SUPPLEMENTARY INFORMATION

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

Qiuchen Guo[1,2,*], Milos Spasic[1,2,*], Adam Maynard[2,*], Gregory J. Goreczny[1,2,*], Amanuel Bizuayehu[1], Jessica F. Olive[1,2,], and Sandra S. McAllister[1-4]

[1]Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[3]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142, USA
[4]Harvard Stem Cell Institute, Cambridge, Massachusetts, 02138, USA

Corresponding Author:
Sandra S. McAllister
Brigham & Women's Hospital
Harvard Institutes of Medicine 742
4 Blackfan Circle
Boston, MA 02115 USA
Phone: 617-525-4929
smcallister1@bwh.harvard.edu

* These authors contributed equally

SUPPLEMENTARY METHODS

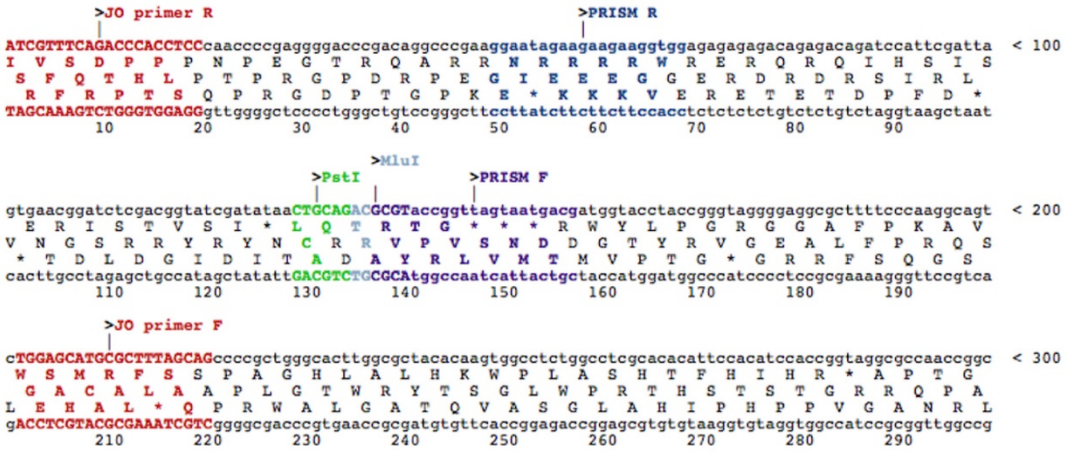
Detection of Lung Metastases Using Bioluminescence Imaging

To analyze metastasis from orthotopic sites, $2.5x10^5$ GFP-Luciferase-labeled parental Met1 l cells (GPF-Luc-Met1) were prepared in 20 µl sterile PBS and injected into the inguinal mammary fat pads. For intravenous injections, $7.5x10^5$ GPF-Luc-Met1 cells were prepared in 100 µl sterile PBS and injected into the tail vein. 21 days after tumor cell injections, mice were euthanized, lungs were dissected and placed into wells, and covered with 300 µg/mL D-luciferin (Perkin-Elmer). Luciferase-positive cells were detected using a Xenogen IVIS imaging system (Caliper Life Sciences). Luminescent signal was detected as radiance (p/sec/cm2/sr) and analyzed using the Living Image Software Version 4.1 (Caliper Life Sciences).

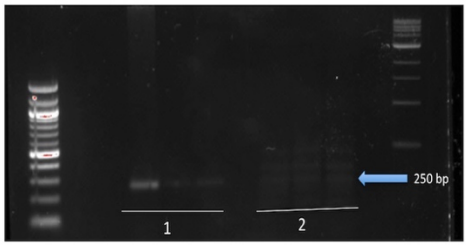
SUPPLEMENTARY FIGURES

Supplemental Figure 1. Barcode Detection by Sanger Sequencing

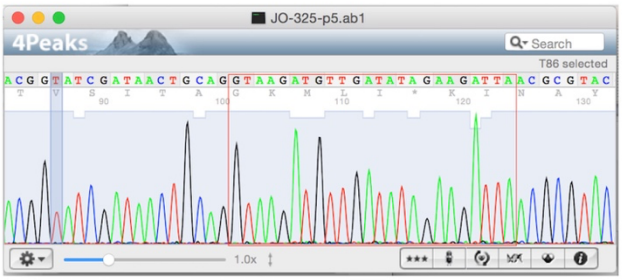
a



b



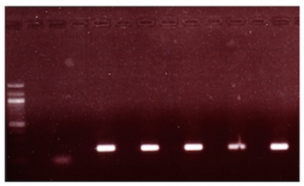
c



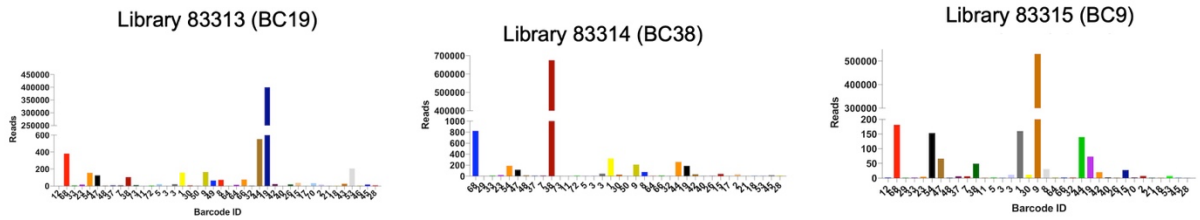
(a) Primers (JO primer F and JO primer R) were designed upstream and downstream of the barcode region (barcodes were inserted between PstI/MluI restriction sites) to use for PCR amplification prior to Sanger sequencing. **(b)** Agarose gel comparing PCR product yields from PCR reactions using gDNA input extracted using a Qiagen kit (designated as 1; lanes contained reactions with DNA input of 500 ng, 250 ng, and 100 ng, respectively) or the PRISM gDNA extraction method (designated as 2; lanes contained reactions with DNA input of 500 ng, 250 ng, and 100 ng, respectively). Expected amplicon size is ~250 bp (blue arrow). **(c)** Representative chromatogram from Sanger sequencing results of a Luminex barcode amplicon (barcode region is boxed in red).

Supplemental Figure 2. BCs are Accurately Detected by NGS

a

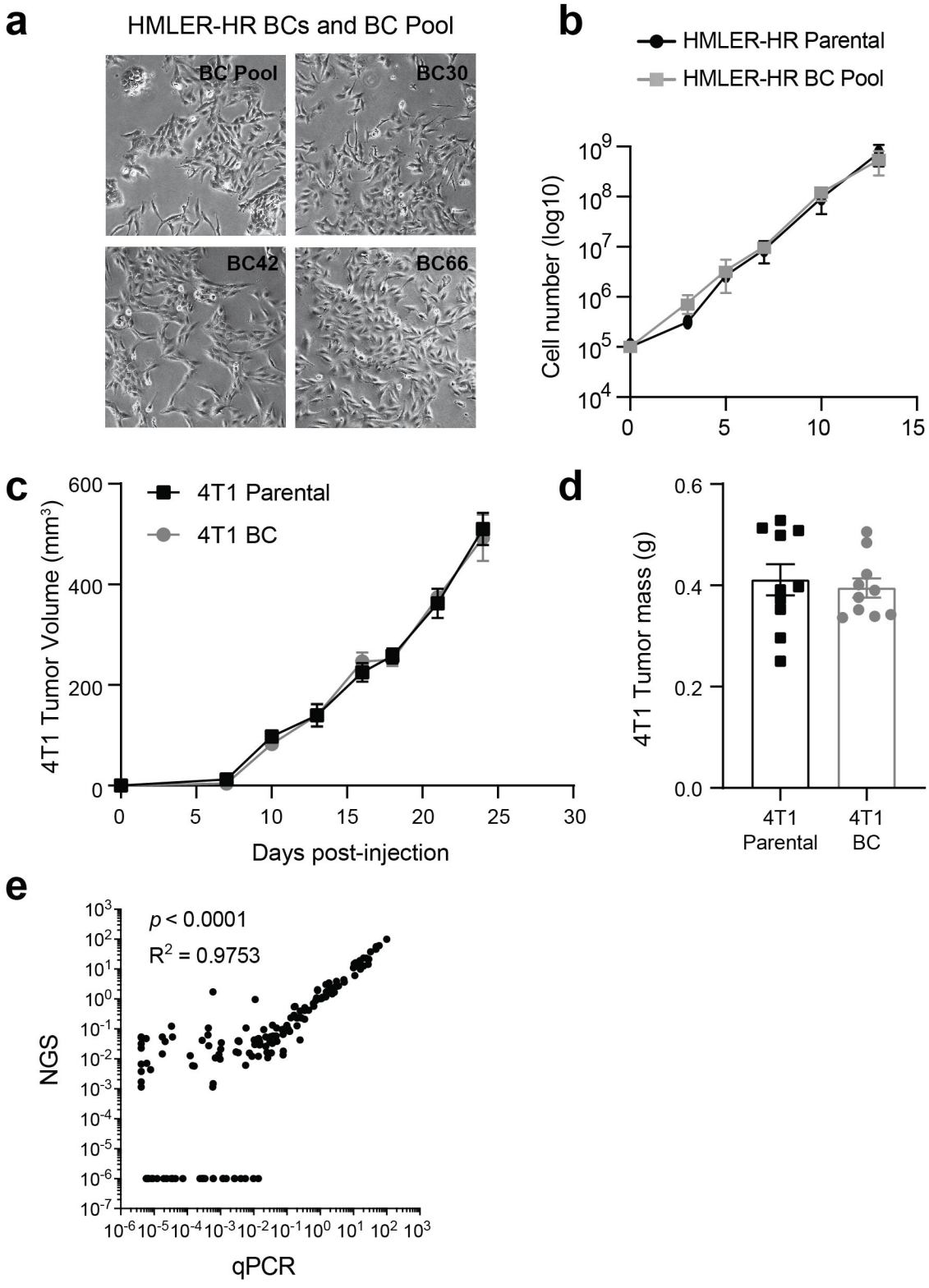


b



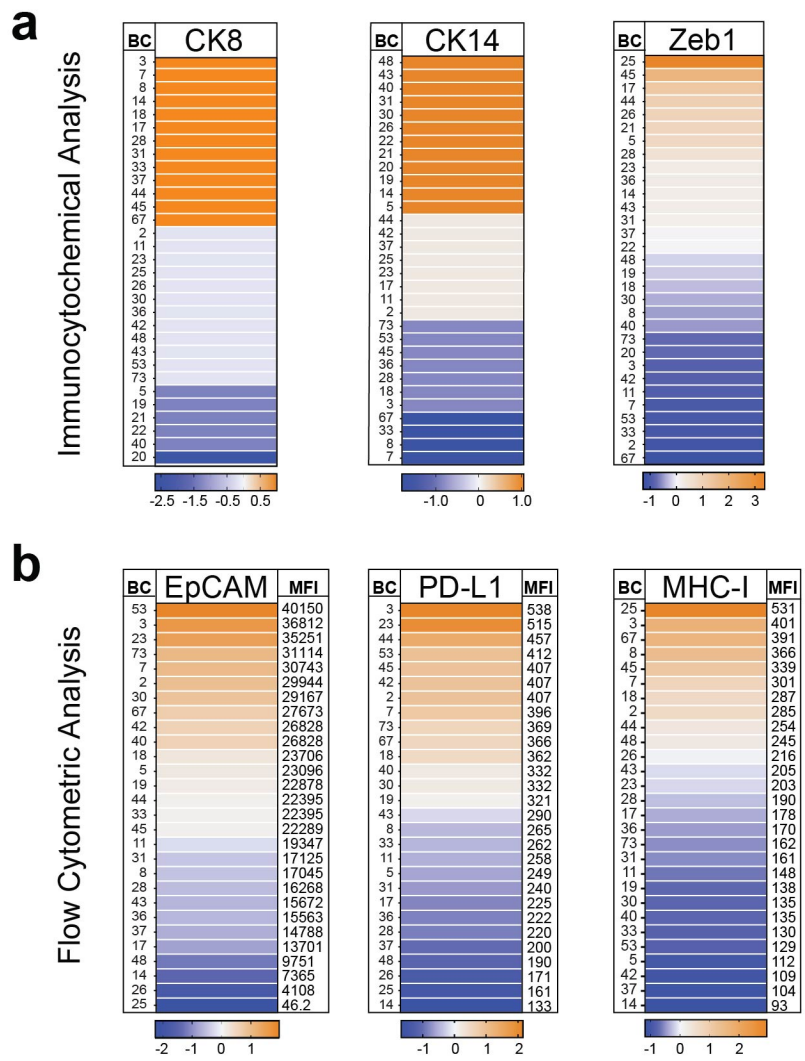
(**a**) Agarose gel of PCR products following attachment of Illumina adaptors to barcode-index amplicons (Lane 1: 100 bp ladder; Lane 2: no DNA control; Lanes 3-7: PCR products). (**b**) Sequencing read counts for the 3 test samples using the PCR-based Illumina library preparation method. Each library corresponds to a single Illumina adaptor, and the expected barcode pair for each library is indicated.

Supplemental Figure 3. Growth Properties of Barcoded Human and Mouse Cancer Cell Lines and Comparision of Deconvolution Methods



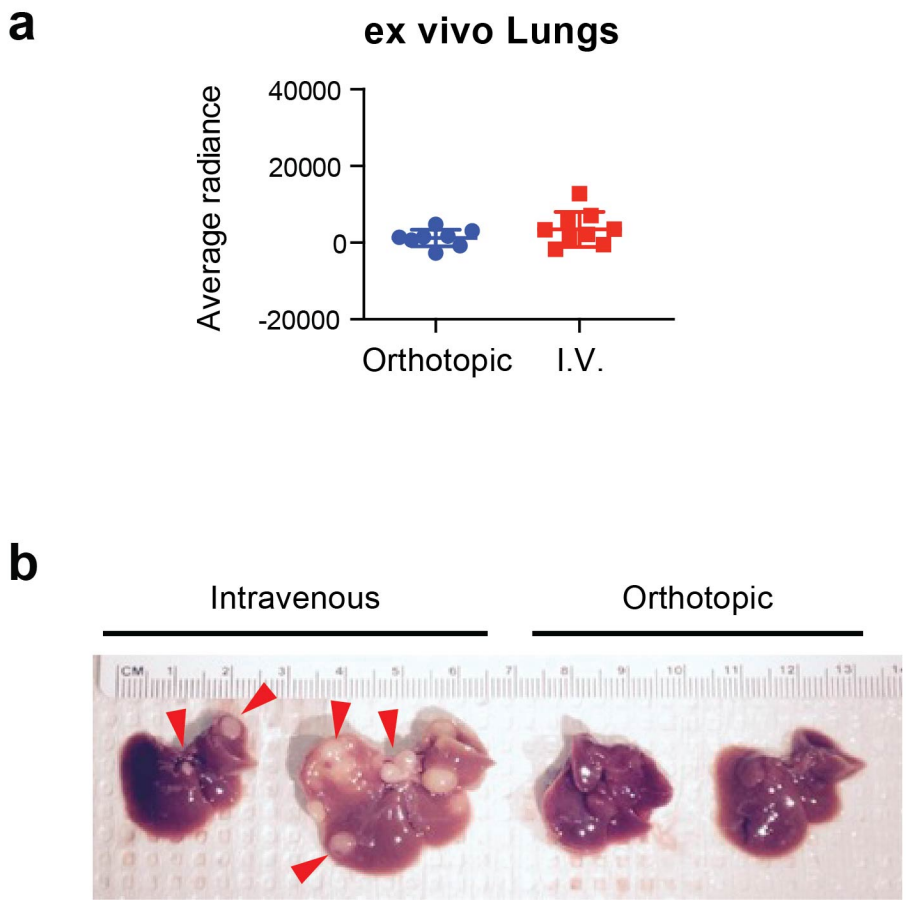
(a) Cell morphology of HMLER-HR barcode cells and the barcode pool. (**b**) Growth kinetics of HMLER-HR Parental (black) and HMLER-HR BC Pool (grey) over 12 days in culture; n=3 replicates per group; error bars = SD. **(c)** Growth of tumors from 4T1 parental cells (black) and 4T1 BC Pool cells (grey) in Balb/C mice (n=10 per cohort); error bars represent S.D. **(d)** Mass (g) of 4T1 from (C) at 25-day experimental end point. (**e**) Genomic DNA was isolated from 6 tumors derived from the McNeu BC Pool after ~5 weeks of tumor growth *in vivo*; each barcode was quantified as a percentage of total barcodes in a given tumor by both qPCR and next-generation sequencing. Each data point represents the value for an individual barcode from an individual tumor by each method. Linear regression was used to calculate the data correlation between NGS and qPCR methods.

Supplemental Figure 4. BC Phenotypic Analyses



(a) Heatmaps showing ranked relative expression levels of cytokeratin 8 (CK8), cytokeratin 14 (CK14), and nuclear Zeb 1 as quantified by immunofluorescence image analysis for each BC (**b**) Heatmaps showing ranked relative expression of EpCAM, PD-L1, and MHC-I as quantified by flow cytometry for each BC. Median fluorescence intensity (MFI) is indicated for each BC.

Supplemental Figure 5. IVIS imaging of GFP-Luc Met1 Lung Metastases



(a) GFP-Luciferase-labeled Met1 parental cells either orthotopically or intravenously (I.V.) into cohorts of mice (n=9) and analyzed the lungs ex vivo 21 days later using IVIS Lumina III imaging system (Perkin Elmer). (**b**) Representative images of lungs from intravenous or orthotopic injection mouse models.

Supplementary Table 1. Collection of Clonally Barcoded Cell Lines

Supplementary Table 2. Oligonucleotide Primer Sequences for qPCR-based BC Detection

Supplementary Table 3. Barcode-index reverse (R) Primer Sequences for NGS-Based BC Detection.

Supplementary Table 4. Illumina Library Preparation Primer Sets.

Supplementary Table 5. Composition of Met1 BC Pool at Start and End of Injections. BC composition of the Met1 BC Pool was calculated at the initiation of injections ("pre-injection") and again approximately 2 hr later, when the final injection was performed ("post-injection"). The BC Pool cells were kept in suspension on ice during the injection time period.

Supplementary Table 6. Raw Data for Seven Biological Features of Individual Met1 BCs.

Supplementary Table 7. Z Scores for Seven Biological Features of Individual Met1 BCs (from data in Table S6).

Supplementary Table 8. Pearson correlation Matrix (from data in Table S7)

Supplementary Table 9. Multiple Variable Z Scores Used for Principal Component Analysis.

Supplementary Table 10. Multiple Variable Principal Component Analysis (from data in Table S9).

Supplementary Table 11. Pearson Correlation Matrix of 7 Biological Features