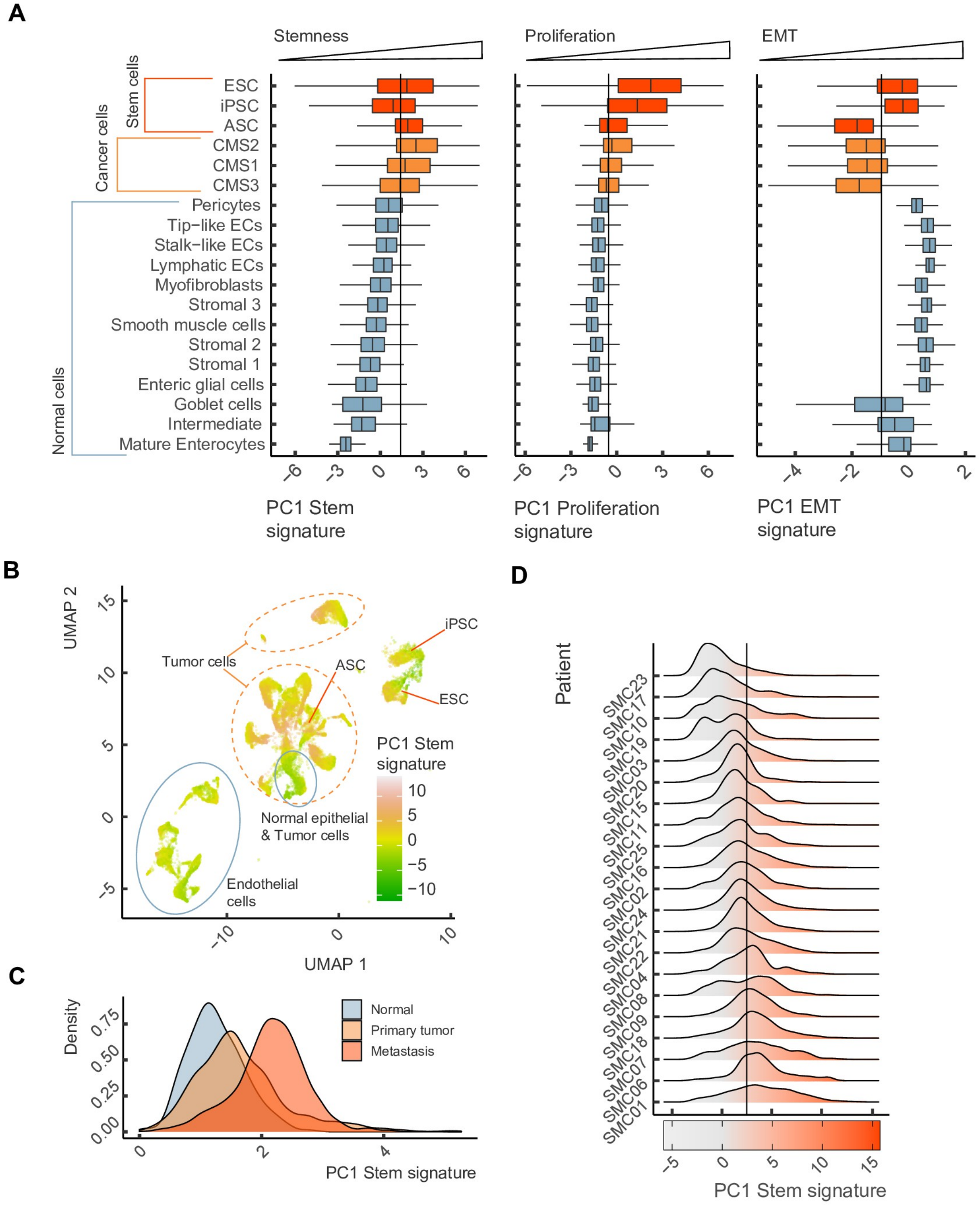


Fig 8. Cancer cells show heterogeneity in stemness, proliferation and EMT scores. (A) Stemness, proliferation and EMT scores in single cells by cell type. Boxplots colors correspond to normal (blue), colorectal cancer (orange), and stem (red) cells. (B) Two dimensional embedding derived from UMAP colored by the stemness score from panel A. (C) Stemness scores computed for single cells from primary lung adenocarcinomas (orange), metastatic lesions (red), and surrounding normal tissues (blue). (D) Stemness scores of colorectal cancer cells from different patients.

https://doi.org/10.1371/journal.pone.0268626.g008

genes that were chosen initially. For instance, the use of different gene sets led to opposite outcomes with tissue-dominated and species-dominated clustering [77]. In this work, we identified a gene set corresponding to the stemness signature encoded in PC1, which followed the stemness gene expression gradient from the normal tissues through tumors and ASC to ESC and iPSCs. This gene set provides a better representation of the stemness axis compared to previously published gene sets [50, 55, 56].

The TCGA dataset showed a consistent reactivation of stemness in all solid tumors in comparison to normal tissues, in accordance with other reports [78–80]. At the same time, the degree of reactivation of the stemness and proliferation signatures varied greatly between different tumor types and also inter- and intratumorally, thus extending the results of previous studies on stemness heterogeneity to the TCGA dataset [27, 81]. Remarkably, tumors originating from the tissues that actively interact with the outside environment (lungs, urinary system, skin, intestines) show strong reactivation of both stemness and proliferation programs consistently with the hypothesis that tumors inherit their self-renewal capacity from the tissues of origin. The presence and abundance of innate stem cell populations is an important factor that contributes to the heterogeneity of stemness signatures between different tumor types.

An important aspect of the reactivation of gene expression programs in tumors is formulated by the Lineage Addiction Model, which suggests that the mechanisms that promote tumor progression involve master regulatory genes that also exert key survival roles in development [82]. Multiple examples of cancer lineage addiction have been reported [83–86], however the effect of lineage addiction on the reactivation of the stemness program remains to be studied in detail. In our analysis, only a few tumor types (Breast Cancer, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma) spanned the entire PC1 axis, while most tumor types were fully localized to either Stem or Normal cluster. This observation supports the idea that the majority of tumors are restricted to specific phenotypic spaces with respect to the reactivation of the stemness program.

In the study of Ben-Porath *et al*, high expression of stemness genes was shown to be predictive of poor survival among the patients of three breast cancer cohorts [50]. Other studies also reported reduced survival for the tumors of stem phenotype [87–90]. Overall, as one of the hallmarks of cancer [91], stemness has been suggested before as a universal predictor of patient survival [27]. In our analysis, however, only seven out of sixteen tumors, for which patient survival data were available, showed a significant negative correlation of stemness with survival. These results highlight the absence of universal ties between patient survival and the degree of stemness signature reactivation, indicating a substantial degree of heterogeneity in tumor reaction to the latter [26]. This analysis suggests that in spite of associations with tumor stage and aggressiveness, the stemness is not a universal predictor of survival and could be regarded instead as a function of the molecular profile that is specific for every tumor type.

According to the cancer stem cell hypothesis, cancers arise from transformation of stem or progenitor cells that are capable of multilineage differentiation. The analysis of CSCs in scRNA-seq data cannot be carried out directly since their identities are not known. However, heterogeneous stemness, proliferation, and EMT signatures may partially serve as indicators of CSCs in a large pool of cells. We observed a larger heterogeneity of stemness signature both in cancer cells and stem cells grown in culture, not including ASC, presumably due to inherent transcriptomic instability of cancer cells that can potentially reach extremes in both differentiation and dedifferentiation. The heterogeneity of stemness signature in iPSCs and ESCs seems to arise from stochastic differentiation processes that are known to occur in cell culture [92, 93]. In this regard, transcriptomic instability of cancer cells that occasionally leads to stochastic dedifferentiation could serve as a source of CSCs.

Interestingly, endothelial cells, fibroblasts and stromal cells all showed mesenchymal scores, while iPSCs and ESCs yet again showed high variability covering both mesenchymal and epithelial scores. This stands in line with previous observations that pluripotent stem cells in cell culture persist in a dynamic balance between epithelial and mesenchymal identities [94]. However, most cancer cells are enriched in epithelial state, whereas a fraction of cells (likely CSCs) still possess high mesenchymal scores. This can be explained by the fact that the bulk of cells comprising the tumor maintain epithelial identity with only a few cells undergoing EMT. It is also known that cancer cells undergoing EMT are involved in metastasis of a tumor, in full agreement with our results.

The major limitation of the present study arises from the use of bulk RNA-seq, which is unable to capture cellular heterogeneity. Single-cell RNA-seq experiments provide an orthogonal view to the bulk RNA-seq by measuring the transcriptional profiles of individual cells, however at the expense of sparse coverage. Here, we used colorectal cancer to demonstrate as a proof of principle that the expression of stemness signature is highly heterogeneous not only intertumorally, but also at the level of individual cancer cells, with a larger contribution from intratumoral heterogeneity. Additionally, it was reported that specific patterns of hyper/hypo methylation follow the stem phenotype in cancer cells [95, 96]. Therefore, another source of stemness heterogeneity may come from the epigenetic component or from other factors such as variability in pre-mRNA splicing [97] and expression of non-coding genes, e.g. transposons [98]. A combined analysis of single-cell transcriptomes, pre-mRNA splicing by bulk RNA-seq, epigenetic assays based on ChIP-seq, and non-coding RNA quantification is needed for the detailed characterization of the stemness heterogeneity landscape. Growing amounts of this information in the public domain enable such characterization from the multi-omics perspective and open new directions for future studies.

Conclusion

Tissue-independent reactivation of the stemness program represents a unifying feature which ties together tumors of different origins. Nonetheless, the degree to which this program becomes reactivated strongly fluctuates between different tumor types and also inter- and intratumorally, thus suggesting that the effect of the stemness program on the tumor phenotype is highly heterogeneous. Multiple studies, including those based on single cell technology, have described tumor heterogeneity arising at different levels through various mechanisms. Here, we pinpoint yet another and previously unappreciated aspect of heterogeneity, one that is related to the reactivation of the stemness program, thus adding to an already complex picture of tumorigenesis and potentially impacting the diagnostic protocols and the development of new anticancer treatments.

Supporting information

S1 Fig. The distribution of the consensus purity estimate (CPE) values across the 19 tumors. The tumor types on the x axis are listed by in the descending order by the mean CPE value.
(TIF)

S2 Fig. Correction of gene expression values. (A) The distribution of $\log_2(1 + CPM)$, where CPM denotes median counts per million, for the raw and CPE-corrected gene expression values. The gene expression counts were first normalized by edgeR and then corrected for tumor purity (See Methods—Correction for tumor purity). (B) A scatter plot of CPE-corrected vs. raw $\log_2(1 + CPM)$. Each point represents a gene.
(TIF)

S3 Fig. Higher order principal components corresponding to Fig 2. Panels (A), (B), and (C) correspond to the stemness, proliferation, and EMT-MET signatures, respectively.
(TIF)

S4 Fig. Clustering of tumor, normal, and ESC samples according to previously reported stemness signatures: (A)—Ben-Porath *et al* [50], (B)—Bhattacharya *et al* [55], (C)—Wong *et al* [56].
(TIF)

S5 Fig. Expression of transcriptional repressors (*CBX7, SALL1*) and positive regulators of stemness (*POU5F1, SALL4*) in tumor samples along the PC1 axis in Fig 2A. Samples were stratified into quartiles according to PC1.
(TIF)

S6 Fig. The positions of LUSC and LUAD tumors in the clustering diagram in Fig 2A. All tumor samples except for LUSC and LUAD are colored gray. The expression of CBX7 drops (B), and the expression of SOX2 increases (C) in tumors on the way from normal samples to iPSCs and ESCs.
(TIF)

S7 Fig. Differentially expressed genes (DEGs) in the LUSC compared to the LUAD and the respective enrichment of their associated GO-terms.
(TIF)

S8 Fig. PCA clustering of tumor, normal, and ESC samples based on the controls sets of random genes that were matched by expression levels to stemness (A), proliferation (B), and EMT (C) signatures.
(TIF)

S9 Fig. The gradient of expression of mesenchymal (A-D) and epithelial (E-F) genes on the PCA clustering diagram corresponding to EMT signatures (see Fig 2C). Normal tissues are shown as gray background. Stem cells are colored as in Fig 2.
(TIF)

S1 Table. The accession numbers and description of the public RNA-seq datasets that were used in the principal component analysis along with TCGA data; *n* denotes the (effective) dataset size.
(TIF)

S2 Table. GEO accession numbers of the datasets that were used to assemble a set of stemness signature genes. Additional information is reported in Methods (see Signature gene sets section for details). $\log_2 FC$ denotes log-fold-change threshold (Treatment vs Control); $\log_2 E$ denotes log gene expression level threshold (Treatment + Control); Comparisons denotes the list of comparisons for differential gene expression analysis.
(TIF)

S1 File. TCGA sample subtype annotation and Stem cluster attribution.
(TSV)

S2 File. Table of literature sources of manually curated stemness markers.
(TSV)

S3 File. The list of stemness, proliferation and EMT/MET signature genes.
(TSV)

Acknowledgments

We thank Ivan Kulakovskii and Dmitrii Scherbinin for the help with curating and processing the data of Dmitrii Papatsenko. All living authors commemorate the contribution of Dmitrii Papatsenko, a cancer biologist who tragically died of a rapidly progressing lung cancer.

Author Contributions

Conceptualization: Artem Baranovsky, Dmitri Papatsenko, Dmitri D. Pervouchine.

Data curation: Artem Baranovsky, Dmitri Papatsenko, Dmitri D. Pervouchine.

Formal analysis: Artem Baranovsky, Timofei Ivanov, Dmitri D. Pervouchine.

Funding acquisition: Dmitri Papatsenko, Dmitri D. Pervouchine.

Investigation: Artem Baranovsky, Dmitri Papatsenko, Dmitri D. Pervouchine.

Methodology: Dmitri Papatsenko, Dmitri D. Pervouchine.

Project administration: Dmitri Papatsenko, Dmitri D. Pervouchine.

Supervision: Dmitri Papatsenko, Dmitri D. Pervouchine.

Validation: Artem Baranovsky, Dmitri D. Pervouchine.

Visualization: Artem Baranovsky, Timofei Ivanov.

Writing – original draft: Artem Baranovsky, Marina Granovskaya, Dmitri Papatsenko, Dmitri D. Pervouchine.

Writing – review & editing: Artem Baranovsky, Dmitri D. Pervouchine.

References

1. Roth GA, Abate D, Abate KHea. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018; 392(10159):1736–1788. https://doi.org/10.1016/S0140-6736(18)32203-7
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin. 2019; 69(5):363–385. https://doi.org/10.3322/caac.21565
3. Arruebo M, Vilaboa N, Sáez-Gutierrez B, Lambea J, Tres A, Valladares M, et al. Assessment of the evolution of cancer treatment therapies. Cancers (Basel). 2011; 3(3):3279–3330. https://doi.org/10.3390/cancers3033279
4. Baxevanis CN, Perez SA, Papamichail M. Combinatorial treatments including vaccines, chemotherapy and monoclonal antibodies for cancer therapy. Cancer Immunol Immunother. 2009; 58(3):317–324. https://doi.org/10.1007/s00262-008-0576-4 PMID: 18704409
5. Zhou BB, Zhang H, Damelin M, Geles KG, Grindley JC, Dirks PB. Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. Nat Rev Drug Discov. 2009; 8(10):806–823. https://doi.org/10.1038/nrd2137 PMID: 19794444
6. Clara JA, Monge C, Yang Y, Takebe N. Targeting signalling pathways and the immune microenvironment of cancer stem cells—a clinical update. Nat Rev Clin Oncol. 2020; 17(4):204–232. https://doi.org/10.1038/s41571-019-0293-2 PMID: 31792354
7. Visvader JE. Cells of origin in cancer. Nature. 2011; 469(7330):314–322. https://doi.org/10.1038/nature09781 PMID: 21248838
8. Friedmann-Morvinski D, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. EMBO Rep. 2014; 15(3):244–253. https://doi.org/10.1002/embr.201338254 PMID: 24531722
9. Polyak K, Hahn WC. Roots and stems: stem cells in cancer. Nat Med. 2006; 12(3):296–300. https://doi.org/10.1038/nm1379 PMID: 16520777
10. Pardal R, Clarke MF, Morrison SJ. Applying the principles of stem-cell biology to cancer. Nat Rev Cancer. 2003; 3(12):895–902. https://doi.org/10.1038/nrc1232 PMID: 14737120

11. Berenson RJ, Andrews RG, Bensinger WI, Kalamasz D, Knitter G, Buckner CD, et al. Antigen CD34+ marrow cells engraft lethally irradiated baboons. J Clin Invest. 1988; 81(3):951–955. https://doi.org/10.1172/JCI113409 PMID: 2893812

12. Kemper K, Prasetyanti PR, De Lau W, Rodermond H, Clevers H, Medema JP. Monoclonal antibodies against Lgr5 identify human colorectal cancer stem cells. Stem Cells. 2012; 30(11):2378–2386. https://doi.org/10.1002/stem.1233 PMID: 22969042

13. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007; 449(7165):1003–1007. https://doi.org/10.1038/nature06196 PMID: 17934449

14. Holyoake T, Jiang X, Eaves C, Eaves A. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. Blood. 1999; 94(6):2056–2064. https://doi.org/10.1182/blood.V94.6.2056 PMID: 10477735

15. Borovski T, De Sousa E Melo F, Vermeulen L, Medema JP. Cancer stem cell niche: the place to be. Cancer Res. 2011; 71(3):634–639. https://doi.org/10.1158/0008-5472.CAN-10-3220 PMID: 21266356

16. Clevers H. The cancer stem cell: premises, promises and challenges. Nat Med. 2011; 17(3):313–319. https://doi.org/10.1038/nm.2304 PMID: 21386835

17. Pattabiraman DR, Weinberg RA. Tackling the cancer stem cells—what challenges do they pose? Nat Rev Drug Discov. 2014; 13(7):497–512. https://doi.org/10.1038/nrd4253 PMID: 24981363

18. Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat Rev Cancer. 2008; 8(10):755–768. https://doi.org/10.1038/nrc2499 PMID: 18784658

19. Yang F, Cao L, Sun Z, Jin J, Fang H, Zhang W, et al. Evaluation of Breast Cancer Stem Cells and Intratumor Stemness Heterogeneity in Triple-negative Breast Cancer as Prognostic Factors. Int J Biol Sci. 2016; 12(12):1568–1577. https://doi.org/10.7150/ijbs.16874 PMID: 27994520

20. Malta TM, Sokolov A, Gentles AJ, Mariamidze Aea. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell. 2018; 173(2):338–354. https://doi.org/10.1016/j.cell.2018.03.034 PMID: 29625051

21. Ito T, Sato N, Yamaguchi Y, Tazawa C, Moriya T, Hirakawa H, et al. Differences in stemness properties associated with the heterogeneity of luminal-type breast cancer. Clin Breast Cancer. 2015; 15(2):93–103. https://doi.org/10.1016/j.clbc.2014.11.002 PMID: 25481840

22. Yun Z, Lin Q. Hypoxia and regulation of cancer cell stemness. Adv Exp Med Biol. 2014; 772:41–53. https://doi.org/10.1007/978-1-4614-5915-6_2 PMID: 24272353

23. Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, et al. Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. Nat Commun. 2019; 10(1):1787. https://doi.org/10.1038/s41467-019-09853-z PMID: 30992437

24. Birbrair A. Stem Cells Heterogeneity. Adv Exp Med Biol. 2019; 1123:1–3. https://doi.org/10.1007/978-3-030-11096-3_1 PMID: 31016591

25. Hayashi Y, Ohnuma K, Furue MK. Pluripotent Stem Cell Heterogeneity. Adv Exp Med Biol. 2019; 1123:71–94. https://doi.org/10.1007/978-3-030-11096-3_6 PMID: 31016596

26. Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyeuski A, et al. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc Natl Acad Sci U S A. 2019; 116(18):9020–9029. https://doi.org/10.1073/pnas.1818210116 PMID: 30996127

27. Chen H, He X. The Convergent Cancer Evolution toward a Single Cellular Destination. Mol Biol Evol. 2016; 33(1):4–12. https://doi.org/10.1093/molbev/msv212 PMID: 26464125

28. Guo G, von Meyenn F, Rostovskaya M, Clarke J, Dietmann S, Baker D, et al. Epigenetic resetting of human pluripotency. Development. 2017; 144(15):2748–2763. https://doi.org/10.1242/dev.146811 PMID: 28765214

29. Höving AL, Sielemann K, Greiner JFW, Kaltschmidt B, Knabbe C, Kaltschmidt C. Transcriptome Analysis Reveals High Similarities between Adult Human Cardiac Stem Cells and Neural Crest-Derived Stem Cells. Biology (Basel). 2020; 9(12). https://doi.org/10.3390/biology9120435 PMID: 33271866

30. Fraineau S, Palii CG, McNeill B, Ritso M, Shelley WC, Prasain N, et al. Epigenetic Activation of Proangiogenic Signaling Pathways in Human Endothelial Progenitors Increases Vasculogenesis. Stem Cell Reports. 2017; 9(5):1573–1587. https://doi.org/10.1016/j.stemcr.2017.09.009 PMID: 29033304

31. Blake LE, Thomas SM, Blischak JD, Hsiao CJ, Chavarria C, Myrthil M, et al. A comparative study of endoderm differentiation in humans and chimpanzees. Genome Biol. 2018; 19(1):162. https://doi.org/10.1186/s13059-018-1490-5 PMID: 30322406

32. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016; 44(8):e71. https://doi.org/10.1093/nar/gkv1507 PMID: 26704973

33. Burrows CK, Banovich NE, Pavlovic BJ, Patterson K, Gallego Romero I, Pritchard JK, et al. Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. PLoS Genet. 2016; 12(1):e1005793. https://doi.org/10.1371/journal.pgen.1005793 PMID: 26812582

34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

35. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013; 41(10):e108. https://doi.org/10.1093/nar/gkt214 PMID: 23558742

36. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019; 47(8):e47. https://doi.org/10.1093/nar/gkz114 PMID: 30783653

37. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1):139–140. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

38. Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet. 2020; 52(6):594–603. https://doi.org/10.1038/s41588-020-0636-z PMID: 32451460

39. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update—from bulk to single-cell expression data. Nucleic Acids Res. 2019; 47(D1):D711–D715. https://doi.org/10.1093/nar/gky964 PMID: 30357387

40. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallée VP, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat Med. 2020; 26(2):259–269. https://doi.org/10.1038/s41591-019-0750-6 PMID: 32042191

41. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015; 6:8971. https://doi.org/10.1038/ncomms9971 PMID: 26634437

42. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16(5):284–287. https://doi.org/10.1089/omi.2011.0118 PMID: 22455463

43. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004; 101(16):6062–6067. https://doi.org/10.1073/pnas.0400782101 PMID: 15075390

44. Lattin JE, Schroder K, Su AI, Walker JR, Zhang J, Wiltshire T, et al. Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res. 2008; 4:5. https://doi.org/10.1186/1745-7580-4-5 PMID: 18442421

45. Hailesellasse Sene K, Porter CJ, Palidwor G, Perez-Iratxeta C, Muro EM, Campbell PA, et al. Gene function in early mouse embryonic stem cell differentiation. BMC Genomics. 2007; 8:85. https://doi.org/10.1186/1471-2164-8-85 PMID: 17394647

46. Nishiyama A, Sharov AA, Piao Y, Amano M, Amano T, Hoang HG, et al. Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. Sci Rep. 2013; 3:1390. https://doi.org/10.1038/srep01390 PMID: 23462645

47. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007; 23(14):1846–1847. https://doi.org/10.1093/bioinformatics/btm254 PMID: 17496320

48. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004; 20(3):307–315. https://doi.org/10.1093/bioinformatics/btg405 PMID: 14960456

49. Ryaboshapkina M, Hammar M. Tissue-specific genes as an underutilized resource in drug discovery. Sci Rep. 2019; 9(1):7233. https://doi.org/10.1038/s41598-019-43829-9 PMID: 31076736

50. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet. 2008; 40(5):499–507. https://doi.org/10.1038/ng.127 PMID: 18443585

51. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43(Database issue):D447–452. https://doi.org/10.1093/nar/gku1003 PMID: 25352553

52. Cai J, Weiss ML, Rao MS. In search of "stemness". Exp Hematol. 2004; 32(7):585–598. https://doi.org/10.1016/j.exphem.2004.03.013 PMID: 15246154

53. Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, Chevassut T, et al. Comment on "'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". Science. 2003; 302(5644):393; author reply 393. https://doi.org/10.1126/science.1088249 PMID: 14563990

54. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007; 131(5):861–872. https://doi.org/10.1016/j.cell.2007.11.019 PMID: 18035408

55. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, et al. Gene expression in human embryonic stem cell lines: unique molecular signature. Blood. 2004; 103(8):2956–2964. https://doi.org/10.1182/blood-2003-09-3314 PMID: 15070671

56. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. Cell Stem Cell. 2008; 2(4):333–344. https://doi.org/10.1016/j.stem.2008.02.009 PMID: 18397753

57. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612. https://doi.org/10.1038/ncomms3612 PMID: 24113773

58. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012; 30(5):413–421. https://doi.org/10.1038/nbt.2203 PMID: 22544022

59. Andor N, Harness JV, Müller S, Mewes HW, Petritsch C. EXPANDS: expanding ploidy and allele frequency on nested subpopulations. Bioinformatics. 2014; 30(1):50–60. https://doi.org/10.1093/bioinformatics/btt622 PMID: 24177718

60. Zheng X, Zhao Q, Wu HJ, Li W, Wang H, Meyer CA, et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. Genome Biol. 2014; 15(8):419. https://doi.org/10.1186/s13059-014-0419-x PMID: 25103624

61. Ma C, Wang F, Han B, Zhong X, Si F, Ye J, et al. SALL1 functions as a tumor suppressor in breast cancer by regulating cancer cell senescence and metastasis through the NuRD complex. Mol Cancer. 2018; 17(1):78. https://doi.org/10.1186/s12943-018-0824-y PMID: 29625565

62. Gil J, Bernard D, Martínez D, Beach D. Polycomb CBX7 has a unifying role in cellular lifespan. Nat Cell Biol. 2004; 6(1):67–72. https://doi.org/10.1038/ncb1077 PMID: 14647293

63. Forzati F, Federico A, Pallante P, Abbate A, Esposito F, Malapelle U, et al. CBX7 is a tumor suppressor in mice and humans. J Clin Invest. 2012; 122(2):612–623. https://doi.org/10.1172/JCI58620 PMID: 22214847

64. Hammerman PS, Lawrence MS, Thomson Eea. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012; 489(7417):519–525. https://doi.org/10.1038/nature11404

65. Network CGAR, et al. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511 (7511):543–550. https://doi.org/10.1038/nature13385

66. Borcherding N, Bormann NL, Voigt AP, Zhang W. TRGAted: A web tool for survival analysis using protein data in the Cancer Genome Atlas. F1000Res. 2018; 7:1235. https://doi.org/10.12688/f1000research.15789.2 PMID: 30345029

67. Justilien V, Walsh MP, Ali SA, Thompson EA, Murray NR, Fields AP. The PRKCI and SOX2 oncogenes are coamplified and cooperate to activate Hedgehog signaling in lung squamous cell carcinoma. Cancer Cell. 2014; 25(2):139–151. https://doi.org/10.1016/j.ccr.2014.01.008 PMID: 24525231

68. Beerling E, Seinstra D, de Wit E, Kester L, van der Velden D, Maynard C, et al. Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. Cell Rep. 2016; 14(10):2281–2288. https://doi.org/10.1016/j.celrep.2016.02.034 PMID: 26947068

69. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014; 344(6190):1396–1401. https://doi.org/10.1126/science.1254257 PMID: 24925914

70. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018;. PMID: 30531897

71. Alda-Catalinas C, Bredikhin D, Hernando-Herraez I, Santos F, Kubinyecz O, Eckersley-Maslin MA, et al. A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program. Cell Syst. 2020; 11(1):25–41. https://doi.org/10.1016/j.cels.2020.06.004 PMID: 32634384

72. Nguyen QH, Lukowski SW, Chiu HS, Senabouth A, Bruxner TJC, Christ AN, et al. Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. Genome Res. 2018; 28(7):1053–1066. https://doi.org/10.1101/gr.223925.117 PMID: 29752298

73. Celià-Terrassa T, Kang Y. Distinctive properties of metastasis-initiating cells. Genes Dev. 2016; 30 (8):892–908. https://doi.org/10.1101/gad.277681.116 PMID: 27083997

74. Ye Y, Kuang X, Xie Z, Liang L, Zhang Z, Zhang Y, et al. Small-molecule MMP2/MMP9 inhibitor SB-3CT modulates tumor immune surveillance by regulating PD-L1. Genome Med. 2020; 12(1):83. https://doi.org/10.1186/s13073-020-00780-z PMID: 32988398

75. Bongiovanni D, Santamaria G, Klug M, Santovito D, Felicetta A, Hristov M, et al. Transcriptome Analysis of Reticulated Platelets Reveals a Prothrombotic Profile. Thromb Haemost. 2019; 119(11):1795–1806. https://doi.org/10.1055/s-0039-1695009 PMID: 31473989

76. Feng L, Houck JR, Lohavanichbutr P, Chen C. Transcriptome analysis reveals differentially expressed lncRNAs between oral squamous cell carcinoma and healthy oral mucosa. Oncotarget. 2017; 8 (19):31521–31531. https://doi.org/10.18632/oncotarget.16358 PMID: 28415559

77. Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, et al. Gene-specific patterns of expression variation across organs and species. Genome Biol. 2016; 17(1):151. https://doi.org/10.1186/s13059-016-1008-y PMID: 27391956

78. Rad A, Esmaeili Dizghandi S, Abbaszadegan MR, Taghechian N, Najafi M, Forghanifard MM. SOX1 is correlated to stemness state regulator SALL4 through progression and invasiveness of esophageal squamous cell carcinoma. Gene. 2016; 594(2):171–175. https://doi.org/10.1016/j.gene.2016.08.045 PMID: 27576349

79. Aponte PM, Caicedo A. Stemness in Cancer: Stem Cells, Cancer Stem Cells, and Their Microenvironment. Stem Cells Int. 2017; 2017:5619472. https://doi.org/10.1155/2017/5619472 PMID: 28473858

80. Wong DJ, Segal E, Chang HY. Stemness, cancer and cancer stem cells. Cell Cycle. 2008; 7(23):3622–3624. https://doi.org/10.4161/cc.7.23.7104 PMID: 19029796

81. Wang X. Computational analysis of expression of human embryonic stem cell-associated signatures in tumors. BMC Res Notes. 2011; 4:471. https://doi.org/10.1186/1756-0500-4-471 PMID: 22041030

82. Garraway LA, Sellers WR. Lineage dependency and lineage-survival oncogenes in human cancer. Nat Rev Cancer. 2006; 6(8):593–602. https://doi.org/10.1038/nrc1947 PMID: 16862190

83. Witwicki RM, Ekram MB, Qiu X, Janiszewska M, Shu S, Kwon M, et al. TRPS1 Is a Lineage-Specific Transcriptional Dependency in Breast Cancer. Cell Rep. 2018; 25(5):1255–1267. https://doi.org/10.1016/j.celrep.2018.10.023 PMID: 30380416

84. Johannessen CM, Johnson LA, Piccioni F, Townes A, Frederick DT, Donahue MK, et al. A melanocyte lineage program confers resistance to MAP kinase pathway inhibition. Nature. 2013; 504(7478):138–142. https://doi.org/10.1038/nature12688 PMID: 24185007

85. Shi K, Yin X, Cai MC, Yan Y, Jia C, Ma P, et al. PAX8 regulon in human ovarian cancer links lineage dependency with epigenetic vulnerability to HDAC inhibitors. Elife. 2019; 8. https://doi.org/10.7554/eLife.44306 PMID: 31050342

86. Park JW, Lee JK, Sheu KM, Wang L, Balanis NG, Nguyen K, et al. Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. Science. 2018; 362(6410):91–95. https://doi.org/10.1126/science.aat5749 PMID: 30287662

87. Cairo S, Wang Y, de Reyniès A, Duroure K, Dahan J, Redon MJ, et al. Stem cell-like micro-RNA signature driven by Myc in aggressive liver cancer. Proc Natl Acad Sci U S A. 2010; 107(47):20471–20476. https://doi.org/10.1073/pnas.1009009107 PMID: 21059911

88. Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med. 2011; 17(9):1086–1093. https://doi.org/10.1038/nm.2415 PMID: 21873988

89. Liu L, Chen K, Wu J, Shi L, Hu B, Cheng S, et al. Downregulation of miR-452 promotes stem-like traits and tumorigenicity of gliomas. Clin Cancer Res. 2013; 19(13):3429–3438. https://doi.org/10.1158/1078-0432.CCR-12-3794 PMID: 23695168

90. Forghanifard MM, Ardalan Khales S, Javdani-Mallak A, Rad A, Farshchian M, Abbaszadegan MR. Stemness state regulators SALL4 and SOX2 are involved in progression and invasiveness of esophageal squamous cell carcinoma. Med Oncol. 2014; 31(4):922. https://doi.org/10.1007/s12032-014-0922-7 PMID: 24659265

91. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144(5):646–674. https://doi.org/10.1016/j.cell.2011.02.013 PMID: 21376230

92. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, et al. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature. 2009; 462(7273):595–601. https://doi.org/10.1038/nature08592 PMID: 19898493

93. Martin Gonzalez J, Morgani SM, Bone RA, Bonderup K, Abelchian S, Brakebusch C, et al. Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. Stem Cell Reports. 2016; 7(2):177–191. https://doi.org/10.1016/j.stemcr.2016.07.009 PMID: 27509134

94. Li X, Pei D, Zheng H. Transitions between epithelial and mesenchymal states during cell fate conversions. Protein Cell. 2014; 5(8):580–591. https://doi.org/10.1007/s13238-014-0064-x PMID: 24805308

95. Pangeni RP, Yang L, Zhang K, Wang J, Li W, Guo C, et al. G9a regulates tumorigenicity and stemness through genome-wide DNA methylation reprogramming in non-small cell lung cancer. Clin Epigenetics. 2020; 12(1):88. https://doi.org/10.1186/s13148-020-00879-5 PMID: 32552834

96. Huang W, Hu H, Zhang Q, Wang N, Yang X, Guo AY. Genome-Wide DNA Methylation Enhances Stemness in the Mechanical Selection of Tumor-Repopulating Cells. Front Bioeng Biotechnol. 2020; 8:88. https://doi.org/10.3389/fbioe.2020.00088 PMID: 32258002

97. Xiao L, Zou G, Cheng R, Wang P, Ma K, Cao H, et al. Alternative splicing associated with cancer stemness in kidney renal clear cell carcinoma. BMC Cancer. 2021; 21(1):703. https://doi.org/10.1186/s12885-021-08470-8 PMID: 34130646

98. He J, Babarinde IA, Sun L, Xu S, Chen R, Shi J, et al. Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. Nat Commun. 2021; 12(1):1456. https://doi.org/10.1038/s41467-021-21808-x PMID: 33674594