SunCatcher: Clonal Barcoding with qPCR Detection Enables Live Cell Functional Analyses to Study Cancer Heterogeneity, Progression, and Metastasis

Sandra McAllister (✉ smcallister1@bwh.harvard.edu)

Brigham and Women's Hospital / Harvard Medical School <https://orcid.org/0000-0002-5111-5903>

Qiuchen Guo

Brigham and Women's Hospital / Harvard Medical School

Milos Spasic

Brigham and Women's Hospital / Harvard Medical School

Adam Maynard

Brigham and Women's Hospital / Harvard Medical School

Gregory Goreczny

Brigham and Women's Hospital / Harvard Medical School

Amanuel Bizuayehu

Brigham and Women's Hospital / Harvard Medical School

Jessica Olive

Brigham and Women's Hospital / Harvard Medical School

Article

Keywords:

Posted Date: February 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1262178/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on July 4th, 2022.
See the published version at <https://doi.org/10.1038/s41467-022-31536-5>.

Abstract

Single cell tracing and “omics” methods are valuable research tools; however, current approaches do not easily enable live cell retrieval. That is a particular issue when further study of cells that were eliminated or otherwise not selected during experimentation could provide critical information. We report a clonal molecular barcoding method, called SunCatcher, that enables longitudinal tracking and functional analysis of live cells. Single cell-derived clonal populations are generated from complex cell populations, infected with a unique molecular barcode, and stocks of each barcoded clone (BC) are retained. BCs can be combined to create a version of the original cell population or to generate custom BC pools. We developed both quantitative PCR-based and next-generation sequencing methods for identifying and quantifying BCs *in vitro* and *in vivo*. We applied SunCatcher to various breast cancer cell lines. Heterogeneous pools of BCs reliably reproduced the proliferation and cancer progression rates of the original parental cell lines. Individual BCs, however, displayed a diversity of proliferation rates and molecular phenotypes that associated with their positive or negative selection *in vivo*. SunCatcher also enabled identification and quantification of early spontaneous metastases that were not detected by conventional methods. We envision that SunCatcher will serve as a reliable and widely accessible research and discovery tool.

Introduction

Cell fate mapping and lineage tracing techniques have been employed for several decades and have been useful for describing heterogeneity and clonal dynamics of complex tissues in both normal and pathological settings. Early insights into hematopoiesis were derived from analyzing karyotypes, assessing heterozygosity of particular genes, and analyzing viral integration patterns^{1–3}. Although limited by their lack of sensitivity and qualitative nature, those techniques enabled the first effective tracking of clonal populations.

Understanding cellular heterogeneity and clonal fitness is particularly important to cancer research. Clinically, intratumoral diversity inversely correlates with therapeutic response, metastatic ability, and patient survival^{4–6}. Although differences in the mutational profile of tumor cell subpopulations may be the best-documented parameter of tumor heterogeneity, it is ultimately the functional heterogeneity, a result of both genetic and non-genetic sources of heterogeneity, that impacts the course of disease progression^{7–10}. Large-scale sequencing efforts and technologies such as single-cell RNA sequencing and single-nucleus sequencing have revealed the diverse mutational and transcriptional landscape of human cancers^{11–15}. Nevertheless, we still have much to learn about the effects of functional heterogeneity on disease progression.

Accounting for and properly modeling tumor heterogeneity and cell plasticity in experimental and pre-clinical settings is thus crucial for improving cancer treatment outcomes. For example, we previously reported that functional heterogeneity can impact data interpretation in studies that employ gene editing and designed a novel CRISPR/Cas9 protocol that incorporates a single-cell cloning step prior to gene

editing in order to generate appropriately matched control and edited cells¹⁶. Moreover, clonal dynamics—driven by clone frequency, spatial proximity, heterotypic interactions, and/or cell plasticity—can profoundly affect disease progression^{4, 10, 17, 18}.

The advent of molecular barcoding techniques, in which short, non-coding DNA sequences are stably integrated into the genomes of complex cell populations using viral infection, enables simultaneous tracking of thousands of clonal populations^{19–21}. In cancer research, DNA barcoding techniques have been used to successfully trace clonal dynamics in heterogeneous tumor cell populations^{22–29}. Those tracing techniques utilize barcode library infection methods in which 10^4 – 10^7 barcodes are introduced into a population of tumor cells. While such high-complexity barcode libraries enable longitudinal quantification of clonal diversity in various experimental settings, they do not enable characterization, functional analyses, or manipulation of the cellular populations of interest. More recently, a number of elegant and sophisticated methods have been developed to trace as well as study certain select barcoded cells^{30–34}; however, some of those methods are limited in the numbers of barcoded cells that can be selected for study, and the techniques are not widely accessible.

Here, we report a novel molecular barcoding approach, termed SunCatcher, that provides a rapid, inexpensive, and highly sensitive method to detect, identify, quantify, isolate, and study individual clones, or mixtures of clones, of interest. SunCatcher is based on generating single-cell-derived clonal populations from any complex cell population and infecting each clonal population with a unique, heritable molecular barcode. The barcoded clones can be studied individually or combined to generate custom pools of clones. As such, SunCatcher enables both longitudinal and retrospective analysis of the molecular, phenotypic, and functional basis for heterogeneity and clonal fitness during experimentation. Importantly, our approach enables the study of clones that are negatively selected during a given process, allowing their specific vulnerabilities to be uncovered, which is not possible with conventional molecular barcoding techniques.

We applied SunCatcher to barcode individual clones from various mouse and human breast cancer cell lines. Heterogeneous pools of barcoded cells reliably reproduced the original *in vitro* and *in vivo* growth and tumor progression rates of the original parental cell lines. The SunCatcher approach was also highly effective for detecting and identifying early spontaneous skeletal and visceral metastases, and cells that underwent immune editing or escape during disease progression. We envision that SunCatcher can be applied to any cell-based studies and believe that it will serve as a useful, widely accessible tool for research and discovery.

Results

Development of the SunCatcher Clonal Barcoding Method

We developed a clonal barcoding system, which we call SunCatcher, that enables longitudinal and retrospective analyses of live cells of interest. First, single cells are isolated from a heterogeneous

parental population of cells, by either dilution passage or fluorescence-activated cell sorting (FACS) based on cell-surface markers of interest, into individual wells of 96-well plates (Fig. 1A). Each single-cell-derived colony is then expanded, after which aliquots of cells from each clonal population are viably frozen to generate stocks of non-barcoded clones (NBCs). Maintenance of NBC stocks is useful should any challenges arise in subsequent steps of the protocol resulting in loss of a given clonal population, in which case that clone can be recovered.

Each NBC population is then infected with a unique molecular barcode (Fig. 1A) that enables each clonal population to be tracked and quantified. Here, we used the PRISM collection of distinct lentiviral vectors that each contains a unique, heritable 24-base-pair (bp) DNA barcode sequence designed to avoid sequence homology with the genome^{28, 35}. We refer to these cells as barcoded polyclonal populations (BPPs; Fig. 1A) since the barcodes are randomly inserted into the genome. To ensure the clonality of the barcode insertion site, a subsequent round of single cell cloning of each BPP is performed; those subclones are expanded and aliquots of each are viably frozen (Fig. 1A). Next, one population for each barcode is randomly selected to represent the barcoded clone (BC). At this stage, each BC has a single, unique 24-bp barcode and is clonal for the lentiviral insertion site. Each BC population is aliquoted into cryovials and viably frozen to generate BC stocks. The BCs can be analyzed individually or mixed in various combinations to create custom barcoded pools (BC Pools) that can be used in any experiment or assay (Fig. 1A, B).

Our own research is focused on breast cancer; hence, we applied the SunCatcher clonal barcoding method to several murine and human breast cancer cell lines. Breast cancer subtypes are defined based on hormone receptor expression and amplification of the Her2/neu oncogene and our parental cell lines were each originally derived from tumors of a distinct subtype. The cell lines we used include the McNeuA Her2-positive³⁶, 4T1 triple-negative breast cancer (TNBC)³⁷, and Met1 TNBC³⁸ murine mammary carcinoma cell lines, as well as the human HMLER-HR TNBC cell line³⁹. Only NBCs and BCs that were able to proliferate and form a colony were saved; not all single cells survived the subcloning process. We also discarded wells in which we observed more than one cell per well after the cell sorting/dilution plating step. Following the SunCatcher barcoding protocol, we typically achieved 30-40 different BCs per cancer cell line (Table S1). For each cell line, we combined an equal number of each BC to generate breast cancer cell BC Pools (Table S1).

Examining the clonal heterogeneity (i.e., BC composition) before, during, or after any *in vitro* or *in vivo* assay is achieved by first isolating genomic DNA and PCR-amplifying common barcode flanking regions. Several methods can then be used to deconvolute the barcode signals, including the previously described PRISM analysis method,²⁸ a specialized next-generation sequencing (NGS) assay, and/or a low-cost, rapid qPCR-based approach (Fig. 1B) that we developed. The deconvolution methods are described in detail below.

Accurate qPCR-Based Barcode Deconvolution Method

We reasoned that qPCR would be a rapid, inexpensive method to identify and quantify barcodes from our BC Pools owing to the smaller library size of SunCatcher compared to traditional barcode libraries, which typically contain 10^4 - 10^8 barcodes^{24, 27, 32, 34}. We first designed oligonucleotide primers to common barcode flanking sequences to enable a pre-amplification step (Fig. 2A; Table S2). The pre-amplification step has several advantages. First, a separate PCR reaction is required to detect each barcode, so pre-amplification ensures sufficient gDNA input material, particularly from precious samples. Second, it adds a layer of specificity for BC detection above any potential background signal from the non-barcoded genome. Finally, it enriches all barcodes, including any very low abundance barcodes, thus enabling quantification of their contribution to any given sample. We also designed oligonucleotides specific to each barcode sequence in order to detect each individual BC (Fig. 2A; Table S2).

To ensure the specificity of our PCR detection method, we tested each individual barcode oligonucleotide primer against each individual BC population as well as the Met1 BC Pool. Every barcode was detected in the BC pool, while each oligonucleotide primer amplified only its specific barcode, as indicated by analyzing the C_t values for each reaction (Figure 2B). Therefore, qPCR-based deconvolution offered a fast, inexpensive, and reliable method for accurately detecting individual barcodes from a polyclonal mixture of cells.

Multiplex Next-Generation Sequencing-Based Barcode Deconvolution Method

We sought to develop a next-generation sequencing (NGS)-based method of barcode detection in cases where multiplexing is desired. To do so, we first confirmed that BC sequences could be detected by Sanger sequencing using primers against the lentiviral barcode vector further upstream and downstream than those used for PCR amplification prior to PRISM-based barcode detection (Fig. S1A). Using this optimized DNA extraction and PCR amplification protocol, we validated that all 30 distinct HMLER-HR BCs could be amplified and detected using Sanger sequencing (Fig. S1B, C).

To maximize the number of samples that we could sequence per Illumina sequencer lane, we chose to use multiple levels of multiplexing within our Illumina library preparation strategy. Twenty-four uniquely indexed Illumina adaptors are commercially available, which enabled us to pool up to 24 samples within a single sequencing lane (see Methods). Systems that allow for further multiplexing through addition of a second index during PCR amplification have been previously described⁴⁰. We chose to introduce a second PCR step into our library preparation work-flow during which one of 20 unique indexes would be added to the 3' end of the barcode amplicon (Fig. 2C, Table S3). Following PCR purification of the indexed barcode amplicons, we pooled equal molar amounts of each amplicon (Fig. 2C). Each barcode-index pool is then inserted into an indexed Illumina adaptor via a ligation reaction (Fig. 2C). Up to 24 unique barcode-index pools can be paired with the indexed Illumina adaptors, thereby allowing for the sequencing and deconvolution of up to 480 unique samples (20 barcode-index pairs x 24

indexed Illumina adaptors) in a single sequencing lane. Identification of the barcodes present within each sample thus requires deconvolution by first isolating all barcode reads associated with a single Illumina index and then isolating all barcode reads within a single Illumina index that are associated with the specific barcode-index associated with that sample.

We tested each indexed primer for its ability to successfully amplify the barcode region of gDNA extracted from HMLER-HR BCs, and then used Sanger sequencing to verify that the index had been incorporated. All 20 indexes could be added to the 3' end of barcode amplicons using PCR. We then tested whether our library preparation method could be used to accurately detect and deconvolute known barcode-index pairs that had been ligated into indexed Illumina adaptors. We used PCR to pair each of 17 indexes with 2 different barcode sequences to generate 17 barcode-index pools. We used input DNA from single BCs, so that a single known DNA barcode would be present in each sample. Additionally, we indexed DNA samples from the HMLER-HR BC Pool (which contains 30 DNA barcodes) with 3 separate indexes to examine a situation where a larger number of barcodes are associated with each index.

For Illumina library preparation, each of the 20 barcode-index pools was ligated to a uniquely indexed Illumina adaptor; the adaptors have identical P5 regions, and their P7 regions each contain a unique TruSeq Index (Fig. 2C). The 20 multiplexed samples were then mixed and run on a single lane of an Illumina MiSeq sequencer. In order to adjust for the low sequence diversity present in the sample, 40% PhiX was spiked into the sequencing reaction to increase the sample diversity⁴¹.

If ligation efficiency is low, one cannot use PCR to amplify the ligation products using primers against the Illumina P5/P7 adaptor sequences because every barcode in the pooled sample would get equally amplified. Therefore, as an alternative to ligation-based Illumina library preparation methods, we attached the Illumina adaptor regions to the amplicons being sequenced using a PCR-based library preparation method. To pursue this method, we designed a single forward PCR primer with complementarity to the barcode vector and 20 reverse primers that each contained a unique Illumina TruSeq index (Table S4).

In order to evaluate whether this method enabled accurate identification of the expected barcode-Illumina index pairs, we generated a test library containing DNA samples from individual HMLER-HR BCs. The PCR reactions for each primer set were highly specific and efficient, generating a single, strong band of the expected length (Fig. S2A). Importantly, the expected barcode-Illumina index pairs generated signal that was ~233-2,500-fold higher than the highest false positive signal, allowing for simple thresholding of true signal from background sequencer noise (Fig. 2D, S2B). These data demonstrated that the barcodes were amenable to next-generation sequencing-based detection.

Phenotypic and Biological Properties of Parental Tumor Cells are Maintained during SunCatcher Barcoding

It was important for us to understand whether the SunCatcher barcoding approach altered the complexity and important tumorigenic properties of the original parental cancer cell lines. To evaluate population complexity, we monitored cell morphology during the subcloning process. The BCs from each cell line exhibit a spectrum of cell morphologies, ranging from uniformly epithelial (e.g., Met1 BC45; Fig. 3A) to uniformly mesenchymal (e.g. Met1 BC26; Fig. 3A) phenotypes. Some BCs have mixed morphologies (e.g., Met 1 BC73 has both cobblestone epithelial and spindle-like mesenchymal phenotypes; Fig. 3A). We also confirmed that the BC Pools, like their respective parental lines, are morphologically diverse (Fig. 3A, Fig. S3A).

We next evaluated BC Pool composition and growth over a period of 7 days (1 passage) in culture using our qPCR-based deconvolution method. The Met1 BC Pool composition shifted during the time course; while some clones expanded (e.g., BC45), no clones dropped below the limit of detection (Fig. 3B).

Importantly, the proliferation rate of the Met1 BC Pool was nearly identical to that of the parental Met1 cell line (Fig. 3C), as was the HMLER-HR BC Pool relative to the HMLER-HR parental line (Fig S3B).

It was most important for us to learn whether the SunCatcher “deconstruction-reconstruction” approach affected *in vivo* tumor biology. To do so, we orthotopically injected one cohort of FVB mice with 2.5×10^5 Met1 Parental cells and another cohort with 2.5×10^5 Met1 BC Pool cells (n=9 per cohort) prepared after 7 days (1 passage) in culture. There were no significant differences between cohorts in tumor growth kinetics or in the final mass of tumors or spleens at the experimental end point (Fig. 3D-F). Likewise, we observed no significant difference in tumor growth or final tumor mass between the parental 4T1 cells and the 4T1 BC Pool (Fig. S3C, D).

Collectively, these data revealed that the SunCatcher clonal barcoding approach preserved the heterogeneity as well as *in vitro* and *in vivo* growth kinetics of the original tumor cell lines from which they were derived. Moreover, subpopulations of just ~30 clones were sufficient to recapitulate phenotypic and functional properties of their respective parental populations.

Inter- and Intra-Individual Clonal Composition of Met1 TNBC Tumors is Reproducibly Consistent

Having confirmed that the SunCatcher clonal barcoding approach retained the tumor growth kinetics of the parental cell lines, we asked how barcode composition compared across individual tumors of a given cohort and whether clonal composition is stochastic or consistent. To do so, we orthotopically injected 2.5×10^5 Met1 BC Pool cells, prepared after 7 days (1 passage) in culture, into contralateral mammary fat pads of FVB mice (n=10 tumors; n=5 mice). As a control, we analyzed a sample from the injected cells that was collected when the cells were prepared for injection (“pre-injection” sample) and a sample taken after the last mouse was injected (“post-injection” sample) to determine if the BC composition shifted during the time (~2 hr) that it took to perform all injections. The composition of the pre- and post-injection samples were nearly identical (Table S5); hence, all mice in the cohort received the same input (Fig. 3G, Table 1).

After a growth period of 18 days, we harvested the tumors, isolated genomic DNA, and used our qPCR-based method to quantify the contribution of each barcode to each tumor. We observed remarkable similarity in BC composition across all 10 tumors in the cohort despite differences in tumor mass (Fig. 3G), suggesting that the tumor-forming capacity of this pre-clinical TNBC model is not stochastic. Given the consistency between tumors, we were able to represent the average BC composition for the entire cohort by calculating the average contribution of all detected BCs (Fig. 3G, Table 1).

We repeated the orthotopic tumor growth experiment to examine the reproducibility of the results between experiments. For BCs that constituted $>0.5\%$ of the total tumor composition, we observed concordance between results from the 2 different experiments ($R^2=0.9099$; Fig. 3H). There was more variability within and between experiments for BCs that represented $<0.5\%$ of the total tumor BC composition ($R^2=0.5303$; Fig 3I). The reproducibility of our results suggested that there was phenotypic stability among the clones *in vivo*.

We also directly compared results from NGS deconvolution to those from qPCR deconvolution on the same tumors to determine whether there would be differences in sensitivity and/or accuracy between the two methods. For this comparison, we analyzed 6 tumors derived from the McNeu BC Pool (Ave tumor mass $280.3 \text{ mg} \pm 112.6 \text{ mg}$). We found significant concordance and correlation in barcode composition between methods ($R^2 = 0.975$; $p<0.0001$; Figure S3E). The qPCR-based detection method was more sensitive than NGS for low-abundance barcodes because NGS requires thresholding to distinguish between low-level false positive signals and signal generated by true, low-frequency barcode variants (Fig. S3E).

These data indicated that intra- and inter-individual primary tumor formation and heterogeneity is remarkably consistent in the Met1 pre-clinical TNBC model. Such consistency enables the study of both the fittest and negatively selected clones to elucidate not only the properties that enable their formation, but also the vulnerabilities of those clones that were negatively selected.

SunCatcher Enables Analysis of Clonal Dynamics During Proliferation *in vitro* and Tumor Progression *in vivo*

It has been hypothesized that proliferation rates dictate clonal dynamics and that the most proliferative clones will dominate a tumor^{4,42}. However, what provides the greatest selective pressure during disease progression has not been well elucidated. For example, it is not clear whether the most dominant clones in a tumor are inherently more proliferative or whether their emergence occurs via a selective process. As a first step toward answering that question, we analyzed the proliferation rate of each Met1 BC monoclonal population *in vitro*. Cells were counted every 72 hours over a 9-day time course to calculate a population doubling value per day and BCs were ranked from highest to lowest doubling rate (Fig. 4A). The most rapidly proliferating clone, BC25, doubled at a rate ~ 2.3 -fold that of the slowest proliferating clone, BC20 (Fig. 4A, Table S6).

We then examined how BC proliferation rates related to their contribution to the BC Pool *in vitro*. To do so, we calculated the fold-change in contribution of each BC to the BC Pool after 7 days (1 passage) in culture relative to the composition at the time the BC Pool was created (i.e., admixture of equal numbers of each BC) (Table 1). Four clones (BC20, BC45, BC53, and BC67) each underwent >2-fold expansion within the BC Pool *in vitro*; four clones (BC7, BC19, BC36, and BC43) exhibited modest increases or no change in their contribution over the 7-day period; the contribution of the remaining 23 clones decreased to varying degrees (Fig. 4A, Table 1). These analyses revealed that the relative proliferation rates of individual BCs did not predict their behavior within the BC Pool in culture. For example, the most proliferative clone, BC25, underwent a 4-fold decrease in contribution, while the least proliferative clone, BC20, underwent a 2.6-fold expansion within the BC Pool in culture (Fig. 4A, Table 1).

We assessed clonal dynamics *in vivo* by calculating the composition of BC Pool tumors at the 18-day end point and relating those values to the BC Pool composition at the time of injection (Table 1). Those analyses revealed four BCs that underwent expansion *in vivo* (BC2, BC8, BC25, and BC67), one BC that maintained its relative contribution (BC3), and 26 BCs that underwent negative selection to varying degrees (Fig. 4A, Table 1). Unlike the *in vitro* clonal dynamics in which BC25 underwent a 4.1-fold reduction, BC25 expanded 5.5-fold within the tumors (Fig. 4A, Table 1). In contrast, BC20, which had expanded 2.6-fold in culture, experienced a 120-fold decrease in contribution to the tumors (Fig. 4A, Table 1). While the *in vivo* behavior of these two clones (BC20 and BC25) reflected their relative *in vitro* monoclonal proliferation rates, those same trends were not observed in all clones. For example, BC73, the second-most proliferative clone *in vitro*, underwent negative selection within the tumors, with a ~20,000-fold reduction in contribution (Fig. 4A, Table 1).

Three clones together consistently comprised ~90% of total BC signal detected in the tumors at the experimental end point: BC8 ($21.2 \pm 4.8\%$), BC53 ($10.7 \pm 2.8\%$), and BC67 ($58.6 \pm 3.7\%$) (Fig. 4A; Table 1). BC53 and BC67 were among the most dominant clones in the BC Pool at the time of injection (25.6% and 20.5%, respectively) and had similar growth rates as monoclonal cultures and within the BC Pool in culture (7.9- and 6.4-fold increases, respectively) (Fig. 4A, Table 1). Hence, despite their similar behavior in culture, the contribution of BC53 contracted ~58% while BC67 expanded ~185% *in vivo* (Fig. 4A, Table 1). Therefore, to determine whether the outcomes of BC53 and BC67 in the resulting tumors were the consequence of clonal dynamics or their own inherent tumor growth capacity, we orthotopically injected BC53 and BC67 into mice as monoclonal populations. Each of these two clones displayed ~10-day latency and rapid growth kinetics, which were similar to each other and that of the BC Pool (Fig. 4B). Those results suggested that BC53 was particularly susceptible to clonal dynamics *in vivo* and BC67 appeared to be unchanged by *in vivo* selection pressures.

In contrast, BC8 comprised only ~2% of the BC Pool at the time of injection (Table 1) and had a relatively low proliferation rate *in vitro* (Fig. 4A, Fig. S4A). Nevertheless, the relative contribution of BC8 expanded ~10-fold *in vivo* (Fig. 4A, Table 1). When injected as a monoclonal population, BC8 displayed relatively short latency (~10-12 days) and rapid growth kinetics, reflective of the BC Pool tumors (Fig. 4B). Those

results suggested that BC8 experienced positive selective pressure *in vivo*, causing it to become one of the most rapidly proliferating clones.

Some clones appeared to undergo negative selection in the polyclonal setting, either *in vitro* or *in vivo*. For example, BC45, which had a relatively high proliferation rate and represented a dominant clone within the BC Pool at time of injection, represented only <0.1% of the resulting tumors (Fig. 4A; Table 1). Likewise, other clones, such as BC14, BC17, BC21, and BC73, diminished to only <0.0004% of the tumors (Table 1).

One of the most proliferative clones *in vitro*, BC18, underwent a 3-fold reduction in relative contribution to the BC Pool in culture and a 574-fold reduction *in vivo* (Fig. 4A; Table 1). Another clone with a high proliferation rate *in vitro*, BC43, comprised ~4% of the BC Pool at time of injection yet was diminished to <0.2% of the resulting tumors (Fig. 4A, Table 1). Strikingly, those 2 clones, BC18 and BC43, were each capable of forming monoclonal tumors after a latency period of ~40-50 days, and with variable incidence (Fig. 4C). Importantly, those results suggested that the contribution to disease progression of clones with inherent tumorigenic potential after a long latency period (e.g., BC18 and BC43) might have been overlooked because ethical size limitations of the BC Pool tumors dictated end points prior to emergence of those clones from latency.

Collectively, these results suggested that clonal dynamics and/or environmentally driven selection pressures ultimately influence the fate of certain BCs, while the inherent proliferation of other BCs is not impacted in the polyclonal setting. Although these concepts have been previously established^{16,18,43}, SunCatcher enabled us to identify and analyze the specific clones within heterogeneous tumors that are influenced by such dynamic processes *in vitro* and *in vivo*.

SunCatcher Enables Phenotypic Analysis of Positively and Negatively Selected Tumor Cell Clones that Associate with Outcomes

Ostensibly, one of the greatest advantages of the SunCatcher clonal barcoding approach is the ability to obtain useful information from analyses of live cells of interest, regardless of their outcome in any given experiment. To test that concept, we designed experiments to determine whether individual BC phenotypes *in vitro* correlate with their fate *in vivo*, namely, their contribution to the resulting Met1 BC Pool tumors. We therefore evaluated each Met1 BC for phenotypes known to impact tumorigenesis and disease progression⁴⁴. We focused on markers of stratified epithelium (cytokeratin 8 and 14), epithelial-mesenchymal transition (Zeb1 and EpCAM), and immune regulation (MHC-I and PD-L1).

Immunocytochemical analysis of cytokeratin 8 (CK8, “luminal”), cytokeratin 14 (CK14, “basal”), and nuclear Zeb1 staining revealed varying levels of expression among the clones for each factor (Fig. S4A, Table S6). Some BCs expressed a single cytokeratin (e.g., BC67 expressed CK8; BC22 expressed CK14) (Fig. 4D). Other BCs contained cells that stained for either CK8 or CK14 to varying extents (e.g., BC11),

and some BCs (e.g., BC36) were comprised of individual cells that expressed both cytokeratin 8 and 14 (Fig. 4D). Similarly, co-staining for EpCAM and Zeb1 revealed BCs that almost exclusively expressed either nuclear Zeb1 (e.g., BC25) or EpCAM (e.g., BC53) (Fig. 4D). Most BCs (e.g., BC5 and BC42) expressed both factors to varying extents (Fig. 4D). Hence, despite their clonal origin, some BCs gave rise to phenotypically diverse populations.

We also evaluated cell-surface expression of EpCAM, PD-L1, and MHC-I for each BC by flow cytometry. In this case, we quantified baseline levels of PD-L1 and MHC-I in the absence of IFN γ stimulation. BCs were then ranked according to their mean fluorescence intensity for each factor and again, we observed varying levels of expression among the BCs (Fig. S4B, Table S6).

We next asked how the BCs relate to one another with respect to the seven *in vitro* hallmark phenotypes that we analyzed (CK8, CK14, Zeb1, EpCAM, MHC-I, PD-L1, and relative proliferation rate; Fig. S4A, B). To do so, Z-scores were generated for each biological parameter for each BC (Table S7). We then generated a Pearson correlation matrix to visualize the relationships between clones (Fig. 4E, Table S8). The matrix revealed significant heterogeneity among all the BCs as well as clusters of BCs that are significantly correlated with one another with respect to the 7 biological features (Fig. 4E, Table S8). We discovered cases in which phenotypically dissimilar clones shared common fates *in vivo*. For example, BC25 and BC53 are significantly anti-correlated ($r = -0.64$; Fig. 4E, Table S8) yet both were dominant clones in the tumors (Fig. 4A). Additionally, we observed that some clones were highly similar phenotypically, yet their fate diverged *in vivo*. For example, the most highly correlated clones, BC53 and BC73 ($r = 0.95$; Table S8), had different outcomes *in vivo*; BC53 comprised 10.71% of the resulting tumors compared to BC73, which comprised <0.01% of the tumors (Table 1).

We next determined how closely correlated CK8, CK14, Zeb1, EpCAM, MHC-I, PD-L1, and proliferation rate were across all Met1 BCs. We confirmed the expected significant anti-correlations between CK8 and CK14⁴⁵ ($r = -0.61$, $p < 0.022$). Nuclear Zeb1 and cell-surface EpCAM expression were significantly anti-correlated ($r = -0.64$, $p < 0.001$) (Fig. 4F, Table S11). Moreover, PD-L1 and EpCAM expression were highly significantly correlated ($r = 0.87$, $p < 0.001$) (Fig. 4F, Table S11), and cell-surface expression of MHC-I was significantly anti-correlated with CK14 expression ($r = -0.42$, $p < 0.03$) (Fig. 4F, Table S11). None of the phenotypic parameters was significantly correlated with *in vitro* proliferation (Fig. 4F).

To reduce dimensionality and visualize trends in the data, we stratified tumor composition into 3 cohorts (>1%, 0.01-0.99%, <0.01%) and performed principal component analysis (PCA) on the 9 variables that we had measured (CK8, CK14, Zeb1, EpCAM, MHC-I, PD-L1, proliferation rate, *in vitro* fold change in contribution to the BC Pool, and *in vivo* fold change in contribution during tumor progression) (Table S9). PCA revealed separation between those BCs that comprised the highest (>1%) and lowest (<0.01%) proportion of the tumors (Fig. 4G), with two notable outliers: BC25 and BC73 (Fig. 4G). To understand which variables were most influential in the PCA patterns, we visualized the PC loadings (Fig. 4H). That analysis revealed that PC1 is most positively influenced by EpCAM ($r = 0.90$) and PD-L1 ($r = 0.84$) expression and negatively impacted by Zeb1 ($r = -0.69$) and CK14 ($r = -0.60$) expression (Fig. 4H, Table

S10). PC2 is most associated with CK8 ($r = 0.79$) and MHC-I ($r = 0.76$) and is negatively associated with CK14 ($r = -0.63$) (Fig. 4G, Table S10).

Collectively, these results revealed that the ability to analyze individual clones in various contexts and at different time points, yields important information about biological properties that associate with outcomes. Moreover, these findings provided proof-of-concept that the ability to study clones that were eliminated or otherwise not selected during our experimentation yielded important information.

Suncatcher Enables Identification and Quantification of Early Spontaneous Metastasis

Early spontaneous metastases from orthotopic sites are often difficult to detect using conventional experimental detection methods, such as bioluminescence of luciferase-labeled tumor cells, for a variety of reasons⁴⁶. We injected GFP-Luciferase-labeled Met1 parental cells¹⁶ either orthotopically or intravenously into cohorts of mice ($n=9$) and analyzed the lungs ex vivo 21 days later. We did not detect lung metastasis by IVIS imaging or by visual observation in the mice that had been orthotopically injected (Fig. S5A, B). Lungs of mice that had been injected intravenously only had detectable signals in 2/9 mice, although visual observation showed overt metastatic nodules in the majority of the lungs (Fig. S5A, B). Those findings were in line with a recent report demonstrating that immunogenicity of fluorescent proteins used to tag cells often hinders pre-clinical metastasis research in immune competent models⁴⁶.

Given the sensitivity of SunCatcher barcode detection and reproducibility of results, we asked whether we might detect early Met1 BC Pool metastases from orthotopic primary tumor sites using our qPCR-based detection method. We first devised a method that would enable us to extrapolate numbers of metastatic cells based on the total barcode qPCR signal obtained from a fixed mass of tissue (see Methods). Our method also enabled thresholding to distinguish real signal from background signal for each tissue; consequently, we were able to reproducibly detect 1 cell per 0.1 mg of tissue (Fig. 5A).

To evaluate spontaneous metastasis, we injected the Met1 BC Pool into contralateral mammary glands and examined visceral and skeletal tissues 21 days later. Total primary tumor burden was similar between 4/5 mice (mean 1185.5 mg; range 1000-1551 mg), and only one small tumor (256 mg) formed in one mouse (Fig. 5B). Mice were otherwise healthy, and we did not observe any overt metastases upon visual examination of tissues. However, barcode deconvolution by qPCR indicated that the lungs, long bones, and mandibles consistently contained barcode signal (Fig. 5B).

Among the tissues in which barcodes were detected, the leg bones (femur and tibia combined) had the greatest relative numbers of spontaneous metastases, regardless of the primary tumor burden, as 100% of the mice had a barcode signal above the threshold (Fig. 5B). Lung metastasis was detected in 4/5 mice, and metastasis to the mandible was detected in 2/5 mice (Fig. 5C).

BC53 was a dominant clone in 100% of mice that had lung metastasis and the only clone that was represented in more than 50% of lung metastasis (Fig. 5C). In the leg bones (femur and tibia), the inter-individual composition of BC metastases was slightly more variable and diverse (Fig. 5C). For example, unlike in the lung, BC53 was detected in 60% of the mice with metastases to the leg bones (Fig. 5C). The mice with the lowest tumor burden (mouse 7047; 256 mg) and highest tumor burden (mouse 7025; 1551 mg) each had clonal leg bone metastases (BC8 and BC53, respectively; Fig. 5C). BC8 and BC53 were the only barcodes detected in the mandibles, and in that tissue, BC8 was the dominant clone (Fig. 5C).

We were also able to assess intra-individual differences in composition across different metastatic sites. Mouse 7025 displayed relative consistency across all tissues, with BC53 clonal metastasis in the lung and leg bones, and BC8 appearing only in the mandible (Fig. 5C). Interestingly, mouse 7047 had the lowest primary tumor burden yet had BC8 clonal metastasis in the leg bones (Fig. 5C). Those results suggested that BC8 is particularly amenable to bone metastasis.

Collectively, these results indicated that SunCatcher offers a functional barcoding approach to study clonal dynamics during metastatic progression. Importantly, the qPCR-based detection method enabled us to identify and quantify metastases even at very early stages from orthotopic sites.

Discussion

We developed the SunCatcher clonal barcoding method to enable functional analysis of the subclones that comprise complex, heterogeneous populations of cells. With SunCatcher, each clone is specific to a single barcode; there is no chance that a clone contains more than one unique barcode sequence. Therefore, a potential complication of traditional barcoding methods, which rely on infecting heterogeneous populations of cells with libraries of 10^6 - 10^8 barcodes without the ability to isolate and study individual clones, is avoided. While the generation of individual clones is initially labor-intensive, once the NBC and BC stocks are established, they can be resourced indefinitely. A significant benefit of the qPCR-based BC detection method is that it is rapid, sensitive, reliable, and inexpensive. The advantage of the NGS-based deconvolution method is that it allows for multiplexing if larger barcode libraries are desired.

We found that pools of just ~30 heterogeneous clones were sufficient to recapitulate tumor growth kinetics of the respective parental population. Moreover, a limited analysis revealed some of the phenotypes that associated with BC representation—both positive and negative selection—in the resulting tumors. The SunCatcher system should easily enable unbiased transcriptional profiling or more specialized functional assays to further characterize BCs. Nevertheless, whether we captured the entirety of the heterogeneity represented in the original parental population is not clear. Finding ways to estimate the number of clones required to represent the heterogeneity of a complex parental cell line could be an important consideration when applying SunCatcher to other studies.

Our analyses of BC cell phenotypes exposed additional unique and novel benefits of the SunCatcher system. First, our clonal approach enables one to trace clonally related cells, even as the inherent plasticity of those cells gives rise to phenotypic heterogeneity. For example, we found that some BCs were comprised of both CK8+ and CK14+ cells. A common experimental approach is to sort cells based on an expression marker of interest (e.g., CK8 and CK14) and then perform functional analyses on the sorted populations to identify differences in their behavior or responses to treatment. Without using molecularly tagged cells in such an experiment, the clonal relationship between some of the CK8+ and CK14+ cells would not be revealed. Second, because stocks of individual BCs are retained, SunCatcher uniquely enables one to not only identify but also isolate the common ancestor of cells that might otherwise appear to be unrelated by other analysis methods. Third, unlike SunCatcher, most traditional barcoding approaches, in which a heterogeneous cell population is infected with a barcode library, do not enable retrospective isolation and analysis of the clone of origin of cells with relevant phenotypes (e.g., tumor formation, metastatic progression, or drug responses).

Using SunCatcher, we identified clones that constituted a very minor proportion of the BC Pool tumors at the experimental end point yet had the potential to form rapidly growing tumors after a long latency period when injected as a pure clonal population (e.g., BC18 and BC43). Pre-clinical end points are dictated by maximum humane tumor size (1cm in our case) and so analysis of the BC Pool tumor composition could not extend beyond that end point. Our findings therefore raise questions about the fate of those minor clones if our experimental time points were extended or if those clones represented residual disease after surgical resection or drug treatment. In fact, a report investigating melanoma stem cells demonstrated that all patient-derived tumor cells had tumor-initiating potential if given sufficient time (very protracted in some cases) following injection⁴⁷. Moreover, several studies provide evidence that metastases are derived from subclones present at low frequency in the primary tumor^{48–50}. SunCatcher, or clonal approaches like it, could thus offer functional analyses of minor subclones and aid in the identification of therapeutic targets that could prevent their progression.

In our experience, we were previously unable to detect early metastases in our pre-clinical models using conventional detection methods. Moreover, we had assumed that our Met1 cells were unable to spontaneously metastasize to distant organs before primary tumors reached ethical end point. Our detection of spontaneous metastases in lungs and skeletal tissues after only 3 weeks are in line with numerous studies that have detected circulating and disseminated tumor cells in the early stages of cancer progression in both preclinical models and in patients^{51–54}. We envision that SunCatcher could provide valuable information about the properties of individual disseminated tumor cells that make them more or less threatening than their counterparts.

A current limitation to SunCatcher is that we cannot directly isolate discrete live BC cells from the BC Pool tumors. The ability to do so could ostensibly reveal important information about BC gene expression and phenotypic plasticity during disease progression. This limitation may be overcome by incorporating BC identification into single cell analysis technologies. Nevertheless, we found that one of the more important and novel benefits of SunCatcher is the ability to identify and functionally analyze clones that

are negatively selected during experimentation. That ability could be especially important for evaluating drug responses and identifying critical cell vulnerabilities. The ability to study “super responders” could lead to development of therapeutic approaches that prevent disease progression.

Methods

Cell lines

The Met1 TNBC cell line, which was originally generated from a spontaneously arising tumor in an FVB/N-Tg (MMTV-PyVmT) mouse³⁸, was a gift from J. Joyce (University of Lausanne) with permission from A. Borowsky (UC Davis School of Medicine) and maintained in DMEM containing 10% fetal bovine serum, 100 U/mL penicillin-streptomycin, and 4 mM glutamine, as previously described¹⁶. The 4T1 TNBC cell line, which was derived from a spontaneously arising tumor in a Balb/c mouse³⁷, was provided by F. Miller (Wayne State University School of Medicine) and maintained in DMEM containing 10% fetal bovine serum, 1x MEM non-essential amino acids solution (Gibco), 100 U/mL penicillin-streptomycin, and 4 mM glutamine. The McNeuA HER2+ cell line, which was originally derived from a spontaneously arising breast tumor in an MMTV-neu transgenic mouse, was provided by Michael Campbell (University of California, San Francisco) and maintained in DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS) and 100 U/mL penicillin-streptomycin (Gibco) at 37°C with 5% CO₂³⁶. Human HMLER-hygro-H-*ras*V12 (HMLER-HR) cells^{39, 55} were derived from human mammary epithelial cells (HMEC) originally obtained from ATCC. were cultured in advanced DMEM/F12 (Gibco) supplemented with 5% calf serum (CS, HyClone), 0.1% hydrocortisone (Sigma Aldrich), and 100 U/mL penicillin-streptomycin (Gibco) at 37°C with 5% CO₂. 293T cells were cultured in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 100 U/mL penicillin-streptomycin (Gibco) at 37C with 5% CO₂. All cell lines were tested and were negative for mycoplasma (Lonza) and short tandem repeat analysis verified the identity of mouse (Bioassay Methods Group, National Institute of Standards and Technology) and human (Promega GenePrint 10 System) cells.

Lentiviral barcode vectors

Barcode sequences were originally developed by Tm Bioscience (Toronto, Ontario, Canada)⁵⁶ and are described in the Luminex FlexMAP Microspheres Product Information Sheet⁵⁷, as previously described by Peck et al.³⁵. Lentiviral vectors containing barcode sequences were generated as previously described^{28, 35}. 293T cells were transfected with 2.5 µg lentiviral vector barcode DNA, 2.5 µg pCMV-dR8.2 dvpr (Addgene plasmid 8455), 1 µg pCMV-VSVG (Addgene plasmid 8454) in Opti-MEM Reduced Serum Medium (Gibco) with Fugene HD (Promega) or FuGENE6 (Roche Corporation). Viral supernatants were collected 48 and 72 hours after transfection, filtered through a sterile 0.45-µm syringe filter (VWR), and stored at -80°C.

SunCatcher barcoding protocol

Single cells were isolated from cell lines by either FACS or plating 0.5 cells/well in 96-well plates. We verified that each well contained a single cell by phase microscopy under 40x magnification and discarded wells that contained more than one cell. Each single-cell clone was expanded, and these cells were designated non-barcoded clones (NBCs). Aliquots of 5×10^5 cells of each NBC were prepared in appropriate culture medium with 10% DMSO and stored in cryovials in liquid nitrogen.

For barcoding, each NBC population was thawed into a 6-well plate and infected with a unique barcode-containing lentivirus supernatant in appropriate medium with Polybrene (Sigma Aldrich) at a concentration ranging from 1:200 to 1:50. After 24 hours, virus was washed out and replaced with fresh medium with 10 $\mu\text{g}/\text{ml}$ blasticidin (Sigma). After 3 days, cell numbers were counted in each condition and those with an infection rate of $<10\%$ (number of infected cells remaining in selection media / infected cells grown in regular growth media) were selected. Single cells were then isolated from the infected populations by plating 0.5 cells/well in 96-well plates and then expanded in order to generate multiple individual clonal populations with each barcode. A single clonal population was selected at random for each unique barcode, which was referred to as the barcoded clonal population (BC). Aliquots of 5×10^5 cells of each BC were prepared in an appropriate culture medium with 10% DMSO and stored in cryovials in liquid nitrogen.

For each cell line, equal numbers of every BC were mixed to form a BC Pool. Aliquots of 5×10^5 cells of each BC Pool were prepared in appropriate culture medium with 10% DMSO and stored in cryovials in liquid nitrogen.

Genomic DNA preparation from cells and tissues

Cultured cells were detached using 0.25% trypsin, pelleted and resuspended in Buffer AL (Qiagen). Genomic DNA (gDNA) was extracted using a QiaAmp DNA Mini Kit (Qiagen) according to manufacturer's instructions. Tumors and other tissues were flash frozen in liquid nitrogen and pulverized using a mortar and pestle. 25 mg of tissue was used for DNA isolation according to the manufacturer's protocol (Qiagen). Femurs and mandibles were immediately dissected and cleaned to remove muscle and connective tissue and were then flash frozen and pulverized (preparations included marrow). 25 mg of bone preparations were used for DNA isolation according to the manufacturer's protocol (Qiagen). The DNA concentration was quantified using a NanoDrop 8000 Spectrophotometer (Thermo Fisher).

qPCR-based barcode identification and analysis

To detect barcodes, gDNA was first pre-amplified by preparing 50 µl reactions containing 500 ng gDNA, OneTaq 2x MasterMix (New England Biolabs), and 10 µM of each F/R pre-amplification primer to common flanking sequences (Forward: 5'-CGATTAGTGAACGGATCTCG-3'; Reverse: 5'-CCGGTGGATGTGGAATGTG-3') (Table S2). The following PCR program was used to amplify the template DNA: 94°C for 30 seconds, followed by 40 cycles of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds; a final extension was performed at 68°C for 5 minutes. The PCR products were purified by a Monarch® PCR & DNA Cleanup Kit following the manufacturer's instructions. Purified PCR products were eluted in 50 µl water, and the DNA concentration was determined using a NanoDrop 8000 Spectrophotometer (Thermo Fisher). To quantify the abundance of specific BCs in any given sample, the purified pre-amplification PCR products were analyzed by qPCR, for which each barcode sequence³⁵ was used as the forward primer and 5'-CCACTTGTGTAGCGCCAAG-3' was used as the universal reverse primer (Table S2). Each sample was tested for all barcodes by setting up an individual reaction for each barcode primer set for each sample. Each qPCR mixture contained 0.001 ng of purified PCR product, 1 mM of each F/R primer, and 1X iTaq™ Universal SYBR Green Supermix (Bio-Rad) in a 10-µL reaction volume. The following program was used: 50°C for 2 minutes and 94°C for 10 minutes, followed by 40 cycles of 94°C for 30 seconds and 60°C for 30 seconds. Dissociation curves were collected after qPCR, and quality control of qPCR signals was performed to ensure that there was a single peak, thus indicating that amplification of a single barcode occurred in each reaction. The Ct values acquired from the qPCR run were converted to arbitrary units (AU, $AU = 20000000000 \times \text{EXP}(-0.69 \times \text{Ct value})$), and the values of the individual barcodes in every sample were summed. The percentage of each barcode present in the sample was calculated by dividing the individual barcode signal by the sum of all the barcode signals.

Next-generation sequencing-based barcode identification and analysis

From gDNA preparations, the barcode region of the lentiviral barcode vector was amplified using either primer set JO primer F/R (F: 5'-TGGAGCATGCGCTTTAGCAG; R: ATCGTTTCAGACCCACCTCC-3') or indexed primer sets (F: JH p05: 5'-ATC GTT TCA GAC CCA CCT CCC-3'; R: JH p11-14 and JH p31-46, Table S3). PCR mixtures contained 250 ng DNA, OneTaq 2x MasterMix (New England Biolabs), and 10 µM each of F/R primers. The following PCR program was used to amplify the template DNA: 94°C for 30 seconds, followed by 30 cycles of 94°C for 20 seconds, 54°C for 30 seconds, and 68°C for 20 seconds; a final extension was performed at 68°C for 5 minutes. PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's protocol, and purified DNA was eluted in 20-40 µl water.

Ligation-based Illumina library preparation was carried out by the Harvard Biopolymers Facility using an Integex Apollo 324 PrepX ILM kit using Kappa reagents according to manufacturers' protocols. For PCR-based Illumina library preparation, samples that had been amplified using one of the barcode-indexed primer sets were subjected to a second PCR using an Illumina-indexed primer set with regions that were complementary to the barcode lentiviral vector regions (F: JO p50, R: JO p65-p88; Table S4). PCR

mixtures contained 100 ng DNA, OneTaq 2x MasterMix (New England Biolabs), and 10 μ M each of F/R primers. The following PCR program was used to amplify the template DNA: 94°C for 30 seconds, followed by 15 cycles of 94°C for 20 seconds, 54°C for 30 seconds, and 68°C for 20 seconds; a final extension was performed at 68°C for 5 minutes. PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's protocol, and purified DNA was eluted in 20-40 μ l water.

Sequencing Data Analysis

Barcode-specific PCR products with Illumina adaptors were mixed in equimolar ratios and sequenced on an Illumina MiSeq at a depth of 20-30x10⁶ reads. First, reads that did not contain the barcode adaptor sequence ((8 base index)-(17 bases)-ACGCGT-(24 base barcode)-CTGCAG) were filtered out. Samples were grouped based on the 8-base index, and a read count table was generated based on all 24-bp barcode sequences present in each read. To identify barcodes, each barcode on the read-count table was compared to the 24-base barcode sequences in the barcode library. A receiver operating characteristic (ROC) curve was generating by classifying the 24-bp barcodes as "True" (present in the barcode library) or "False" (absent from the library). Thresholding was performed based on the read count that maximized the area under the curve (AUC) from the ROC curve.

Animal experiments

All animal studies were conducted in accordance with regulations of the Institutional Animal Care and Use Committee of the Brigham and Women's Hospital (protocol no. 2017N000056). Female FVB/NJ and BALB/cJ mice were purchased from Jackson Laboratory. Mice were 8-9 weeks of age at the time of injection. For orthotopic injections, 2.5x10⁵ Met1 or Met1BC cells or 1.5x10⁵ 4T1 or 4T1 BC cells were prepared in 20 μ l PBS and injected into the inguinal mammary fat pads of FVB/NH or BALB/cJ mice, respectively. Tumors were measured with calipers 2-3 times per week, and tumor volume was calculated as .

For spontaneous metastasis studies, 2.5x10⁵ Met1 BC Pool cells were prepared in 20 μ l PBS and injected into two contralateral inguinal mammary fat pads of FVB mice. Mouse weight was monitored daily, and animals were sacrificed if their weight decreased by 20%. Otherwise, mice were sacrificed 21 days after injection by CO₂ inhalation and perfused with PBS. Tissues were collected for gDNA preparation and barcode analysis as described above.

Cell proliferation assay

Met1 parental or barcoded clones were plated at a density of 20,000 cells/well in a 6-well plate in DMEM containing 10% fetal bovine serum, 100 U/mL penicillin-streptomycin, and 4 mM glutamine. Every 72 hours over a 9-day time course, cells were counted using a hemocytometer and replated at a density of 20,000 cells/well. Relative proliferation rates of each BC are represented as cell doublings per day.

Immunofluorescence and image analysis

Glass coverslips (#1.5; Election microscopy science) were coated with rat tail collagen I (Thermo-Fisher Scientific) overnight at room temperature. The coverslips were washed once with PBS, and cells were seeded and spread overnight at 37°C and 5% CO₂. The cells were then fixed in 4% paraformaldehyde (Sigma), permeabilized in 1% TX-100 (Sigma) and blocked in 3% bovine serum albumin in PBS (Fisher Scientific). Cells were stained with the indicated primary antibodies for overnight in 4°C, followed by a 1-hour incubation with secondary antibodies at room temperature. Then nuclei were visualized with DAPI (1 µg/ml, Sigma) and mounted on glass slides. Stains and primary antibodies: F-actin (rhodamine phalloidin, 1:1000; Thermo Fisher catalog: R415), CK14 (1:200; Biolegend catalog: 905303), and CK8 (1:10; TROMA-1, DSHB), anti-Zeb1 (1:100; Santa Cruz Biotechnology catalog: sc-25388). Secondary Alexa Fluor antibodies (Invitrogen, 1:250): anti-rat 647 (catalog: A21472), anti-rabbit 488 (catalog: A21206), anti-rat 488 (catalog: A21208), anti-rabbit 647 (catalog: A21244). Slides were imaged using a Nikon Eclipse Ni microscope. Numbers of CK8+ and CK14+ cells were counted by two independent researchers, represented as a percentage of total cells and then stratified into 4 categories whereby 1= 0-5%, 2= 6-30%, 3= 31-60% and 4=61-100% from 5 randomly selected images for each BC. Zeb1 quantification was performed using ImageJ by setting the threshold on the nuclear Zeb1 signal and on the DAPI (nucleus) stained image. The percentage of nuclear Zeb1+ cells was calculated as the number of nuclear Zeb1+ cells/field per total number of cells/field (DAPI).

Flow cytometry

10⁵ cells were harvested using 0.25% trypsin, centrifuged, and seeded into round bottom 96-well plates. The cells were washed with FACS buffer (PBS with 2% FBS) and centrifuged at 300 x g for 5 minutes at 4°C. Cells were incubated with anti-EpCAM (Clone G8.8, APC-Cy7, 1:400 dilution, BioLegend: 118218), anti-MHC-I (clone: KH114, FITC, 1:400 dilution, BioLegend: 115104), and PD-L1 (clone: 10F.9G2, PE-Cy7, 1:100 dilution, BioLegend: 124314) on ice for 30 minutes. The cells were washed twice with FACS buffer and analyzed on a BD Canto II flow cytometer. DAPI (0.1 µg/ml, Sigma) was used as a dead cell marker. Data were analyzed using FlowJo software.

Barcode Association Matrix

Each barcoded clone was assessed for 7 biological features via flow cytometry (EpCAM, MHC-I and PD-L1), Immunofluorescence (Zeb1, CK8, and CK14) and cell proliferation. Z-scores were generated for each quantified biological feature and assigned to each barcoded clone using the formula:

$$Z = \frac{(Value_{clone} - Mean_{all\ clones})}{Standard\ Deviation_{all\ clones}}$$

As an example, flow cytometry z-scores were calculated for each barcode using the mean fluorescence intensity (MFI) of each clone ($Value_{clone}$), the average MFI of all clones ($Mean_{all\ clones}$), and the standard deviation of the MFI for all clones ($Standard\ Deviation_{all\ clones}$). A Pearson correlation matrix was generated by calculating the correlation of each barcode with every other barcode based on the 7 calculated z-scores. Similarly, a Pearson correlation matrix was generated by calculating the correlation of each biological feature with every other biological feature using the z-scores from the barcoded clones. Hierarchical clustering was performed using one minus Pearson's correlation.

Metastasis detection in the lung and bone

The total metastatic burden in various tissues was calculated by analyzing gDNA preparations as follows. The barcode region was pre-amplified from gDNA using the forward primer 5'-CGATTAGTGAACGGATCTCG-3' and the reverse primer 5'-CCGGTGGATGTGGAATGTG-3'. PCR mixtures contained 100 ng DNA, OneTaq 1x MasterMix (New England Biolabs), and 10 µM each of F/R primers. The following PCR program was used to amplify the template DNA: 94°C for 30 seconds, followed by 15 cycles of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds; a final extension was performed at 68°C for 5 minutes. The PCR products were purified with a Monarch® PCR & DNA Cleanup Kit following manufacturer's protocol. Purified PCR products were eluted in 17 µl water. Purified PCR products were used as input for qPCR to quantify all barcodes. For each 10-µl qPCR reaction, 4 µl of the purified PCR product was added to 10 µM primers and 2 µM probe (forward: 5'-TACCGGTTAGTAATGAC-3'; reverse: 5'-TAAAGCGCATGCTCCAG-3'; probe: 5'-FAM-AAAAGCGCCTCCCCTACCCGGTAGGTA-3'-Eclipse) and 5 µl TaqMan™ Fast Advanced Master Mix (Applied Biosystems™). The following PCR program was used: 50°C for 2 minutes and 95°C for 2 minutes, followed by 40 cycles of 95°C for 3 seconds and 60°C for 30 seconds.

Calibration curve to estimate metastatic burden

To extrapolate the metastatic burden based on the total barcode qPCR signal obtained from different types of tissues, we first generated calibration curves by spiking known amounts of gDNA isolated from barcoded cells into gDNA isolated from a fixed amount (0.1 gram) of a given tissue from a tumor-free FVB mouse. Based on the estimated amount of gDNA (6 pg) per diploid cell,⁵⁸ we calculated the estimated barcoded cell number spiked into the gDNA isolated from 0.1 g of indicated tissue. The serially diluted samples were then subjected to PCR-based barcode detection as described above (including pre-amplification, PCR product purification, and qPCR detection using primers and probes with TaqMan Fast Advanced Master Mix). Calibration curves were generated by linear regression analysis of the log-transformed estimated input of barcoded cells and log-transformed qPCR signals. The background signal was defined as the qPCR signal of the normal tissue without barcoded tumor cell gDNA. Samples with a signal higher than background were defined as positive for metastasis, and the metastatic burden was extrapolated from the calibration curve.

Statistical analysis

The data are represented as the mean $\pm$ SEM, unless otherwise indicated. Data were analyzed by two-way ANOVA with the Sidak multiple comparison correction, one-way ANOVA with Tukey's multiple comparison test for significance, or an unpaired two-tailed t test with Welch's correction, as indicated in the figure legends. All data were analyzed using GraphPad Prism 7.0. $P < 0.05$ was considered statistically significant.

Data availability

Data are available upon reasonable request.

Declarations

ACKNOWLEDGMENTS

Lentiviral barcode vectors were a generous gift from Dr. Todd R. Golub. We thank Lisa Situ and Drs. Channing Yu, Yuanbo Qin, and Molly DeCristo for technical assistance. This work was supported by the NIH/NHLBI T32 Hematology Training Grant Fellowship to M.S.; Landry Cancer Biology Research Fellowship to A.M.; and DOD/CDMRP/BCRP W81XWH-14-1-0191 Era of Hope Scholar Award, NIH/NCI R01CA166284 Presidential Early Career Award for Scientists and Engineers, and AACR Gertrude B. Elion Cancer Research Award to S.S.M.

AUTHOR CONTRIBUTIONS

S.S.M. provided the study objective. Q.G., M.S., A.M. and G.J.G. designed and optimized the qPCR barcode detection system in primary tumor and metastasis under the supervision of S.S.M. A.B. provided technical support in conducting experiments for in vitro and in vivo models. J.F.O. designed SunCatcher barcoding protocol and optimized the NGS barcode detection system under the supervision of S.S.M. S.S.M. composed the manuscript. All authors critically reviewed and approved the final manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

References

1. Wu, A.M., Till, J.E., Siminovitch, L. & McCulloch, E.A. Cytological evidence for a relationship between normal hematopoietic colony-forming cells and cells of the lymphoid system. *J Exp Med* **127**, 455–464 (1968).
2. Harrison, D.E., Astle, C.M. & Lerner, C. Number and continuous proliferative pattern of transplanted primitive immunohematopoietic stem cells. *Proc Natl Acad Sci U S A* **85**, 822–826 (1988).
3. Capel, B., Hawley, R., Covarrubias, L., Hawley, T. & Mintz, B. Clonal contributions of small numbers of retrovirally marked hematopoietic stem cells engrafted in unirradiated neonatal W/W^v mice. *Proc Natl Acad Sci U S A* **86**, 4564–4568 (1989).
4. Marusyk, A. & Polyak, K. Cancer. Cancer cell phenotypes, in fifty shades of grey. *Science* **339**, 528–529 (2013).
5. Meacham, C.E. & Morrison, S.J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
6. Dagogo-Jack, I. & Shaw, A.T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81–94 (2018).
7. Kreso, A. et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* **339**, 543–548 (2013).
8. Chaffer, C.L. et al. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* **154**, 61–74 (2013).
9. Castano, Z. et al. IL-1β inflammatory response driven by primary breast cancer prevents metastasis-initiating cell colonization. *Nat Cell Biol* **20**, 1084–1097 (2018).
10. Campbell, N.R. et al. Cooperation between melanoma cell states promotes metastasis through heterotypic cluster formation. *Dev Cell* (2021).
11. Navin, N.E. & Hicks, J. Tracing the tumor lineage. *Mol Oncol* **4**, 267–283 (2010).

12. van Galen, P. et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281 e1224 (2019).
13. Acosta, J., Ssozi, D. & van Galen, P. Single-Cell RNA Sequencing to Disentangle the Blood System. *Arterioscler Thromb Vasc Biol* **41**, 1012–1018 (2021).
14. Xu, J. et al. Using single-cell sequencing technology to detect circulating tumor cells in solid tumors. *Mol Cancer* **20**, 104 (2021).
15. Wu, S.Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334–1347 (2021).
16. Olive, J.F. et al. Accounting for tumor heterogeneity when using CRISPR-Cas9 for cancer progression and drug sensitivity studies. *PLoS One* **13**, e0198790 (2018).
17. Chaffer, C.L. et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc Natl Acad Sci U S A* **108**, 7950–7955 (2011).
18. Janiszewska, M. et al. Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. *Nat Cell Biol* **21**, 879–888 (2019).
19. Gerrits, A. et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010).
20. Lu, R., Neff, N.F., Quake, S.R. & Weissman, I.L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol* **29**, 928–933 (2011).
21. Schepers, K. et al. Dissecting T cell lineage relationships by cellular barcoding. *J Exp Med* **205**, 2309–2318 (2008).
22. Nolan-Stevaux, O. et al. Measurement of Cancer Cell Growth Heterogeneity through Lentiviral Barcoding Identifies Clonal Dominance as a Characteristic of In Vivo Tumor Engraftment. *PLoS One* **8**, e67316 (2013).
23. Nguyen, L.V. et al. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat Commun* **5**, 5871 (2014).
24. Nguyen, L.V. et al. Clonal analysis via barcoding reveals diverse growth and differentiation of transplanted mouse and human mammary stem cells. *Cell Stem Cell* **14**, 253–263 (2014).
25. Porter, S.N., Baker, L.C., Mittelman, D. & Porteus, M.H. Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo. *Genome Biol* **15**, R75 (2014).
26. Bhang, H.E. et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21**, 440–448 (2015).
27. Wagenblast, E. et al. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature* **520**, 358–362 (2015).
28. Yu, C. et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat Biotechnol* (2016).

29. Mathis, R.A., Sokol, E.S. & Gupta, P.B. Cancer cells exhibit clonal diversity in phenotypic plasticity. *Open Biol* **7** (2017).
30. Feldman, D. et al. CloneSifter: enrichment of rare clones from heterogeneous cell populations. *BMC Biol* **18**, 177 (2020).
31. Jin, X. et al. A metastasis map of human cancer cell lines. *Nature* **588**, 331–336 (2020).
32. Oren, Y. et al. Cycling cancer persister cells arise from lineages with distinct programs. *Nature* (2021).
33. Berthelet, J. et al. The site of breast cancer metastases dictates their clonal composition and reversible transcriptomic profile. *Sci Adv* **7** (2021).
34. Gutierrez, C. et al. Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. *Nature Cancer* **2**, 758–772 (2021).
35. Peck, D. et al. A method for high-throughput gene expression signature analysis. *Genome Biol* **7**, R61 (2006).
36. Campbell, M.J., Wollish, W.S., Lobo, M. & Esserman, L.J. Epithelial and fibroblast cell lines derived from a spontaneous mammary carcinoma in a MMTV/neu transgenic mouse. *In Vitro Cell Dev Biol Anim* **38**, 326–333 (2002).
37. Miller, F.R. Tumor subpopulation interactions in metastasis. *Invasion Metastasis* **3**, 234–242 (1983).
38. Borowsky, A.D. et al. Syngeneic mouse mammary carcinoma cell lines: two closely related cell lines with divergent metastatic behavior. *Clin Exp Metastasis* **22**, 47–59 (2005).
39. Elenbaas, B. et al. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* **15**, 50–65 (2001).
40. Wong, K.H., Jin, Y. & Moqtaderi, Z. Multiplex Illumina sequencing using DNA barcoding. *Curr Protoc Mol Biol* **Chapter 7**, Unit 7 11 (2013).
41. Mitra, A., Skrzypczak, M., Ginalski, K. & Rowicka, M. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform. *PLoS One* **10**, e0120520 (2015).
42. Aleckovic, M., McAllister, S.S. & Polyak, K. Metastasis as a systemic disease: molecular insights and clinical implications. *Biochim Biophys Acta Rev Cancer* **1872**, 89–102 (2019).
43. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* **12**, 323–334 (2012).
44. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
45. Dontu, G. & Ince, T.A. Of mice and women: a comparative tissue biology perspective of breast stem cells and differentiation. *J Mammary Gland Biol Neoplasia* **20**, 51–62 (2015).
46. Grzelak, C.A. et al. Elimination of fluorescent protein immunogenicity permits modeling of metastasis in immune-competent settings. *Cancer Cell* (2021).
47. Quintana, E. et al. Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. *Cancer Cell* **18**, 510–523 (2010).

48. Ding, L. et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
49. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892 (2012).
50. Lindstrom, L.S. et al. Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol* **30**, 2601–2608 (2012).
51. Luzzi, K.J. et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* **153**, 865–873 (1998).
52. Pantel, K., Brakenhoff, R.H. & Brandt, B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. *Nat Rev Cancer* **8**, 329–340 (2008).
53. Hosseini, H. et al. Early dissemination seeds metastasis in breast cancer. *Nature* (2016).
54. Lucci, A. et al. Circulating tumour cells in non-metastatic breast cancer: a prospective study. *Lancet Oncol* **13**, 688–695 (2012).
55. McAllister, S.S. et al. Systemic endocrine instigation of indolent tumor growth requires osteopontin. *Cell* **133**, 994–1005 (2008).
56. Bioscience, T. <http://www.tmbioscience.com>.
57. Luminex <http://www.luminexcorp.com>.
58. Gillooly, J.F., Hein, A. & Damiani, R. Nuclear DNA Content Varies with Cell Size across Human Cell Types. *Cold Spring Harb Perspect Biol* **7**, a019091 (2015).

Tables

Table 1 is not available with this version.

Figures

Fig. 1: SunCatcher Clonal Barcoding and Functional Analyses.

Figure 1

(a) Schematic of the SunCatcher Clonal Barcoding Method. SunCatcher utilizes two rounds of single cell cloning to ensure that each subclone has only 1 unique barcode and that each cell within that subclone contains the same barcode insertion site. Due to the single cell cloning approach, custom BC pools of any combination can be designed. (FACS, fluorescence-activated cell sorting; NBC, non-barcoded clone; BPP, barcoded polyclonal population (polyclonal for barcode insertion site); BC, barcoded clone; BC Pool, population containing multiple BCs). (b) BCs or BC Pools (input) can be entered into any experiment and detected at end point (output) using various deconvolution approaches. Deconvolution by qPCR provides a cost-effective, rapid, and highly sensitive method for detecting and quantifying BCs. SunCatcher enables retrieval of all clones, including those negatively selected during experimentation, for further analyses and/or design and testing of custom BC Pools.

Fig. 2: BC Detection by qPCR and Next-Generation Sequencing-Based Methods.

Figure 2

(a) Design of pre-amplification primers (red arrows) and BC-specific primers (black arrows) for qPCR detection. (b) Heatmap of qPCR C_T values when testing each indicated barcode oligonucleotide primer against every Met1 BC population as well as the Met1-BC pool. (c) Multi-level multiplexing for NGS is

achieved by attaching 1 of 20 unique barcode-indexes to the 3' end of each sample. Up to 20 indexed samples can be pooled into a single barcode-index library. Each barcode-index library can be ligated into 1 of 24 uniquely indexed Illumina adaptors. Purple region represents PstI, light blue represents MluI. **(d)** Sequencing read counts for 2 examples of test samples using the Illumina library preparation method. Each library corresponds to a single Illumina adaptor, and the expected barcode pair for each library is indicated.

Fig. 3: SunCatcher Approach Maintains Tumorigenic Properties and Enables Analysis of Heterogeneity in vitro and in vivo.

Figure 3

(a) Phase contrast microscopic images of indicated Met1 BCs; 20x magnification. **(b)** Sand plot showing clonal composition (cumulative percentage) of Met1 BC Pool over 2 passages (7 days) *in vitro*. **(c)** Growth kinetics of Met1 Parental (black; $Y = 0.3360x + 0.1699$) and Met1 BC Pool (grey; $Y = 0.3562x + 0.4121$) over 8 days in culture; $n=4$ per group; error bars = SD. **(d)** Growth of tumors from Met1 parental cells and Met1 BC Pool cells in FVB mice; $n=5$ per cohort, error bars represent S.E.M. **(e, f)** Mass (g) of

tumors (e) and spleens (f) from mice in experiment represented in (d). **(g)** Quantitative PCR assessment of barcode composition in the Met1 BC Pool at time of injection (left), in each of 10 tumors (n=5 mice) after 18 days (center), and calculation of the average composition of all tumors (right). Composition bars show indicated barcodes as a percent of total barcode signal (100%) within each sample; tumor mass (mg) is indicated above each bar. Color code for each BC is indicated. **(h, i)** Average representation of each BC in n=10 tumors for each of 2 experiments. BCs that constituted >0.5% of total BC signal are represented in (h) and <0.5% are shown in (i). The correlation coefficients (R^2 , simple linear regression) are shown; error bars represent S.E.M.

Fig. 4: SunCatcher Enables Functional and Phenotypic Analysis of Individual Clones and Assessment of Clonal Dynamics

Figure 4

(a) Heatmap representing functional analyses of Met1 BCs in monoculture and within the BC Pool. Column 1 shows proliferation rate (doubling time) of each individual BC over a 9-day period in culture as ranked from highest to lowest. Each row represents data for the BC indicated in column 1. Clonal dynamics in vitro (column 2) are represented as the fold change in BC composition of the Met1 BC Pool after 7 days in culture relative to initial admixture. Clonal dynamics in vivo (column 3) are represented as

the fold change in BC composition of tumors at experimental end point relative to the Met1 BC Pool composition at the time of injection. The average contribution of each BC as a percent of total BCs in each resulting tumor (n=10) is shown in column 4. Data values for columns 2-4 are shown in Table 1. **(b)** Growth of tumors derived from the Met1 BC Pool (n=4) and indicated individual BCs (n=5 per cohort). These BCs represent the 3 most dominant clones in the resulting BC Pool tumors at experimental end point (associated with Fig. 4A and Table 1). **(c)** Growth of tumors derived from indicated BCs (n=6 per cohort). These BCs represent 2 of the clones that comprised <0.3% of the Met1 BC Pool tumors at the 18-day experimental endpoint (associated with Fig. 4A and Table 1). **(d)** Left panels: Immunofluorescence images of indicated BCs stained for cytokeratin 8 (CK8, green), cytokeratin 14 (CK14, red), f-actin (white). Right panels: Immunofluorescence images of indicated BCs stained for epithelial cell adhesion molecule (EpCAM, green), Zeb1 transcription factor (red), f-actin (white). All nuclei were counterstained with DAPI (blue); 20x magnification. **(e)** Pearson correlation matrix for individual Met1 BCs analyzed for 7 *in vitro* phenotypes: proliferation rate, CK8, CK14, Zeb1, EpCAM, PD-L1, MHC-I (associated with Fig. S4 and Tables S6-S8). **(f)** Pearson correlation matrix for the 7 phenotypes assessed *in vitro*: proliferation rate, CK8, CK14, Zeb1, EpCAM, PD-L1, MHC-I (associated with Table S10). *p<0.022, 1-tailed, 90% CI; **p<0.030, ***p<0.001, 2-tailed, 90% CI. **(g)** Principal component scores from analysis of BCs based on 9 parameters (proliferation, in vitro dynamics, in vivo dynamics, CK8, CK14, Zeb1, EpCAM, PD-L1, MHC-I) stratified by indicated percent composition of Met1 BC Pool tumors at experimental end point (associated with Tables S9, S10). **(h)** Loadings plot from PCA shown in 4f.

Fig. 5: SunCatcher Barcoding to Study Early Metastasis from Orthotopic Sites.

Figure 5

(a) Calibration curves were generated for indicated tissues by serially diluting known amounts of barcoded tumor cell gDNA into a fixed amount of normal tissue gDNA. From left to right: lung, long bones (from femur and tibia), mandible. **(b)** 2.5×10^5 barcoded Met1 tumor cells were injected bilaterally into the mammary fat pads (n=5 animals) and tumors were allowed to grow for 21 days, at which point tissues were harvested and metastasis burden was calculated. Dashed lines indicate the background signals

from each indicated tissue type. Tissues with signal above the background were considered as positive for metastasis and estimated tumor cell number per 0.1mg tissue was calculated based on the calibration curve for that specific tissue. **(c)** Barcode composition analysis on tissues with positive metastasis signal. Bars represent percent of total barcode signal (100%) within each sample. Also shown are mouse identities, total primary tumor burden for each animal, and estimated numbers of metastases per tissue; N.D., not detected.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementalmaterials.pdf](#)