

## RESEARCH ARTICLE

# High accuracy gene expression profiling of sorted cell subpopulations from breast cancer PDX model tissue

Warren Porter<sup>☉</sup>, Eileen Snowden<sup>☉</sup>, Friedrich Hahn, Mitchell Ferguson, Frances Tong, W. Shannon Dillmore, Rainer Blaesius<sup>ID</sup>\*

BD Technologies and Innovation, Research Triangle Park, NC, United States of America

☉ These authors contributed equally to this work.

\* [Rainer.Blaesius@bd.com](mailto:Rainer.Blaesius@bd.com)



## OPEN ACCESS

**Citation:** Porter W, Snowden E, Hahn F, Ferguson M, Tong F, Dillmore WS, et al. (2020) High accuracy gene expression profiling of sorted cell subpopulations from breast cancer PDX model tissue. *PLoS ONE* 15(9): e0238594. <https://doi.org/10.1371/journal.pone.0238594>

**Editor:** Randall J. Kimple, University of Wisconsin, UNITED STATES

**Received:** April 20, 2020

**Accepted:** August 19, 2020

**Published:** September 10, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0238594>

**Copyright:** © 2020 Porter et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are within the manuscript and supporting information files. The complete RNASeq raw data can be accessed at the EMBL-EBI ArrayExpress server under the

## Abstract

Intratumor Heterogeneity (ITH) is a functionally important property of tumor tissue and may be involved in drug resistance mechanisms. Although descriptions of ITH can be traced back to very early reports about cancer tissue, mechanistic investigations are still limited by the precision of analysis methods and access to relevant tissue sources. PDX models have provided a reproducible source of tissue with at least a partial representation of naturally occurring ITH. We investigated the properties of phenotypically distinct cell populations by Fluorescence activated cell sorting (FACS) tissue derived cells from multiple tumors from a triple negative breast cancer patient derived xenograft (PDX) model. We subsequently subjected each population to in depth gene expression analysis. Our findings suggest that process related gene expression changes (caused by tissue dissociation and FACS sorting) are restricted to Immediate Early Genes (IEGs). This allowed us to discover highly reproducible gene expression profiles of distinct cellular compartments identifiable by cell surface markers in this particular tumor model. Within the context of data from a previously published model our work suggests that gene expression profiles associated with hypoxia, stemness and drug resistance may reside in tumor subpopulations predictably growing in PDX models. This approach provides a novel opportunity for prospective mechanistic studies of ITH.

## Introduction

Phenotypic differences between patients with the same tumor type (intertumor heterogeneity) and among cancer cells within the same patient (intratumor heterogeneity/ITH) have always been integral to descriptions of tumor appearance [1], yet they were neglected as a source of useful information for a long time. While intertumor heterogeneity has been integrated into clinical practice, particularly in breast cancer [2], ITH remains to be included as a standard part of cancer tissue analysis. The technical difficulty to assess ITH has hindered efficient ways to build on observations of inherent complexity described in the 1970s and 80s and arguably a general resistance to this concept [3] prevented an unbiased appreciation. The past 10–15 years have seen a “re-discovery” of ITH as a mainstream topic. Models of “tumor as an ecosystem” [4] or “tumor as an organ” [5] explain ITH not as incidental phenomenon but as a

accession number “E-MTAB-9428” with the title “High accuracy gene expression profiling of sorted cell subpopulations from breast cancer PDX model tissue” (<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9428/>).

**Funding:** This study was completely funded by BD Technologies and Innovation. The funder provided support in the form of salaries for authors WP, ES, FT, WSD, FH, MF and RB, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the ‘author contributions’ section.

**Competing interests:** WP, ES, FT, WSD and RB are employees of BD Technologies and Innovation and FH and MF were employees of BD Technologies and Innovation at the time the work was carried out. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the ‘author contributions’ section. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

foundational property. Focusing on transcriptional heterogeneity in particular has become an effective way to link ITH to clinically meaningful phenomena [6]. Maximizing and understanding the accuracy of this approach will be an important factor in its efficient translation into the clinic.

scRNASeq has arguably been the main tool to accelerate discovery of transcriptional heterogeneity in tumors and healthy tissues alike. The results have prompted models which transcend the limits set by traditional cell type definitions and replace the artificially fixed categories with continuous gene expression spaces of individual cells [7, 8]. This is likely a much more accurate depiction of cell biological reality since each organism develops from a single cell in a continuous process rather than discrete “jumps”. The drawback, however, lies mainly in a technical limitation: scRNASeq shows a ‘dropout’ effect due to inefficient mRNA capture [9] which makes it very difficult to create a high resolution model for a given sample. For elucidation of organisms (especially with high genetic homogeneity such as in inbred strains) this can be counteracted by sequencing large numbers of cells; maybe the most impressive example is represented in the “Tabula Muris”, a high resolution analysis of 20 murine organs generated from over 100,000 cells [10]. This publication highlighted the enormous resources required but also showed that scRNASeq can be powerfully enhanced by combining it with cell sorting. Our investigations cause us to argue that the two methods are complementary rather than competing and demonstrate that this approach enables uncovering of highly reproducible gene expression patterns with striking functional implications.

Patient derived xenograft models allow propagation of human tumor tissue in immune compromised mice in ways that seem much superior to tissue culture conditions, particularly with regard to maintaining ITH [11]. Previously we had established methods in our lab that clearly demonstrate the presence of distinct subpopulations within breast cancer tissue emerging reproducibly in separate host animals implanted with tissue from the identical primary tumor [12]. Here we addressed the possible influence of fluorescence activated cell sorting (FACS) on transcriptional profiling and investigated the presence of multiple different subpopulations within the same breast cancer model tissue. As in the previous work we chose surface markers identifying cell compartments with sufficient numbers for molecular analysis of the subpopulations, for this model CD49f and CD133. Those two markers as well as CD184 which we had employed for the previous tumor model have been shown in numerous studies to allow identification of cancer cells with distinct roles in tumorigenesis and have frequently been used in the search of cancer stem cells [12–14]. Using quadruplicate samples we were able to demonstrate the potential this approach provides for detailed studies of specific cellular compartments within a complex tumor tissue. In addition, our data indicates that RNASeq of sorted PDX tumor cell populations complements scRNASeq through a) a much higher accuracy by avoiding the drop-out effects of single cell sequencing and b) providing a concept for replenishable sources of cell subpopulations with predictable gene expression signatures.

## Methods

All experimental procedures were essentially as previously described [12], any differences are noted below.

### PDX tumor extraction

The protocol was approved by the BD Institutional IACUC committee. BD is licensed by the USDA and is accredited by AALAC. All methods were carried out in accordance to best practices outlined in the *Guide for the care and use of Laboratory Animals* [15].

Mice were sourced from The Jackson Laboratory. The properties of the mouse model used in the study was a triple negative breast cancer model (BRC13, Jackson Lab designation TM00999) previously characterized for surface marker expression in our lab [12]. Female NOD/SCID/IL2Rg<sup>-/-</sup> (NSG) mice age 6 to 7 weeks were purchased from The Jackson Laboratory with tumor material already implanted. All animal studies were performed under an Institutional Animal Care and Use Committee–approved protocol. A dedicated room to safely house immune compromised mice was used. Positive air flow and filtration was maintained through a BioBubble system. Mice were socially housed in groups of five within a micro isolation cage rack. Food and water were available ad lib. Each cage contained environmental enrichments, including nestlets and igloos. Tumors were assessed weekly for growth and measured with a digital caliper. Tumor volume was estimated by the formula  $0.5 \times (\text{length} \times \text{width}^2)$ . Animals were monitored daily with additional measurements taken if needed. Animals were euthanized if grafted tumor tissue showed signs of necrosis. Additional health signs were monitored with the following termination points: “rapid” weight loss of >10%, specific clinical signs of severe organ involvement (e.g. circling, head-tilt, respiratory rates, paralysis). Also, Body Condition Score (BCS) was monitored, with a score of 2 or below serving as a humane endpoint. Once tumor size reached >500 mm<sup>3</sup> (several weeks to several months after inoculation), they were harvested for analysis. Mice were humanely euthanized via exsanguination under anesthesia at the time of tumor extraction, and the tumor was immediately removed via blunt dissection and placed in cold PBS for dissociation and analysis.

### Preparation of single-cell suspensions from tumor tissue

Single cell suspensions were prepared as previously described. Briefly, excised tissue was weighed and then minced using a scalpel or scissors to a size not exceeding 1 to 2 mm<sup>3</sup>. Tissue was then enzymatically dissociated into single cells using Horizon™ Dri Tumor & Tissue Dissociation Reagent (BD Biosciences, San Jose, CA) in a 37°C water bath for 30 minutes with frequent agitation. The enzymatic reaction was stopped by rinsing in PBS (Cellgro) and 1% BSA (Sigma-Aldrich). Dissociated tissue was filtered through a 70 μm sieve and treated with ACK Buffer (Life Technologies) to remove contaminating red blood cells. Cells were resuspended in PBS, without magnesium and calcium.

### Cell staining for flow cytometry and analysis

Before staining, single-cell suspensions were blocked for 30 minutes, on ice in FcR block (Miltenyi Biotec). For surface marker characterization, cells were diluted in PBS and distributed into wells of a 96-well plate containing viability dye (LIVE/DEAD Fixable Near-IR or Aqua, Invitrogen) and Hoechst 33342 (Invitrogen; at 0.2 mg/mL) for viable nucleated cell identification. The basic sorting strategy was as described previously [12]. Monoclonal antibodies for CD49f (clone G0H3) were purchased from BD Biosciences, for CD133 from Miltenyi (Bergisch Gladbach, Germany). A complete list of reagents is provided in S1 Table. Cells were incubated with antibody solutions for 30 minutes at 4°C, in the dark, and then rinsed twice in PBS before acquisition on BD LSRII flow cytometer. For sorting, cells were stained using a four-color panel combining amine viability reagent, anti-mouse cocktail, and markers identified from the primary characterization plate. After staining, cells were rinsed, and resuspended in BD FACS Pre-Sort Buffer, filtered again through a 70 μm sieve, and populations of interest were sorted on the BD FACSAria II (BD Biosciences) at 20 p.s.i. using a 100-μm nozzle. Cell doublets or clumps were removed with electronic doublet discrimination gating. Post-sorting analysis of sorted subpopulations regularly showed purity >95%. Analysis of results was performed using FACSDiva software (version 6.1.3 LSRll and version 8.0 FACSAria II). Graphical representations of the data were performed using R or MS Excel.

## RNA isolation, RNASeq and sequence data analysis

20–50,000 cells stained for a specific surface marker were collected during FACS sorting and processed in bulk for RNA extraction. Quality assessments were determined using Agilent BioAnalyzer and NanoDrop readings. RNA specimens underwent QC analysis to determine their suitability for target labeling. QC analyses included determination of concentration and volume, A260/A280 ratio, and 28S/18S rRNA ratio (where applicable), and an RNA integrity summary score (Agilent RIN or Caliper RQS). RNASeq was performed using the Illumina RNASeq TruSeq Stranded mRNA SQ905 Kit. All samples were sequenced in one run across four lanes on an Illumina HiSeq with 2 x 50 paired end reads. The primary sequencing analysis was done through a 3rd party (EA Genomics, Morrisville, NC) using STAR v2.4 for sequence alignment and RSEM v1.2.14 for transcription quantification.

RSEM counts from the sequencing results were used for the secondary analysis performed in R. Non-human genes were filtered out of the data set. Differential expression was calculated using the edgeR v3.14.0 package:

- 1.) Groups were created for each of the sorted populations ("Tissue", "Unsorted\_0hr", "Unsorted\_3hr", "Human\_Only", "CD49f\_Low", "CD49f\_High", "CD133\_Low", "CD133\_High").
- 2.) Normalization factors were calculated with calcNormFactors using the default settings, to convert observed library sizes to effective library sizes.
- 3.) Counts per million (CPM) and log CPM were calculated from the normalized data.
- 4.) glmQLFit was used to model the fit of the normalized data.
- 5.) glmQLFTest was used to model pairwise comparisons between the groups.
- 6.) topTags was used to extract the most differentially expressed genes from each comparison object created from step 5 (glmQLFTest).
- 7.) The resulting differential expression data was used for comparison to 3rd party gene lists.

Statistical significance was calculated and displayed in all figures using false discovery rate (FDR).

To assess the biological interpretation of gene expression patterns all DEGs with  $FDR < 0.1$  and  $x\text{-fold} \geq 2$  were analyzed using the Enrichr tool [16, 17].

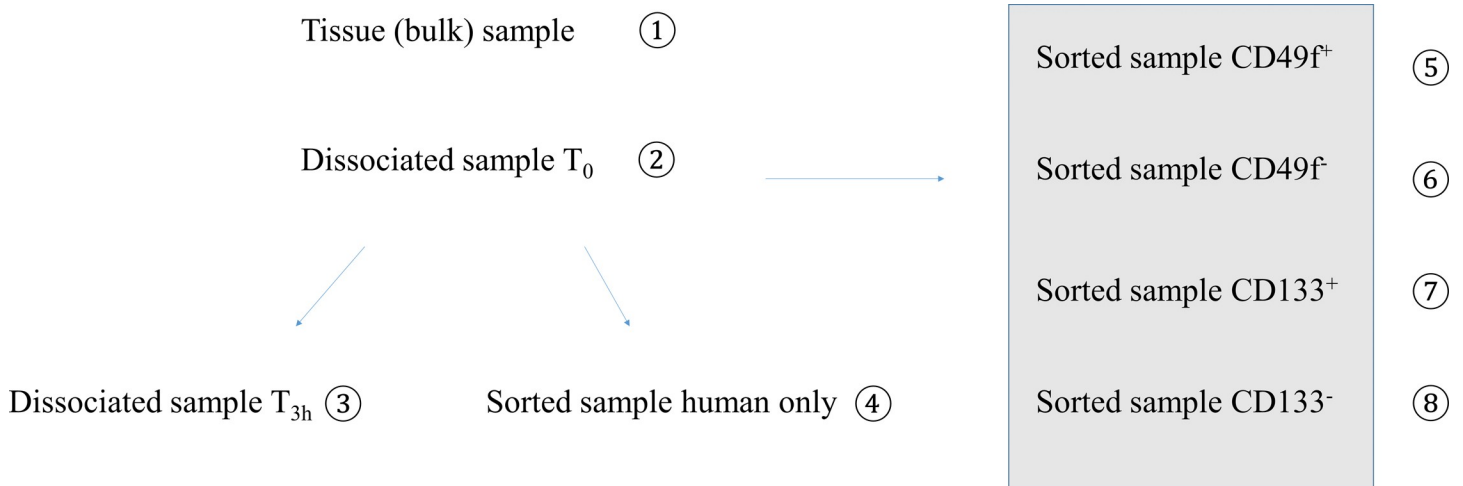
## Results

### Influence of tissue dissociation and FACS sorting on gene expression profiles

Our study was based on pairwise comparisons of gene expression profiles for (8) different cell populations from a triple negative breast cancer PDX model. The first (4) samples were used to validate our sample preparation and gene expression analysis processes, while subsequent samples were FACS sorted using (2) surface markers to compare biological similarities and differences in gene expression (Figs 1 and S1). Variables investigated in the sample preparation included tissue dissociation, duration of the method and flow sorting as possible sources of bias. RNA isolated from freshly collected tissue using a standard method served as a reference.

To control for the technical variability as well as biological variability between individual host mice, we carried out the experiment in quadruplicate, (i.e. each of the 8 conditions shown in Fig 1 represents 4 separate tumors from the same model grown in individual host animals harvested on different days) resulting in a total of 32 samples. Three samples did not pass QC due to limited amounts of RNA and/or poor RNA integrity.

Out of the 3 major variables (tissue dissociation, protocol duration and flow sorting) tissue dissociation was the only sample processing variable introducing significant changes in expression patterns (Fig 2). Among 24 differentially expressed genes (DEGs) after dissociation, 23 were upregulated above a cutoff of  $FDR < 0.05/4\text{-fold}$  and only a single one was

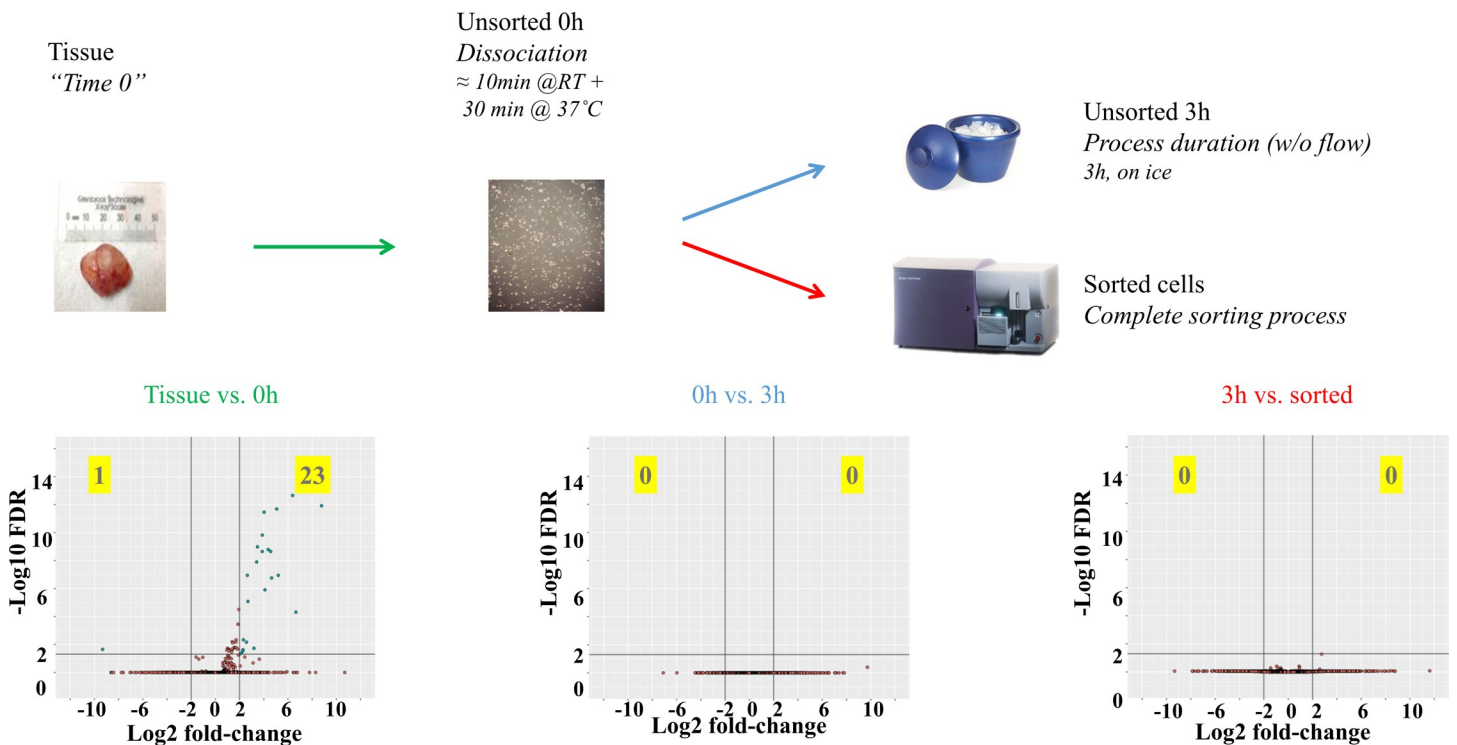


**Fig 1. Schematic of experimental setup.** Tumor tissue was extracted and gene expression profiles from 8 different samples per tumor were generated from 4 mice (i.e. total of 8x4 = 32 RNASeq profiles) of the same breast cancer PDX model, BRC13. The left part addressed the sample handling aspects of the procedure while the samples marked by the grey box were generated to investigate ITH present in this model. See text for details.

<https://doi.org/10.1371/journal.pone.0238594.g001>

downregulated, suggesting *de novo* RNA synthesis rather than RNA degradation as the main effect triggered by this sample preparation step.

Investigation of the characteristics of the significantly changed genes implicated a specific mechanism causing these observations. A database query of the significantly changed genes



**Fig 2. Pairwise comparisons between gene expression profiles generated through bulk RNASeq for 4 different conditions.** 0h time point represents cells immediately after dissociation; 3h time point represents cells left undisturbed for 3 hours after dissociation while sorted cells underwent flow sorting which generally took up to 3 hours. Volcano plots show the log<sub>2</sub> fold-change (x-axis) and the statistical significance as -log(FDR) (y-axis). Lines indicate the cutoff values chosen (4-fold and FDR < 0.05). FDR values were calculated across the entire data set.

<https://doi.org/10.1371/journal.pone.0238594.g002>

(Table 1) revealed that a great majority (20/24) have been described as immediate early genes (IEG). Of the 4 genes not identified in the IEG literature, 2 (AL832891 and AK024092) have 99% homology to other IEGs (Cyr61 and CHD2) and the remaining 2 are non-coding RNAs. Similar trends were seen with gene sets analyzed with lower stringency cutoff values (S2 Table).

### Gene expression profiles of cell subpopulations identified by CD49f and CD133

The second half of the experiments addressed ITH and whether subpopulations identified by surface markers differed in their gene expression profiles. Previously, we identified surface marker ITH in nine triple negative breast cancer models [12]. Each model had a unique and reproducible marker profile. Indication of ITH (i.e. presence of both positive and negative cells for a given marker) existed with respect to many of the markers but also differed among the models. Investigation of CD184<sup>hi</sup> and CD184<sup>lo</sup> sorted cells in one model revealed major differences in gene expression profiles between marker-high and low populations [12]. In the current study we focused on another PDX triple negative breast cancer model exhibiting varying expression levels of the surface markers CD49f and CD133. Both markers have been described as indicators of functionally distinct roles in cancer tissue such as drug resistance, metastasis, EMT and stemness [28–30]. For our investigation their choice was also opportunistic since they allowed isolation of sufficient cell numbers for both marker positive and negative

**Table 1. List of genes identified as DE before and immediately after tissue dissociation (FDR<0.05, x-fold>4).**

	Gene	FDR	x-fold change	Reference identifying gene as IEG
1	FOS	2.3E-15	263.0	[18]
2	EGR1	2.2E-13	83.5	[18]
3	FOSB	1.2E-12	437.0	[18]
4	DUSP1	2.0E-12	33.5	[18]
5	IER2	3.5E-12	16.3	[19]
6	JUN	1.5E-10	14.7	[18]
7	JUNB	1.0E-09	11.1	[18]
8	ZFP36	1.6E-09	20.6	[18]
9	NR4A2	2.2E-09	23.7	[18]
10	RHOB	2.2E-09	14.7	[20]
11	KLF2	1.2E-08	10.6	[21]
12	GADD45B	1.1E-07	6.2	[18]
13	EGR2	1.1E-07	36.7	[22]
14	NR4A1	1.8E-07	25.0	[18]
15	ATF3	1.2E-06	17.1	[18]
16	SOCS3	8.3E-06	6.5	[23]
17	FAM71A	4.9E-05	100.9	[24]
18	AL832891	4.6E-03	5.0	*
19	LOC284454	6.8E-03	5.9	ncRNA
20	BTG2	1.9E-02	9.1	[25]
21	CTD-2311B13.7:ENSG00000257931	2.2E-02	-640.1	ncRNA
22	CXCL2	2.4E-02	4.9	[26]
23	AK024092	2.9E-02	4.8	*
24	DUSP2	4.3E-02	4.4	[27]

\* no reference found

<https://doi.org/10.1371/journal.pone.0238594.t001>



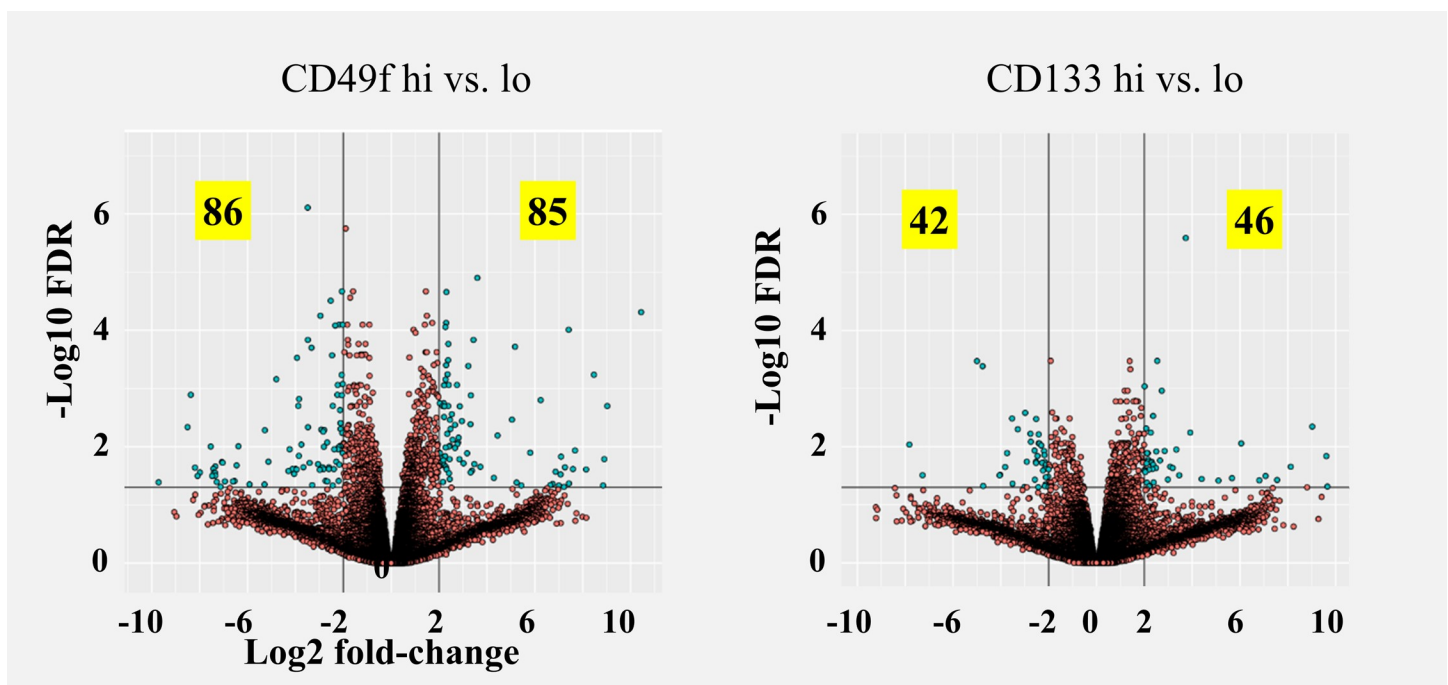
phenotypes. The previous model as well as the one investigated here contained 15% of mouse cells which were excluded from the subsequent gene expression analysis [12].

We FACS sorted the dissociated tumor cell population into CD49f<sup>hi</sup> and CD49f<sup>lo</sup> as well as CD133<sup>hi</sup> and CD133<sup>lo</sup>, and performed (whole transcriptome) RNASeq on each pool. Both sorting approaches identified a number of DEGs, even at very stringent statistical significance levels (Fig 3 and Table 2), suggesting significant differences between the respective subpopulations. The finding that gene expression correlated with protein expression (i.e. ITGA6 for CD49f and PROM1 for CD133 were significantly upregulated in the marker positive population) suggests gene variance is due to biological changes and not due technical noise. Moreover, the statistical significance for both genes was highest among upregulated genes in the marker positive population in both experiments (S2 Fig).

Among all genes identified in both experiments the majority were unique to one arm but among the ones in common (ca. 1/3 of the CD133 gene set overlapped with 1/8 of the CD49f set) an unexpected inverse correlation was detected (Fig 4). 120 of the 121 genes in common were inversely correlated at a qualitative level (i.e. identical direction of change) and at a quantitative level the correlation coefficient was  $R = -0.85$ , strongly suggesting that this was not caused by random noise.

A priori an inverse correlation could be a sign of a simple negative correlation between the two markers. However both the flow analysis (Fig 4A) as well as the finding that more DE genes are unique to one arm of the experiment than common in both arms argue for a more complex relationship between the populations.

To investigate the possible roles for the respective subpopulations we determined GO terms for each of the 6 upregulated gene sets (S5–S10 Tables). Possibly the best defined functional indicator could be determined for the BRC13/CD133<sup>hi</sup> population which—very much like the BRC12/CD184<sup>hi</sup> population contained many GO terms associated with cell proliferation (S6 and S10



**Fig 3. Volcano plots of DEGs identified between tumor subpopulations sorted with either CD49f (left) or CD133 (right).** Cutoff values shown (black lines, for illustration only) are >4-fold and FDR <0.05.

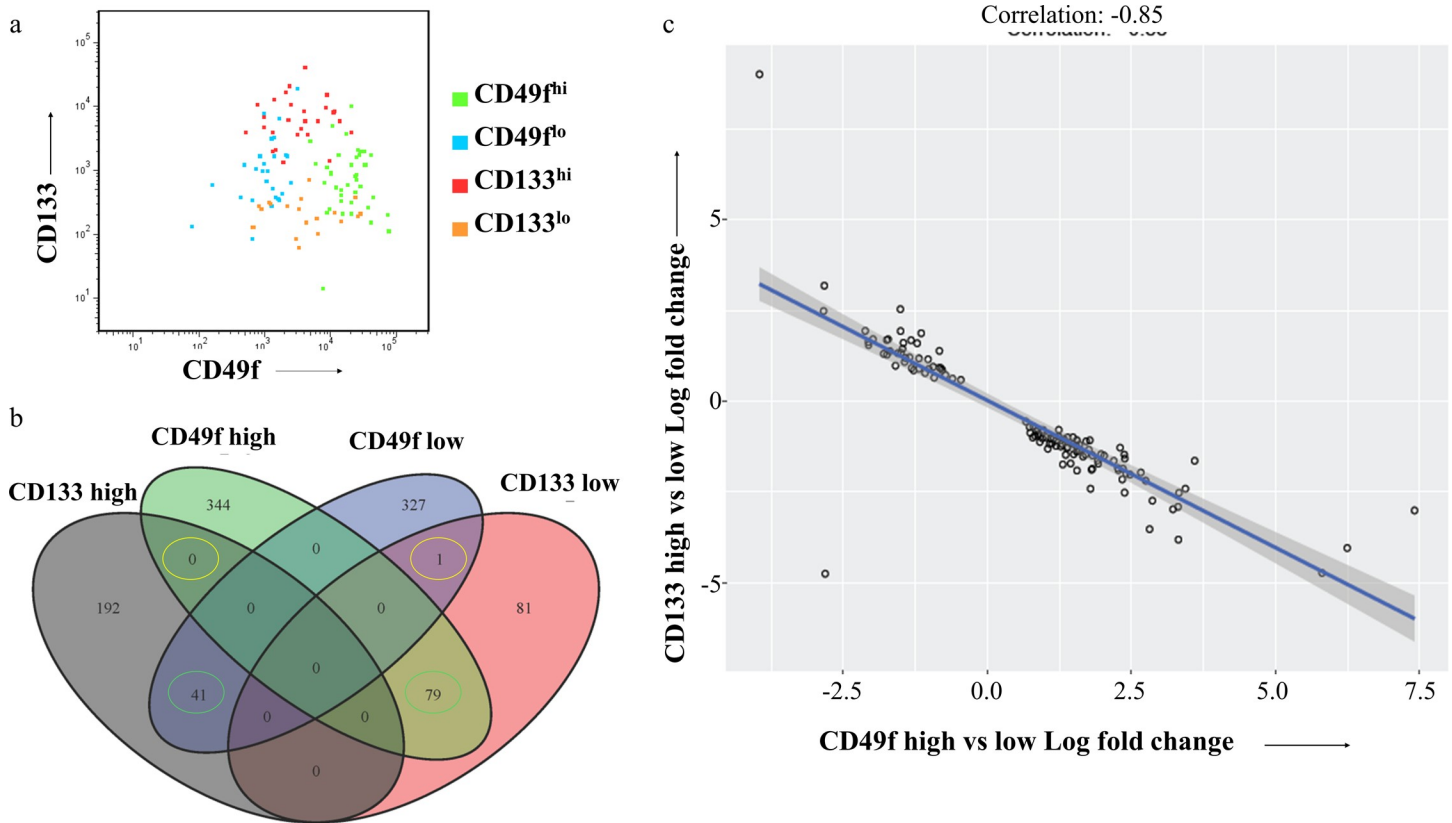
<https://doi.org/10.1371/journal.pone.0238594.g003>

**Table 2. Number of differentially expressed genes at different cutoff levels.**

Cutoff	CD49f <sup>hi</sup>	CD49f <sup>lo</sup>	CD133 <sup>hi</sup>	CD133 <sup>lo</sup>
0.1	668	620	432	274
0.05	423	369	233	161
0.01	150	129	57	33

<https://doi.org/10.1371/journal.pone.0238594.t002>

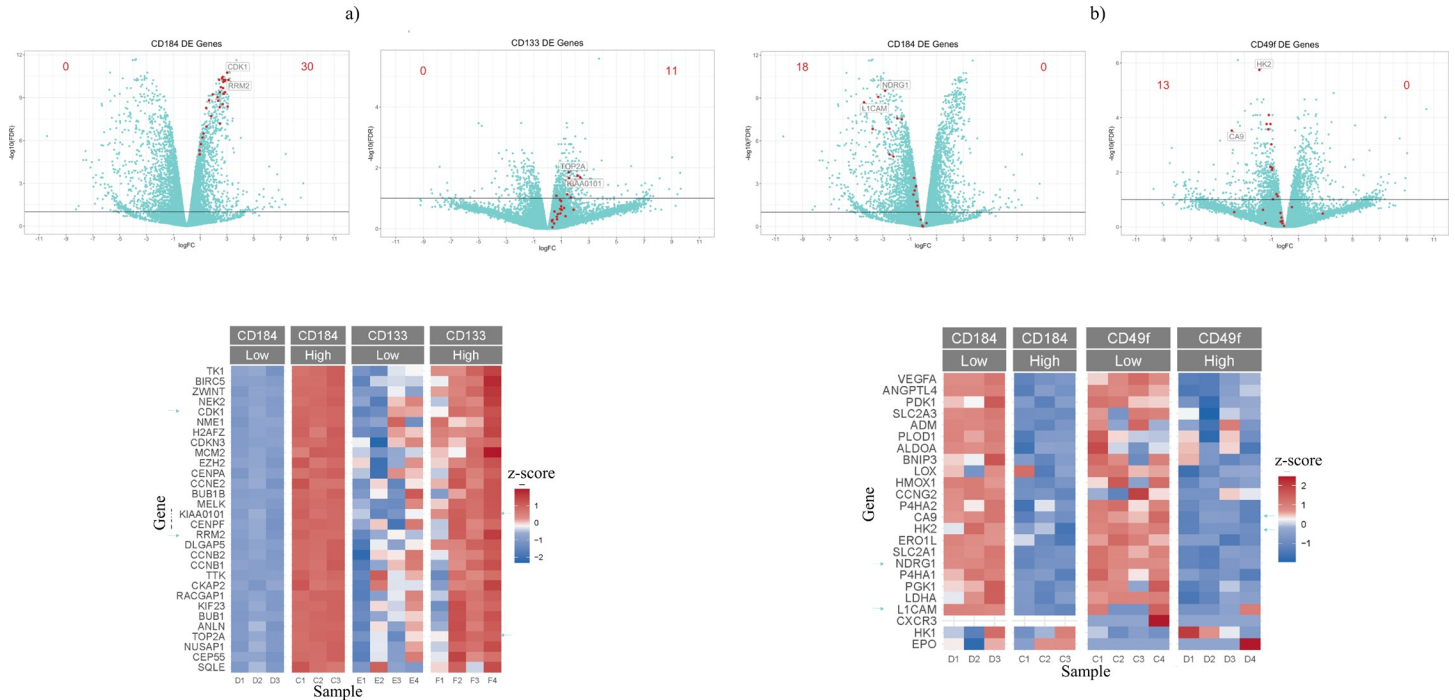
Tables). The BRC13/CD49<sup>lo</sup> population shows a variety of signs consistent with organization of the microenvironment, such as ECM organization, angiogenesis and regulation of cell migration which all appear as cellular processes indicated for the BRC12/CD184<sup>lo</sup> subpopulation (S7 and S9 Tables). The genes upregulated in BRC13/CD49<sup>hi</sup> and BRC13/CD133<sup>lo</sup> populations did not provide GO terms which would indicate a clear consistent function but—consistent with the inverse correlation described above—8 out of the top 50 GO terms in those two lists are identical (S5 and S8 Tables). While neither BRC13 DEG list provided evidence for a strong and unambiguous functional interpretation of the gene profiles, the CD133<sup>hi</sup> and CD49<sup>lo</sup> lists mirror a number of gene sets identified in breast cancer as well as other solid cancers, indicating the possibility of functional relevance of the subpopulations described. Particularly, the BRC12/CD184<sup>lo</sup> and BRC13/CD49<sup>lo</sup> populations contained many overexpressed genes in common with profiles described under conditions of hypoxia (Fig 5A, add'l examples in supplementary data). In contrast, the BRC12/CD184<sup>hi</sup> and BRC13/CD133<sup>hi</sup> populations contained equally pronounced similarities to profiles identified in connection with high proliferation in healthy breast tissue, regulation by the E2F transcription



**Fig 4.** Flow cytometric analysis of sorted populations using CD49f and CD133 (a) and correlations between DEGs identified by using either CD49f or CD133 as sorting criteria (b and c, FDR<0.05).

<https://doi.org/10.1371/journal.pone.0238594.g004>





**Fig 5. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations.** A hypoxia gene signature composed of 3 published signatures [30–32] was found in BRC12 CD184<sup>lo</sup> and BRC13 CD49f<sup>lo</sup> subpopulations (a) while a proliferative gene signature identified in histologically normal breast tissue from cancer patients [41] was present in BRC12 CD184<sup>hi</sup> and BRC13 CD133<sup>hi</sup> subpopulations (b). Genes with lowest FDR and highest x-fold values are labeled in volcano plots and marked with arrows in heatmaps, respectively.

<https://doi.org/10.1371/journal.pone.0238594.g005>

factor family, DNA damage response and stem cell properties (Fig 5B and supplementary data). Such parallels to previously documented results could also be identified within functional gene families; especially within the keratin family of genes 8 out of 9 members DE with  $FDR < 0.1$  in both BRC12/CD184<sup>lo</sup> and BRC13/CD49f<sup>lo</sup> cells tracked in almost identical patterns (S3 Table).

### Discussion

Tissue transcriptional profiling is dependent on the accurate reflection of cells in their physiological environment. In other words, distortion caused by sample handling or signal amplification are a frequent hurdle to identifying genes that are consistently expressed in a differential state that reflects physiology [31]. We established a workflow that accurately identifies distinct subpopulations in tissue from breast cancer PDX models [12]. In this study, we set out to determine if sample handling or flow cytometry introduced bias in transcriptional profiling and show different subpopulation profiles indicating complex ITH in tumor tissue from a breast cancer PDX model. The surface markers were chosen based on their expression levels within this model rather than for known biological roles.

To test accuracy we compared expression profiles assessed by whole genome RNASeq from dissociated tissue cells before and after FACS sorting with the “gold standard” profile of bulk fresh tissue. Using quadruplicate samples (i.e. 4 separate tumors from the same tumor model) allowed us to statistically assess random fluctuations in individual samples from consistent gene expression differences caused by the steps investigated. We found no evidence of compromised mRNA integrity in any of the steps including flow sorting. Our results indicate that tissue dissociation caused significant changes in a limited set of genes. However, outside of this specific group (IEGs representing less than 0.5% of the human genome) no significant

expression differences were detected when compared to fresh tissue. In contrast, transcriptional profiling of sorted subpopulations derived from this tissue identified distinct gene clusters differentially expressed in cells identified by specific surface markers. Furthermore, our data provides evidence that those differences can be detected with extremely high precision.

IEGs have been extensively described in the literature as fast responders to multiple stimuli. It's been shown that IEGs in a variety of cell types are activated within minutes after exposure to various perturbations, including mechanical stress [32]. Their ability to rapidly activate transcription likely represents part of a cascade of events that are involved in tight gene regulation. The cascade is initiated with a distinct set of immediate early genes, followed by a set of delayed early genes [18, 33], a nomenclature derived from the sequence of viral gene expression after successful entry of the virus into its host cell. Due to this cascade of early events, it's unlikely any published tissue dissociation protocol (or detachment of adherent tissue culture cells) will prevent activation of IEGs. However, our data suggests artefactual changes to the rest of the transcriptome can be avoided, presumably by limiting the ability of cells to execute the full cascade. The specific list of IEGs involved in a given experiment is context dependent [32]. Our survey of published literature revealed that many IEGs overlap between different experiments and their total number is unlikely to exceed 150–200 genes. As a consequence, we have shown that our protocol allows us to effectively limit process related bias and accurately detect gene expression changes in >99% of genes. Bioinformatic filtering of all known IEGs and inclusion of control samples offer additional measures to mitigate artefacts when processing cells derived from tissue. It is also remarkable that aside from the IEGs discussed above, fewer than 200 genes would be categorized as DE even up to a FDR <0.9 cutoff. This strongly argues that no systematic distortion is caused by the methods used here and experimental noise is minimal (see S3 Fig). In conclusion, our data suggests that our methods detect DEGs between different experimental conditions with high accuracy and reproducibility.

The exploratory segment of our experiment targeted the discovery of gene expression profiles of interest in tumor tissue subpopulations identified by various surface markers. A previously published example from our lab had validated this approach [12] and we aimed to uncover whether using different surface markers as well as a different tumor model may result in similar or different subpopulation profiles.

We successfully identified distinct DEG sets with each marker containing hundreds of DEGs even at stringent statistical cutoff values (Table 2). Several arguments derived from our results would indicate that RNASeq performed on FACS sorted cell populations enables sensitive and precise identification of genes unique to these cells: 1) The genes for both surface markers used in this study (ITGA6/CD49f and Prom1/CD133) as well as our previous study (CXCR4/CD184) were among the top 3 DEGs (ranked by FDR, S2 Fig). 2) Among the range of differentially expressed genes with very high statistical significance (FDR<0.01) a number were detected at lower than 2-fold differential expression (8 for CD133 and 39 for CD49f sorting). 3) The DEGs common to both CD133 sorting as well as CD49f sorting were strongly correlated; the 121 genes in common between each sort (at FDR<0.05) have a coefficient of  $R = -0.85$  (Fig 4). Such a clear correlation is much more consistent with an underlying biological control of those genes as opposed to a random occurrence and further supports the notion of a very low noise level. All of these observations are consistent with “clean” separation of expression profiles in subpopulations isolated through FACS sorting.

When we compared the two DEG lists from this work with our previously published list (using a different TNBC model labeled BRC12) and with gene profiles from other investigators we found a number of surprising parallels. For the two cancer models used in our lab there was substantial overlap between DEGs in the BRC12/CD184<sup>lo</sup> and BRC13/CD49f<sup>lo</sup> populations on one hand and BRC12/CD184<sup>hi</sup> and BRC13/CD133<sup>hi</sup> on the other (Figs 5 and S4–

S13). This was not only far more than would be expected by chance but each overlap also aligned with gene expression patterns described by other investigators related to biological functions important to tumorigenesis. BRC12/CD184<sup>lo</sup> and BRC13/CD49<sup>do</sup> cells expressed many genes related to hypoxia and drug resistance [34–36], (Figs 5A and S10–S13) while BRC12/CD184<sup>hi</sup> and BRC/CD133<sup>hi</sup> cells had high expression in genes indicating proliferation, DNA Damage Response and stemness (Figs 5B and S4–S9).

Focusing on specific gene families we found further evidence of co-regulation of genes in a manner consistent with functional relevance. The best such example is presented in the keratin gene family. The KRT genes 5,6, 14, 16 and 17 which appear to be co-regulated and have been used as markers for basal breast cancer [37–40] are all upregulated in BRC12/CD184<sup>lo</sup> and BRC13/CD49<sup>do</sup> cells with remarkable quantitative consistency (S3 Table). Their upregulation is also consistent with the activity of Hypoxia Inducible Factor (HIF), i.e. the presence of hypoxia [41, 42]. Furthermore, especially some genes for which evidence for functional roles in hypoxia has been demonstrated repeatedly and in cancer cells (e.g. LICAM, ANGPTL4, CA9, VEGFA, LDHA, P4HA1, LOX, LOXL2, LOXL4) are almost all significantly upregulated in both BRC12/CD184<sup>lo</sup> and BRC13/CD49<sup>do</sup> cells (S3 Table). Hypoxia is often correlated with therapeutic resistance [43, 44]. In accordance with this, gene expression signatures of drug resistance can also be found in those same populations (S8 Fig).

While functional tests clearly are required to confirm phenotypic relevance of our results we believe that these distinct distributions of overexpressed genes in separate compartments point to some very intriguing possibilities. The two mouse models we investigated showed a high degree of similarities, especially considering the functions the gene patterns are implicated in. scRNASeq has led to a much better appreciation of subtle gene expression differences between otherwise similar tissue cells and a number of those studies with TNBC result in similar gene patterns as we describe here (S9–S11 Figs). Often gradients are found across different locations within the tissue but they don't necessarily exist across straight lines which is one among multiple challenges when investigating their functional relevance. It may be useful to expand on the historical notion of what a "cell type" is by considering "cell states" with distinct functional properties [45]. In a recent publication Uri Alon and colleagues proposed a model where cells exist along different axes within a multidimensional space of functionality and each end of a given axis is defined by the presence of "specialist" cells with respect to the function in question [8]. scRNASeq is highly efficient at detecting such gradients but gene dropout effects limit its resolution and function can only be deduced from correlations. A variety of spatial methods have been developed to locate cells and reveal the tissue architecture but similar limitations apply and those studies are highly resource intensive. We believe our approach is complementing those tools by providing direct access to such "specialist" cells in quantities sufficient to carry out in depth functional investigations. Acquired cancer drug resistance might be a primary example of where this could be of major clinical use. Together with the findings of other investigators our data argue that such "division of labor" exists in breast cancer tissue where hypoxia/drug resistance may reside in one compartment while proliferation/stemness/DNA damage response may be driven by another. We show for the first time that such compartments emerge spontaneously and reproducibly in different PDX models. Some overlaps of our profiles with drug resistance signatures in other cancer types (e.g. lung, melanoma; see S8 Fig) suggests that this type of ITH may be common across solid cancer tissues. This provides powerful options for setting up prospective studies in vivo. Recent developments of PDX models with humanized immune systems [46] will likely further improve those options. Studies targeting drug response variations have been conducted in cell culture [44] and resulted in changes of many of the genes related to hypoxia and TGFβ activity described in our study.

A number of studies have described ITH regarding both phenotype as well as genotype and suggested interactions between the different compartments [47–49]. The successive engraftment of heterogeneous patient derived cancer tissue through multiple passages [47] showed continued co-existence of distinct genotypic clones growing in each new generation of host mice. The implausibility of perfectly synchronized independent genotypes suggests that clones are interdependent which is consistent with numerous reports of non-cell autonomous proliferation [50, 51]. Our data do not include DNA sequencing and it is thus unclear whether the surface marker identified subpopulations are genotypically distinct, identical or mixed themselves. The overlapping gene expression patterns we observed with the different subpopulations identified in BRC13 (Fig 4) are more consistent with a model of gradients within multidimensional gene expression space as the one pioneered by Uri Alon and colleagues [7] than with a mixture of uniform clones. In our experience surface marker profiles across different mice from the same model, including BRC12 and BRC13 remain quite stable [12]. Although this could be interpreted as a sign of stable relative ratios but it is important to keep in mind that for technical reasons most PDX tissue is analyzed within a fairly narrow range of tumor sizes. Some investigators reported observations that both the relative clonal ratios and total combined cell numbers influence growth kinetics [48] and it is certainly possible that the relative ratios might differ if the same tumors are harvested at a different stage of their growth. Studies combining our approach with DNA sequencing, lineage barcoding and propagation of clonal mixtures through multiple passages as well as analysis at different stages of tumor development provide many opportunities to clarify the respective contribution of each category of ITH described. One of the two PDX models investigated by Jeff Chuang and colleagues [47] happens to be the same one we used in our previous study (BRC12 aka TM00096) and thus would offer a particularly rich data set as background for further work. The use of short-term cultures of defined mixtures of cells as described by Bruna et al. [49] offers an additional efficient way towards such investigations. We believe our work describes a near optimal source of cells for this purpose since reproducibly emerging subpopulations with well defined gene expression patterns grown in the same tumor eliminate the confounding factors afflicting cells grown on plastic or derived from multiple patients. The two most important implications of our findings are technical and biological in nature, respectively: 1) our approach is an efficient option to investigate subpopulations at a functional level due to their accessibility through well established methods of surface marker directed FACS sorting. 2) subpopulations with distinct gene expression signatures related to hypoxia/drug response and proliferation/DNA damage response may be common across different breast cancer models prior to drug challenges. We have successfully applied FACS analysis based on these methods to patient derived tissues [52] raising the possibility that once the functional implications of the various gene profiles can be better understood optimized assays targeted to clinical samples could likely be developed.

## Supporting information

### S1 File.

(DOCX)

**S1 Fig. Flow cytometric analysis of representative BRC 13 sample.** Plots represent sample before (left) and after (right, overlay of 4 samples) sorting with CD49f and CD133, resp. These plots are representative of all replicate tumors.

(PDF)

**S2 Fig. Surface markers used for sorting and respective gene expressions.** Genes encoding the surface markers which were used to sort the populations (CD184/CXCR4, CD49f/ITGA6,

CD133/Prom1) were found to be differentially expressed with high statistical relevance in each experiment. CD184 (left) data is from a prior publication [12]. The horizontal line marks FDR = 0.1.

(PDF)

**S3 Fig. Number of genes on the y-axis shown as a function of the FDR.** Even up to very high FDR values the number of genes assessed to be different is very low for all comparisons between samples during the sample prep phase while the number of genes differentially expressed between sorted subpopulations grows continuously with increasing FDR.

(PDF)

**S4 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for breast tissue proliferation.** Overlap of a proliferative gene signature from histologically normal breast tissue (30 genes, [53], top; same as in Fig 5) and "hallmark E2F target genes" (34 genes, [54], bottom) with resp. gene sets within this publication are shown in volcano plots. Numbers provided are genes above FDR cutoff of 0.1 up (right) or down (left) regulated.

(PDF)

**S5 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for breast tissue proliferation.** Overlap of a proliferative gene signature from histologically normal breast tissue (30 genes, [53], left; same as in Fig 5) and "hallmark E2F target genes" (34 genes, [54], right) with resp. gene sets within this publication are shown in heatmaps. 2 genes (TOP2A and MCM2, see arrows) overlap between the signatures.

(PDF)

**S6 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for stemness and genomic instability.** Overlap of a stem cell gene signature [55], (top/) and a genomic instability signature identified in glioma [56] (bottom) are shown in volcano plots.

(PDF)

**S7 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for stemness and genomic instability.** Overlap of a stem cell gene signature [55], (left) and a genomic instability signature identified in glioma [56] (right) are shown in heatmaps (b).

(PDF)

**S8 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for poor outcome and DNA damage response.** Overlap of a subpopulation signature associated with poor outcome in lung adenocarcinoma [57] (top) and a signature associated with DNA damage response (DDR) and resistance to PARP inhibition [58] (bottom) are shown in volcano plots.

(PDF)

**S9 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for poor outcome and DNA damage response.** Overlap of a subpopulation signature associated with poor outcome in lung adenocarcinoma [57] (left) and a signature associated with DNA damage response (DDR) and resistance to PARP inhibition [58] (right) are shown in heatmaps (b).

(PDF)

**S10 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for hypoxia.** Overlap of a hypoxia signature composed from 3 publications (see Fig 5a, [34–



36], top) and a larger hypoxia gene set described in [59], (bottom) are shown in volcano plots. (PDF)

**S11 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for hypoxia.** Overlap of a hypoxia signature composed from 3 publications (see Fig 5a, [34–36], left) and a larger hypoxia gene set described in [59], (right) are shown in heatmaps (b). (PDF)

**S12 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for drug resistance.** Overlap of a drug resistance signature described in lung carcinoma and melanoma [60] are shown in a volcano plot. (PDF)

**S13 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations for drug resistance.** Overlap of a drug resistance signature described in lung carcinoma and melanoma [60] are shown in a heatmap. (PDF)

**S14 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations with overlap of clusters identified in scRNASeq experiments on TNBC samples.** Overlap of 19 genes associated with proliferation in a scRNASeq analysis of TNBC tissue [61]. (PDF)

**S15 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations with overlap of clusters identified in scRNASeq experiments on TNBC samples.** Overlap of 18 genes associated with an epithelial phenotype in a scRNASeq study on TNBC patient samples [62]. (PDF)

**S16 Fig. Gene expression signatures identified in BRC12 and BRC13 tumor cell subpopulations with overlap of clusters identified in scRNASeq experiments on TNBC samples.** Overlap genes identified in different clusters within single cells from the same tissue in PDX models of TNBC [63] A) 24 combined genes most upregulated in clusters 1,3,6 B) 20 combined genes most upregulated in clusters 2,4. (PDF)

**S1 Table. List of reagents used in flow sorting.** (XLSX)

**S2 Table. IEGs identified with statistical cutoff of FDR<0.1.** (XLSX)

**S3 Table. List of GO terms generated in Enrichr for DEGs upregulated in the BRC13/CD133<sup>lo</sup> subpopulation.** (XLSX)

**S4 Table. List of GO terms generated in Enrichr for DEGs upregulated in the BRC13/CD133<sup>hi</sup> subpopulation.** (XLSX)

**S5 Table. List of GO terms generated in Enrichr for DEGs upregulated in the BRC13/CD49<sup>do</sup> subpopulation.** (XLSX)

**S6 Table. List of GO terms generated in Enrichr for DEGs upregulated in the BRC13/CD49<sup>hi</sup> subpopulation.**

(XLSX)

**S7 Table. List of GO terms generated in Enrichr for DEGs upregulated in the BRC12/CD184<sup>lo</sup> subpopulation.**

(XLSX)

**S8 Table. List of GO terms generated in Enrichr for DEGs upregulated in the BRC12/CD184<sup>hi</sup> subpopulation.**

(XLSX)

**S9 Table. Select genes differentially expressed in BRC12/CD184<sup>lo</sup> and BRC13/CD49<sup>lo</sup>.** A number of members of the keratin family of genes (a) as well as functionally relevant hypoxia genes (b) were found to be regulated in a very similar fashion in BRC12 and BRC13 subpopulations.

(XLSX)

**S10 Table. Gene lists with expression profiles selected from literature.**

(XLSX)

**S11 Table. DEG list from previous experiment using BRC12 and CD184.**

(XLSX)

**S12 Table. DEG list from 2<sup>nd</sup> experiment using BRC13 and CD49f.**

(XLSX)

**S13 Table. DEG list from 2<sup>nd</sup> experiment using BRC13 and CD133.**

(XLSX)

## Acknowledgments

The authors wish to thank Drs. Jeff Baker, Eric Dixon, BDTI and Ben Stanger, University of Pennsylvania for helpful comments.

## Author Contributions

**Conceptualization:** Warren Porter, Eileen Snowden, W. Shannon Dillmore, Rainer Blaesius.

**Data curation:** Warren Porter, Frances Tong.

**Formal analysis:** Warren Porter, Eileen Snowden, Frances Tong, Rainer Blaesius.

**Investigation:** Warren Porter, Eileen Snowden, Friedrich Hahn, Mitchell Ferguson, Rainer Blaesius.

**Methodology:** Warren Porter, Eileen Snowden, Friedrich Hahn, Mitchell Ferguson, Rainer Blaesius.

**Project administration:** W. Shannon Dillmore, Rainer Blaesius.

**Software:** Warren Porter.

**Supervision:** W. Shannon Dillmore, Rainer Blaesius.

**Validation:** Warren Porter, Eileen Snowden, Frances Tong.

**Visualization:** Warren Porter, Eileen Snowden, Frances Tong, Rainer Blaesius.

**Writing – original draft:** Rainer Blaesius.

**Writing – review & editing:** Rainer Blaesius.

## References

1. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012; 12(5):323–34.
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406(6797):747–52. Epub 2000/08/30. <https://doi.org/10.1038/35021093> PMID: 10963602.
3. Heppner GH. Tumor heterogeneity. *Cancer Res*. 1984; 44(6):2259–65. Epub 1984/06/01. PMID: 6372991.
4. Heppner GH, Shekhar M. Tumor heterogeneity is fundamental to the tumor ecosystem. *Oncology (Williston Park, NY)*. 2014; 28(9):780–1. Epub 2014/09/17. PMID: 25224476.
5. Egeblad M, Nakasone ES, Werb Z. Tumors as organs: complex tissues that interface with the entire organism. *Dev Cell*. 2010; 18(6):884–901. <https://doi.org/10.1016/j.devcel.2010.05.012> PMID: 20627072.
6. Hinohara K, Polyak K. Intratumoral Heterogeneity: More Than Just Mutations. *Trends in cell biology*. 2019; 29(7):569–79. <https://doi.org/10.1016/j.tcb.2019.03.003>
7. Korem Y, Szekely P, Hart Y, Sheftel H, Hausser J, Mayo A, et al. Geometry of the Gene Expression Space of Individual Cells. *PLoS computational biology*. 2015; 11(7):e1004224. Epub 2015/07/15. <https://doi.org/10.1371/journal.pcbi.1004224> PMID: 26161936; PubMed Central PMCID: PMC4498931.
8. Adler M, Korem Kohanim Y, Tendler A, Mayo A, Alon U. Continuum of Gene-Expression Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. *Cell Systems*. 2019; 8(1):43–52.e5. <https://doi.org/10.1016/j.cels.2018.12.008>
9. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*. 2018; 174(3):716–29.e27. Epub 2018/07/03. <https://doi.org/10.1016/j.cell.2018.05.061> PMID: 29961576; PubMed Central PMCID: PMC6771278.
10. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. 2018; 562(7727):367–72. Epub 2018/10/05. <https://doi.org/10.1038/s41586-018-0590-4> PMID: 30283141; PubMed Central PMCID: PMC6642641.
11. Byrne AT, Alferez DG, Amant F, Annibaldi D, Arribas J, Biankin AV, et al. Interrogating open issues in cancer precision medicine with patient-derived xenografts. *Nat Rev Cancer*. 2017; 17(4):254–68. Epub 2017/01/21. <https://doi.org/10.1038/nrc.2016.140> PMID: 28104906.
12. Snowden E, Porter W, Hahn F, Ferguson M, Tong F, Parker JS, et al. Immunophenotyping and Transcriptomic Outcomes in PDX-derived TNBC Tissue. *Molecular Cancer Research*. 2017; 15(4):429–38. Epub 12/30/2016.
13. Krebsbach PH, Villa-Diaz LG. The Role of Integrin  $\alpha 6$  (CD49f) in Stem Cells: More than a Conserved Biomarker. *Stem cells and development*. 2017; 26(15):1090–9. Epub 2017/05/13. <https://doi.org/10.1089/scd.2016.0319> PMID: 28494695; PubMed Central PMCID: PMC5563922.
14. Brugnoli F, Grassilli S, Al-Qassab Y, Capitani S, Bertagnolo V. CD133 in Breast Cancer Cells: More than a Stem Cell Marker. *Journal of oncology*. 2019; 2019:7512632. Epub 2019/10/23. <https://doi.org/10.1155/2019/7512632> PMID: 31636668; PubMed Central PMCID: PMC6766124 publication of this paper.
15. National Research Council (US) Committee for the Update of the Guide for the Care And Use of Laboratory Animals. *Guide for the Care and Use of Laboratory Animals. Guide for the Care and Use of Laboratory Animals*. 8 ed. Washington (DC): National Academies Press (US) National Academy of Sciences.; 2011.
16. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*. 2016; 44(W1):W90–7. Epub 2016/05/05. <https://doi.org/10.1093/nar/gkw377> PMID: 27141961; PubMed Central PMCID: PMC4987924.
17. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013; 14:128. Epub 2013/04/17. <https://doi.org/10.1186/1471-2105-14-128> PMID: 23586463; PubMed Central PMCID: PMC3637064.
18. Tullai JW, Schaffer ME, Mullenbrock S, Sholder G, Kasif S, Cooper GM. Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *The Journal of biological*

- chemistry. 2007; 282(33):23981–95. Epub 2007/06/19. <https://doi.org/10.1074/jbc.M702044200> PMID: 17575275; PubMed Central PMCID: PMC2039722.
19. Moriya S, Chourasia D, Ng KW, Khel NB, Parhar IS. Cloning and localization of immediate early response 2 (ier2) gene in the brain of medaka. *Journal of chemical neuroanatomy*. 2016; 77:24–9. Epub 2016/05/03. <https://doi.org/10.1016/j.jchemneu.2016.04.005> PMID: 27134039.
  20. Jahner D, Hunter T. The ras-related gene rhoB is an immediate-early gene inducible by v-Fps, epidermal growth factor, and platelet-derived growth factor in rat fibroblasts. *Molecular and cellular biology*. 1991; 11(7):3682–90. Epub 1991/07/01. PMID: 1710770; PubMed Central PMCID: PMC361128.
  21. Cullingford TE, Butler MJ, Marshall AK, Tham el L, Sugden PH, Clerk A. Differential regulation of Kruppel-like factor family transcription factor expression in neonatal rat cardiac myocytes: effects of endothelin-1, oxidative stress and cytokines. *Biochim Biophys Acta*. 2008; 1783(6):1229–36. Epub 2008/04/15. <https://doi.org/10.1016/j.bbamcr.2008.03.007> PMID: 18406357; PubMed Central PMCID: PMC2396231.
  22. Rossi SL, Lumpkin CJ, Harris AW, Holbrook J, Gentillon C, McCahan SM, et al. Identification of early gene expression changes in primary cultured neurons treated with topoisomerase I poisons. *Biochemical and biophysical research communications*. 2016; 479(2):319–24. Epub 2016/09/20. <https://doi.org/10.1016/j.bbrc.2016.09.068> PMID: 27641670; PubMed Central PMCID: PMC5048596.
  23. Waelput W, Verhee A, Broekaert D, Eyckerman S, Vandekerckhove J, Beattie JH, et al. Identification and expression analysis of leptin-regulated immediate early response and late target genes. *The Biochemical journal*. 2000; 348 Pt 1:55–61. Epub 2000/05/05. PMID: 10794713; PubMed Central PMCID: PMC1221035.
  24. Pirola L, Balcerzyk A, Tothill RW, Haviv I, Kaspi A, Lunke S, et al. Genome-wide analysis distinguishes hyperglycemia regulated epigenetic signatures of primary vascular cells. *Genome Res*. 2011; 21(10):1601–15. Epub 2011/09/06. <https://doi.org/10.1101/gr.116095.110> PMID: 21890681; PubMed Central PMCID: PMC3202278.
  25. Qiao X, Lu JY, Hofmann SL. Gene expression profiling in a mouse model of infantile neuronal ceroid lipofuscinosis reveals upregulation of immediate early genes and mediators of the inflammatory response. *BMC neuroscience*. 2007; 8:95. Epub 2007/11/21. <https://doi.org/10.1186/1471-2202-8-95> PMID: 18021406; PubMed Central PMCID: PMC2204004.
  26. Qiu LQ, Lai WS, Bradbury A, Zeldin DC, Blackshear PJ. Tristetraprolin (TTP) coordinately regulates primary and secondary cellular responses to proinflammatory stimuli. *Journal of leukocyte biology*. 2015; 97(4):723–36. Epub 2015/02/07. <https://doi.org/10.1189/jlb.3A0214-106R> PMID: 25657290; PubMed Central PMCID: PMC4370050.
  27. Kelly K, Siebenlist U. Immediate-early genes induced by antigen receptor stimulation. *Current opinion in immunology*. 1995; 7(3):327–32. Epub 1995/06/01. PMID: 7546396.
  28. Sansone P, Berishaj M, Rajasekhar VK, Ceccarelli C, Chang Q, Strillacci A, et al. Evolution of Cancer Stem-like Cells in Endocrine-Resistant Metastatic Breast Cancers Is Mediated by Stromal Microvesicles. *Cancer Res*. 2017; 77(8):1927–41. Epub 2017/02/17. <https://doi.org/10.1158/0008-5472.can-16-2129> PMID: 28202520; PubMed Central PMCID: PMC5392366.
  29. Gomez-Miragaya J, Palafox M, Pare L, Yoldi G, Ferrer I, Vila S, et al. Resistance to Taxanes in Triple-Negative Breast Cancer Associates with the Dynamics of a CD49f+ Tumor-Initiating Population. *Stem cell reports*. 2017; 8(5):1392–407. Epub 2017/05/02. <https://doi.org/10.1016/j.stemcr.2017.03.026> PMID: 28457887; PubMed Central PMCID: PMC5425727.
  30. Vassilopoulos A, Chisholm C, Lahusen T, Zheng H, Deng CX. A critical role of CD29 and CD49f in mediating metastasis for cancer-initiating cells isolated from a Brca1-associated mouse model of breast cancer. *Oncogene*. 2013; 33(47). Epub 2013/12/10. <https://doi.org/10.1038/onc.2013.516> PMID: 24317509; PubMed Central PMCID: PMC24317509.
  31. Wang E. RNA amplification for successful gene profiling analysis. *Journal of translational medicine*. 2005; 3:28-. <https://doi.org/10.1186/1479-5876-3-28> PMID: 16042807.
  32. Bahrami S, Drablos F. Gene regulation in the immediate-early response process. *Advances in biological regulation*. 2016; 62:37–49. Epub 2016/05/26. <https://doi.org/10.1016/j.jbior.2016.05.001> PMID: 27220739.
  33. Sironen RK, Karjalainen HM, Elo MA, Kaarniranta K, Torronen K, Takigawa M, et al. cDNA array reveals mechanosensitive genes in chondrocytic cells under hydrostatic pressure. *Biochim Biophys Acta*. 2002; 1591(1–3):45–54. Epub 2002/08/17. PMID: 12183054.
  34. Casillas AL, Toth RK, Sainz AG, Singh N, Desai AA, Kraft AS, et al. Hypoxia-Inducible PIM Kinase Expression Promotes Resistance to Antiangiogenic Agents. *Clin Cancer Res*. 2018; 24(1):169–80. Epub 2017/11/01. <https://doi.org/10.1158/1078-0432.ccr-17-1318> PMID: 29084916; PubMed Central PMCID: PMC6214353.
  35. Amelio I, Mancini M, Petrova V, Cairns RA, Vikhrev P, Nicolai S, et al. p53 mutants cooperate with HIF-1 in transcriptional regulation of extracellular matrix components to promote tumor progression.

- Proc Natl Acad Sci U S A. 2018; 115(46):E10869–e78. Epub 2018/11/02. <https://doi.org/10.1073/pnas.1808314115> PMID: 30381462; PubMed Central PMCID: PMC6243248.
36. Lu H, Tran L, Park Y, Chen I, Lan J, Xie Y, et al. Reciprocal Regulation of DUSP9 and DUSP16 Expression by HIF1 Controls ERK and p38 MAP Kinase Activity and Mediates Chemotherapy-Induced Breast Cancer Stem Cell Enrichment. *Cancer Res.* 2018; 78(15):4191–202. Epub 2018/06/09. <https://doi.org/10.1158/0008-5472.can-18-0270> PMID: 29880481.
  37. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome biology.* 2007; 8(8):R157-R. <https://doi.org/10.1186/gb-2007-8-8-r157> PMID: 17683518.
  38. Hoggarth ZE, Osowski DB, Freeberg BA, Garrett SH, Sens DA, Sens MA, et al. The urothelial cell line UROtsa transformed by arsenite and cadmium display basal characteristics associated with muscle invasive urothelial cancers. *PLoS One.* 2018; 13(12):e0207877. Epub 2018/12/15. <https://doi.org/10.1371/journal.pone.0207877> PMID: 30550540; PubMed Central PMCID: PMC6294394.
  39. Pan X, Hobbs RP, Coulombe PA. The expanding significance of keratin intermediate filaments in normal and diseased epithelia. *Current opinion in cell biology.* 2013; 25(1):47–56. Epub 2012/12/29. <https://doi.org/10.1016/j.ceb.2012.10.018> PMID: 23270662; PubMed Central PMCID: PMC3578078.
  40. Loschke F, Seltmann K, Bouameur JE, Magin TM. Regulation of keratin network organization. *Current opinion in cell biology.* 2015; 32:56–64. Epub 2015/01/17. <https://doi.org/10.1016/j.ceb.2014.12.006> PMID: 25594948.
  41. Xing D, Liu L, Marti GP, Zhang X, Reinblatt M, Milner SM, et al. Hypoxia and hypoxia-inducible factor in the burn wound. *Wound Repair Regen.* 2011; 19(2):205–13. <https://doi.org/10.1111/j.1524-475X.2010.00656.x> PMID: 21362088.
  42. Krishnamachary B, Berg-Dixon S, Kelly B, Agani F, Feldser D, Ferreira G, et al. Regulation of colon carcinoma cell invasion by hypoxia-inducible factor 1. *Cancer Res.* 2003; 63(5):1138–43. Epub 2003/03/05. PMID: 12615733.
  43. Sullivan R, Pare GC, Frederiksen LJ, Semenza GL, Graham CH. Hypoxia-induced resistance to anti-cancer drugs is associated with decreased senescence and requires hypoxia-inducible factor-1 activity. *Molecular cancer therapeutics.* 2008; 7(7):1961–73. Epub 2008/07/23. <https://doi.org/10.1158/1535-7163.mct-08-0198> PMID: 18645006.
  44. O'Brien SK, Chen L, Zhong W, Armellino D, Yu J, Loreth C, et al. Breast cancer cells respond differentially to modulation of TGFbeta2 signaling after exposure to chemotherapy or hypoxia. *Cancer Res.* 2015; 75(21):4605–16. Epub 2015/09/06. <https://doi.org/10.1158/0008-5472.can-15-0650> PMID: 26340918.
  45. Su Y, Wei W, Robert L, Xue M, Tsoi J, Garcia-Diaz A, et al. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. *Proc Natl Acad Sci U S A.* 2017; 114(52):13679–84. Epub 2017/12/13. <https://doi.org/10.1073/pnas.1712064115> PMID: 29229836; PubMed Central PMCID: PMC5748184.
  46. Wang M, Yao LC, Cheng M, Cai D, Martinek J, Pan CX, et al. Humanized mice in studying efficacy and mechanisms of PD-1-targeted cancer immunotherapy. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology.* 2018; 32(3):1537–49. Epub 2017/11/18. <https://doi.org/10.1096/fj.201700740R> PMID: 29146734; PubMed Central PMCID: PMC5892726.
  47. Kim H, Kumar P, Menghi F, Noorbakhsh J, Cerveira E, Ryan M, et al. High-resolution deconstruction of evolution induced by chemotherapy treatments in breast cancer xenografts. *Scientific reports.* 2018; 8(1):17937. Epub 2018/12/19. <https://doi.org/10.1038/s41598-018-36184-8> PMID: 30560892; PubMed Central PMCID: PMC6298990.
  48. Nguyen LV, Cox CL, Eirew P, Knapp DJ, Pellacani D, Kannan N, et al. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nature communications.* 2014; 5:5871. Epub 2014/12/24. <https://doi.org/10.1038/ncomms6871> PMID: 25532760; PubMed Central PMCID: PMC4284657.
  49. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. *Cell.* 2016; 167(1):260–74.e22. Epub 2016/09/20. <https://doi.org/10.1016/j.cell.2016.08.041> PMID: 27641504; PubMed Central PMCID: PMC5037319.
  50. Janiszewska M, Tabassum DP, Castano Z, Cristea S, Yamamoto KN, Kingston NL, et al. Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. *Nature cell biology.* 2019; 21(7):879–88. Epub 2019/07/03. <https://doi.org/10.1038/s41556-019-0346-x> PMID: 31263265; PubMed Central PMCID: PMC6609451.
  51. Weber RJ, Desai TA, Gartner ZJ. Non-autonomous cell proliferation in the mammary gland and cancer. *Current opinion in cell biology.* 2017; 45:55–61. Epub 2017/03/18. <https://doi.org/10.1016/j.ceb.2017.02.009> PMID: 28314237.



52. Sen M, Hahn F, Black TA, DeMarshall M, Porter W, Snowden E, et al. Flow based single cell analysis of the immune landscape distinguishes Barrett's esophagus from adjacent normal tissue. *Oncotarget*. 2019; 10(38):3592–604. Epub 2019/06/21. PMID: [31217895](https://pubmed.ncbi.nlm.nih.gov/31217895/); PubMed Central PMCID: PMC6557213.
53. Chen DT, Nasir A, Culhane A, Venkataramu C, Fulp W, Rubio R, et al. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. *Breast cancer research and treatment*. 2010; 119(2):335–46. Epub 2009/03/07. <https://doi.org/10.1007/s10549-009-0344-y> PMID: [19266279](https://pubmed.ncbi.nlm.nih.gov/19266279/); PubMed Central PMCID: PMC2796276.
54. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc Natl Acad Sci U S A*. 2015; 112(47):E6544–52. Epub 2015/10/16. <https://doi.org/10.1073/pnas.1518007112> PMID: [26460041](https://pubmed.ncbi.nlm.nih.gov/26460041/); PubMed Central PMCID: PMC4664352.
55. Shats I, Gatz ML, Chang JT, Mori S, Wang J, Rich J, et al. Using a stem cell-based signature to guide therapeutic selection in cancer. *Cancer Res*. 2011; 71(5):1772–80. Epub 2010/12/21. <https://doi.org/10.1158/0008-5472.can-10-1735> PMID: [21169407](https://pubmed.ncbi.nlm.nih.gov/21169407/); PubMed Central PMCID: PMC3049992.
56. Zhang Y, Li J, Yi K, Feng J, Cong Z, Wang Z, et al. Elevated signature of a gene module coexpressed with CDC20 marks genomic instability in glioma. *Proc Natl Acad Sci U S A*. 2019; 116(14):6975–84. Epub 2019/03/17. <https://doi.org/10.1073/pnas.1814060116> PMID: [30877245](https://pubmed.ncbi.nlm.nih.gov/30877245/); PubMed Central PMCID: PMC6452696.
57. Min JW, Kim WJ, Han JA, Jung YJ, Kim KT, Park WY, et al. Identification of Distinct Tumor Subpopulations in Lung Adenocarcinoma via Single-Cell RNA-seq. *PLoS One*. 2015; 10(8):e0135817. Epub 2015/08/26. <https://doi.org/10.1371/journal.pone.0135817> PMID: [26305796](https://pubmed.ncbi.nlm.nih.gov/26305796/).
58. Hassan S, Esch A, Liby T, Gray JW, Heiser LM. Pathway-Enriched Gene Signature Associated with 53BP1 Response to PARP Inhibition in Triple-Negative Breast Cancer. *Molecular cancer therapeutics*. 2017; 16(12):2892–901. Epub 2017/09/30. <https://doi.org/10.1158/1535-7163.mct-17-0170> PMID: [28958991](https://pubmed.ncbi.nlm.nih.gov/28958991/); PubMed Central PMCID: PMC5765867.
59. Harris AL. Hypoxia—a key regulatory factor in tumour growth. *Nat Rev Cancer*. 2002; 2(1):38–47. Epub 2002/03/21. <https://doi.org/10.1038/nrc704> PMID: [11902584](https://pubmed.ncbi.nlm.nih.gov/11902584/).
60. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016; 352(6282):189–96. <https://doi.org/10.1126/science.aad0501>
61. Yang B, Chou J, Tao Y, Wu D, Wu X, Li X, et al. An assessment of prognostic immunity markers in breast cancer. *NPJ Breast Cancer*. 2018; 4:35-. <https://doi.org/10.1038/s41523-018-0088-0> PMID: [30393759](https://pubmed.ncbi.nlm.nih.gov/30393759/).
62. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nature communications*. 2018; 9(1):3588. Epub 2018/09/06. <https://doi.org/10.1038/s41467-018-06052-0> PMID: [30181541](https://pubmed.ncbi.nlm.nih.gov/30181541/); PubMed Central PMCID: PMC6123496.
63. Savage P, Blanchet-Cohen A, Revil T, Badescu D, Saleh SMI, Wang YC, et al. A Targetable EGFR-Dependent Tumor-Initiating Program in Breast Cancer. *Cell reports*. 2017; 21(5):1140–9. Epub 2017/11/02. <https://doi.org/10.1016/j.celrep.2017.10.015> PMID: [29091754](https://pubmed.ncbi.nlm.nih.gov/29091754/).