AWARD NUMBER: W81XWH-18-1-0420

TITLE: Targeting Resistance in Colorectal Cancer with a Novel Lineage-Tracking Technology

PRINCIPAL INVESTIGATOR: Amy Brock

CONTRACTING ORGANIZATION: The University of Texas at Austin

REPORT DATE: November 2022

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOVEMBER 2022	FINAL	1 Aug 2018 - 31 July 2022

4. TITLE AND SUBTITLE
Targeting Resistance in Colorectal Cancer with a Novel Lineage-Tracking Technology

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-18-1-0420

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Amy Brock, Ph.D.

E-Mail: amy.brock@utexas.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Texas at Austin
110 Inner Campus Dr, MN 13
Austin, TX 78712-1139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This proposal addresses the FY17 PRCRP Topic Area of Colorectal Cancer and the Military Relevance Focus Area "Gaps in cancer prognosis, treatment and/or survivorship" that impact the health of military service members, veterans, and their beneficiaries. Treatment of colorectal cancer is complicated by the high degree of intratumoral heterogeneity. Recent studies have demonstrated the utility of DNA barcode libraries in monitoring heterogeneous cell populations. This is achieved by labeling each cell with a unique, heritable sequence; lineage abundance is tracked over time by sequencing the barcode ensemble. Changes in clonal dynamics after perturbations, such as treatment with a pharmacological agent, reveal variation in lineage survival or growth rate. This approach enables the simultaneous observation of many lineage trajectories to reveal high-resolution details of cancer population dynamics. With this tool, the ability to concurrently track clonal fitness dynamics and generate lineage-specific genomic and transcriptomic data over longitudinal studies provides new insight into the response to therapeutics.

15. SUBJECT TERMS
Colorectal carcinoma, tumor heterogeneity, chemotherapy, chemoresistance, clonal dynamics, lineage tracing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	35	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

	Page
1. Introduction	4
2. Keywords	4
3. Accomplishments	4
4. Impact	7
5. Changes/Problems	8
6. Products	9
7. Participants & Other Collaborating Organizations	10
8. Special Reporting Requirements	11
9. Appendices	11

1. INTRODUCTION:

This proposal addresses the FY17 PRCRP Topic Area of Colorectal Cancer and the Military Relevance Focus Area "Gaps in cancer prognosis, treatment and/or survivorship" that impact the health of military service members, veterans, and their beneficiaries. Treatment of colorectal cancer is complicated by the high degree of intratumoral heterogeneity. Recent studies have demonstrated the utility of DNA barcode libraries in monitoring heterogeneous cell populations. This is achieved by labeling each cell with a unique, heritable sequence; lineage abundance is tracked over time by sequencing the barcode ensemble. Changes in clonal dynamics after perturbations, such as treatment with a pharmacological agent, reveal variation in lineage survival or growth rate. This approach enables the simultaneous observation of many lineage trajectories to reveal high-resolution details of cancer population dynamics. With this tool, the ability to concurrently track clonal fitness dynamics and generate lineage-specific genomic and transcriptomic data over longitudinal studies provides new insight into the response to therapeutics.

2. KEYWORDS:

Colorectal carcinoma, tumor heterogeneity, chemotherapy, chemoresistance, clonal dynamics, lineage tracing

3. ACCOMPLISHMENTS:

What were the major goals of the project?

These are described below, broken down by Specific Aim, Major Task, and Subtask from the approved Statement of Work (SOW). There are two Specific Aims of the project, as described in the approved SOW.

Aim 1: To utilize a novel cell lineage tracking system BAAR to measure cell lineage dynamics and isolate specific lineages of interest in cultures of patient-derived colorectal carcinoma.

Aim 2: To demonstrate the utility of BAAR for the identification of tumor subpopulations that vary in chemosensitivity and resistance.

Specific Aim 1: To utilize a novel cell lineage tracking system *BAAR* to measure cell lineage dynamics and isolate specific lineages of interest in cultures of patient-derived colorectal carcinoma cells.

Major Task 1: To validate the BAASE lineage tracking and isolation in patient-derived colorectal carcinoma cells (mutant KRAS)

Subtask 1: Complete local IRB review has been completed and submit all documents have been submitted for HRPO review and approval. No work will commence on colorectal cells until HRPO review and approval is complete.

Result: This Subtask was reported in the Year 1 report.

Subtask 2: 3 CRC cells lines will be obtained from Drs. Eckhardt and Van Den Berg at the Developmental Therapeutics Lab and maintained as organo-cultures. These organo-cultures will be labeled with reference barcode libraries.

Result: This Subtask was reported in the Year 2 report.

Subtask 3: Recall plasmids for the reference barcode sets will be constructed and confirmed by Sanger sequencing.

Result: This Subtask was completed in the Year 2 report.

Subtask 4: Transfection parameters will be optimized for Recall plasmid delivery.

Result: This Subtask was reported in the Year 2 report.

Subtask 5: Activation of a GFP reporter gene by the Recall plasmid will be measured by flow cytometry. The percent of GFP+ cells will be compared to the known composition of the reference populations to measure the false positive/false negative rates of lineage activation.

Result: This Subtask was reported in the Year 2 report.

Major Task 2: Label CRC cell with high diversity barcode library

Subtask 1: An additional set of 3 cell lines will be obtained from the Dev. Therapeutics Lab for organoculture

Result: This task was reported in Year 2.

Subtask 2: All cell lines from MT1 and MT2 ST1 will be labeled with a high diversity BAASE barcode library, using the parameter conditions optimized in Major Task 1.

Result: This task was reported in Year 2.

Subtask 3: Activation of Recall plasmid GFP expression will be assessed by flow cytometry. Transfection conditions, dCas9-VPR ratios, and timing of isolation from Major Task 1 will be confirmed in these 6 cell lines.

Result: Transfection was performed using using 1.5µl Lipofectamine™3000, 1µl P3000™ Reagent, 150ng of Recall GFP plasmid and 250ng of dCas9-VPR plasmid. ells were analyzed for GFP expression via flow cytometry 48 h post-transfection. Error load was quantified by comparatively tallying the values of the GFP-histogram, from high to low GFP intensity, of matching and mismatching recall samples. Specifically, sum totals of matching recall events were tabulated with respect to and along with each new accruing mismatch recall event. Transfection rates were consistent across samples and varied from 19.1- 29.6

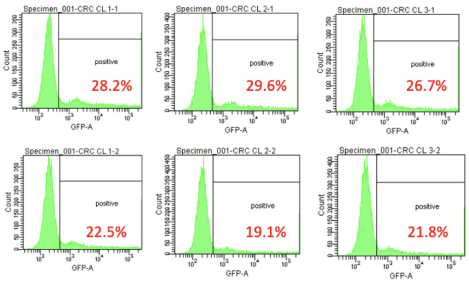


Figure 1. Optimization of barcode labeling in patient-derived CRC.

Specific Aim 2: To demonstrate the utility of *BAAR* for the identification of tumor subpopulations that vary in chemoresponsiveness and resistance.

Major Task 1: To isolate lineages resistant to irinotecan, oxaliplatin and 5-FU

Subtask 1: The 3 cell lines in MT1 will be used to establish baseline dose response curves. IC50 will be measured for CRC organo-cultures treated with irinotecan, oxaliplatin and 5-FU alone and in combination.

Results: PDO cultures were seeded according to standard Developmental Therapeutics Lab protocols. Briefly, 5000 cells are embedded in Matrigel droplets (5-10 uL) per well of a 96 well plate and cultured in DMEM/F12 supplemented with N-acetyl-l-cysteine, WNT3a, R-spondin-1, Noggin, EGF and FGF. Cultures were maintained for 4 days before dosing in triplicate. After 48 hours in the presence of drug, viability was measured by Cell Titer Glo. Mean IC50 were 4.04 uM oxaliplatin, 66.8 uM 5-fluorouracil, and 88 uM irinotecan.

Subtask 2: Barcode labeled organo-cultures of all 6 cell lines will be treated with irinotecan, oxaliplatin and 5-FU, alone and in combination. A control population will be monitored without any drug treatment (vehicle only).

Results: These cell treatments were performed and the analysis comprises Subtask 3 in this section.

Subtask 3: For each of 6 cell populations, barcode ensembles will be quantified by NGS and between 2 and 3 high abundance resistant lineages will be identified.

Results: Cells were harvested from PDO cultures post-treatment and genomic DNA was extracted using the PureLink® Genomic DNA Mini Kit (Thermo Fisher). Barcode sequences were amplified using PCR, extracted by Ampure bead cleaup and prepared for SE 150 bp sequencing on an Illumina NovaSeq platform by the UT Austin Genomics and Sequencing Analysis Facility (GSAF). Primer sequences contained both flanking barcode annealing regions and Illumina adaptor/index sequence. Functional gRNA units were identified with a custom Python script and gRNA sequences grouped using the ustacks program provided as part of the stacks package. Barcodes with 1 mismatch (per 20 bp barcode sequence) were pooled, corresponding to an allowed sequencing error rate of 5%. The distribution of barcode abundance is shown in Figures 2 and 3 below.

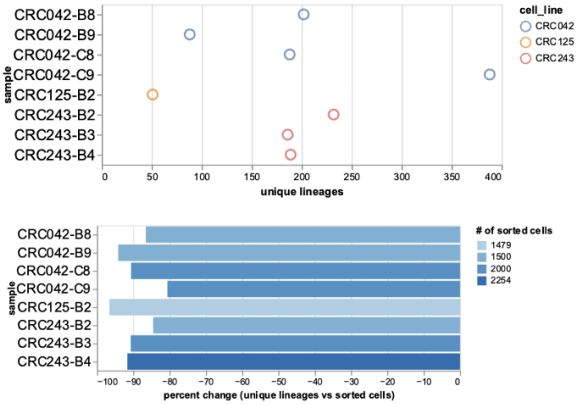


Figure 2. Quantitation of barcode diversity in patient-derived CRC populations. **A)** Targeted sequencing of genomic DNA isolated from biological replicate cells libraries of each patient-derived CRC. All samples were expanded to at least 5 million cells before gDNA isolation. **B)** Population CRC125 displayed reduced cell proliferation rates, potentially due to low capture of initial sorted cells.

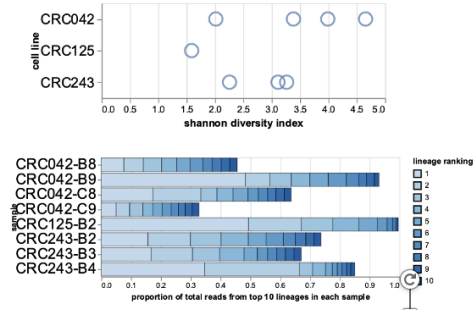


Figure 3. Quantitation of Shannon index diversity in three patient-derived CRC populations. **A)** CRC042 barcoded cell libraries display the largest range in Shannon diversity compared to CRC243 and CRC125; **B)** Quantitation of the 10 most abundant lineages; in all samples except CRC042-C9 these represent greater than 50% of the library composition.

Subtask 4: Recall plasmids will be constructed to isolate high abundance resistant lineages.

Results: Using methods previously established in the lab, Recall plasmids were constructed by type IIS restriction cloning to introduce barcode specific landing pad sequences upstream of a GFP reporter and were verified by Sanger sequencing.

Subtask 5: Resistant lineages will be isolated by GFP activation and FACS.

Results: To isolate specific lineages of interest, we then transfected the heterogeneous cell population with Recall plasmids to drive lineage-specific green fluorescent protein (GFP) expression and enable collection of purified lineages by FACS. Here we found that although this protocol works well for conventional adherent cell culture models (immortalized cell lines and also primary cells), there were significant challenges in collecting sufficient cells of a single lineage from the small cell numbers associated with organoid culture. In addition to low capture of rare lineages, we found that shifting the culture from 3D to dispersed cells for FACS resulted in significant cell death.

To address this challenge we first attempted to pool multiple replicate samples, however cell viability was still less than 20%. We then shifted to a strategy of plating the PDO cells in adherent 2D culture conditions for the GFP activation step and maintaining in these culture conditions until lineage isolation. With this approach we achieved the recovery and expansion of cells surviving each drug treatment, and were able to harvest 1 million cells per sample for subsequent characterization. However, barcode sequencing of the control population revealed that lineage diversity was significantly reduced by the change in culture conditions (Figure 4). This confounding effect resulted in control cultures which no longer maintained the rare lineages that contributed to the survivor cell populations. On the one hand, this provides support for the utility of PDO models in capturing the rich diversity of phenotypes within a clinical sample. However, we conclude there remains a technical challenge in lineage isolation from this highly heterogeneous background and this will need to be tackled in future refinements to the approach.

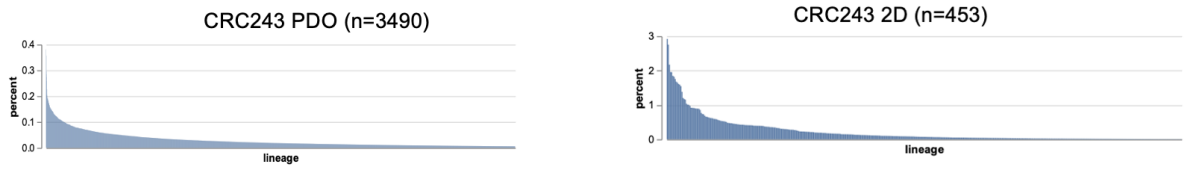


Figure 4. Targeted NGS of genomic barcode ensembles reveals a significantly loss of lineage diversity upon a shift to adherent culture conditions. (*n,* number of unique barcodes)

Major Task 2: To characterize resistant lineages by RNASeq and determine their sensitivity to other therapies.

Subtask 1: Transcriptomes of isolated resistant lineages will be measured by RNASeq to identify altered signaling pathways that may be targeted by other therapies. The CTEP database will be examined for matching drug targets.

Results: Despite the troubleshooting detailed above in MT 1, Subtask 5 we were not able to isolate sufficient quantities of the lineages identified from PDO cultures to enable downstream RNA-Seq analysis. We conclude that lineage-resolved transcriptomic analysis of PDO may require additional optimization or technological refinements to the ClonMapper system.

Subtask 2: Isolated lineages will be subcultured and LD50 will be measured to the other drugs and combinations in this proposal (irinotecan, oxaliplatin and 5-FU

Results: Although we were not able to isolate individual lineages for subculturing and downstream analysis, we were able to compare LD50 of the heterogeneous cell populations culture in PDO and in adherent 2D culture (Figure 5). We conclude that the shift in culture conditions not only results in selection of specific lineages but also alters the drug sensitivity phenotype of the cell population. This aspect warrants further follow-up investigation, as it will be critical to understand how the selection and maintenance of patient models impacts the distribution of molecular phenotypes and treatment profiles in vitro.

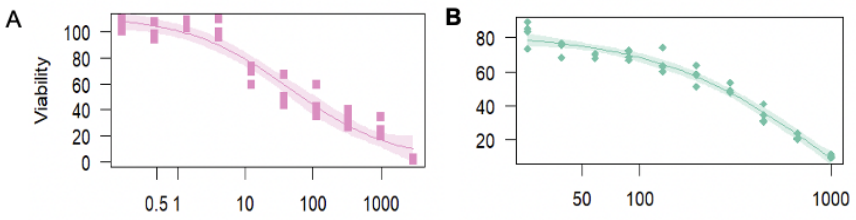


Figure 5. Shift in LD50 is associated with a change in culture conditions. CRC cultured as PDO (A) and 2D adherent culture (B) demonstrate a shift in sensitivity to irinotecan from LD50= 88 uM to LD50= 550 uM.

What opportunities for training and professional development has the project provided?

Graduate students supported by this grant participate in ongoing training and professional development activities. Within the department, trainees participate in weekly lab group meetings to share technical skills and gain practical experience in critical evaluation of the literature, oral communication and experiment design.

At the department level, students participate in two biweekly seminar series—one on Cancer Biology and one focused on Computational Oncology. A student-led troubleshooting session provides an additional setting for peer mentoring in computational work and statistical analyses. This session includes research scientists and postdocs, as well as graduate students, and is highly interactive.

At the college/university level, trainees have participated in two short courses through The University of Texas at Austin Genomics and Sequencing Analysis Core and the Texas Advanced Computing Center to improve technical skills in bioinformatics and computational biology. In combination with foundational course work, these classes have enabled students to gain new skills in informatics; this has proved useful for analysis of the NGS data in this project.

How were the results disseminated to communities of interest?

The PI has presented invited seminars and conference podium talks to educate clinicians and researchers about the utility of the tools developed in this work. Audiences have included dana Farber Cancer Research Institute/Harvard Medical School, Moffitt Cancer Center, University of North Carolina at Chapel Hill, Texas A&M University, the American Association of Cancer Research Evolutionary Dynamics, among others. In addition, the team has participated in local symposia and workshops to interact with the graduate student and postdoc trainees of the University of Texas.

What do you plan to do during the next reporting period to accomplish the goals?

This is the final report.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

This project has accomplished the important step of demonstrating that this barcoding technology can be utilized in patient-derived tumor cells. This is a key step, as our earlier work has used only established, immortalized cell lines which have been maintained in laboratories for many years and thus have some biological differences from cancer cells that are treated in the clinic. This proof-of-technology has opened up feasibility for other collaborators to also use this system in patient-derived tumor cells and we will continue to make these tools available for wide sharing.

What was the impact on other disciplines?

Nothing to report.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

Nothing to report.

Actual or anticipated problems or delays and actions or plans to resolve them

In year 1, we experienced a significant delay in HRPO review of our documents (supporting documents uploaded via eBRAP Aug 2018, supporting documents submitted by email Dec 27, 2018; HRPO clarification requested by email May 20, 2019; final decision received by email July 15, 2019). HRPO agreed with the previous recommendation of our local institutional review board and this research was deemed exempt.

Like most research groups worldwide, we experienced significant COVID-19 related delays due to illness among lab and university personnel and supply chain issues. We were able to work through many of these challenges.

Changes that had a significant impact on expenditures

Nothing to report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nothing to report

Nothing to report

Significant changes in use of biohazards and/or select agents

Nothing to report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Howland K., Brock A. Cellular barcoding tracks heterogeneous clones through selective pressures and phenotypic transitions. *Trends in Cancer*. (2023, *in review)*.

Books or other non-periodical, one-time publications.

Kaitlyn Johnson, Eric Brenner, Amy Brock. Implications of non-genetic heterogeneity in cancer drug resistance and malignant progression. Phenotypic Switching, Academic Press, 2020,

Other publications, conference papers and presentations.

- Eric A. Brenner, Daylin Morgan, Aziz Al'Khafaji, Catherine Gutierrez, Catherine J. Wu, Amy Brock; Abstract PO-097: High resolution analysis of clonal dynamics using lineage tracing and single cell transcriptomics. *Cancer Res* 1 November 2020; 80 (21_Supplement): PO–097.
- Amy Brock; Abstract IA004: Tracking population heterogeneity and chemoresistance with functionalized cell barcodes. *Cancer Res* 15 May 2022; 82 (10_Supplement): IA004.

- **Website(s) or other Internet site(s)**

Nothing to report

- **Technologies or techniques**

Nothing to report

- **Inventions, patent applications, and/or licenses**

Nothing to report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Amy Brock, Ph.D.
Project Role:	PI
Researcher Identifier (e.g. ORCID ID):	amybrock (eRA Commons)
Nearest person month worked:	1 month/year
Contribution to Project:	Dr. Brock oversees all aspects of the project, coordinates and plans studies, contributes to data analysis and interpretation and communication.
Funding Support:	n/a

Name:	Carla Van Den Berg, PharmD
Project Role:	Co-I
Researcher Identifier (e.g. ORCID ID):	vandenbc (eRA Commons)
Nearest person month worked:	1 month/year
Contribution to Project:	Dr. Van Den Berg…
Funding Support:	n/a

Name:	Gail Eckhardt, MD
Project Role:	Co-I
Researcher Identifier (e.g. ORCID ID):	Eckhardt.Gail (eRA Commons)
Nearest person month worked:	1 month/year
Contribution to Project:	Dr. Eckhardt…
Funding Support:	n/a

Name:	Daylin Morgan
Project Role:	Graduate Student
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	6 months/year
Contribution to Project:	Mr. Morgan established the plasmids for the study and optimized transfection conditions of PDO cells.
Funding Support:	n/a

Name:	Russell Durrett
Project Role:	Graduate Student
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2 months, year 2
Contribution to Project:	Mr. Durrett assisted in design and construction of barcode libraries and performed analysis of library diversity.
Funding Support:	n/a

Name:	Eric Brenner
Project Role:	Graduate Student
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2 months/ years 1-2
Contribution to Project:	Mr. Brenner developed bioinformatic analyses for clonally resolved single cell measurements.
Funding Support:	n/a

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to report.

What other organizations were involved as partners?

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: N/A

QUAD CHARTS: N/A

9. APPENDICES: Please find attached manuscript, accepted at *Trends in Cancer*.

Cellular barcoding tracks heterogeneous clones through selective pressures and phenotypic transitions

Kennedy K. Howland[1] and Amy Brock[1]

[1]Department of Biomedical Engineering, The University of Texas at Austin, Austin TX 78734

Corresponding author email: amy.brock@utexas.edu

Abstract:

Genomic DNA barcoding has emerged as a sensitive and flexible tool to measure the fates of clonal sub-populations within a heterogeneous cancer cell population. Coupling cellular barcoding with single cell transcriptomics permits the longitudinal analysis of molecular mechanisms with detailed clone-level resolution. Numerous recent studies have employed these tools to track clonal cell states in cancer progression and treatment response. With these new technologies comes the opportunity to examine longstanding questions about the origins and contributions of tumor cell heterogeneity and the roles of selection and phenotypic plasticity in disease progression and treatment.

Keywords: barcoding, tumor heterogeneity, tumor evolution, drug resistance, phenotypic plasticity,

Introduction

Heterogeneity is a fundamental property of all cell populations [1] and results from the stochasticity of molecular interactions [2]–[5]. In cancer, as in normal cell development, the inherent phenotypic diversity in a cell population is a substrate for the forces of selection [6]–[8]. Cellular interactions and interactions with the physical tissue environment stabilize particular phenotypes and sculpt the overall population structure [5], [9], [10]. In addition, tumors consist of cells that are genetically heterogeneous, even when derived from a single clone, and genetic instability may contribute to the generation of additional variation with disease progression [11]. A longitudinal understanding of tumor heterogeneity is essential for improving clinical treatments and outcomes [12], [13]. In recent years, cell-based methods have been complemented by high-throughput sequencing and methods for the genomic and transcriptomic analysis of individual single cells. Advances in these areas have also facilitated the development of cellular barcoding methods that permit the tracking of individual cell trajectories with exquisite detail. Barcoding technologies enable the quantification of heterogeneous populations and the evaluation of changes in phenotypic state, in the context of selective pressures arising from the cell population, tissue and environmental interactions, and external perturbations, such as drug treatment.

This review will explore the development and applications of barcoding technology in recent years, including biological questions, techniques, challenges, and outcomes. Specifically, we will delve into the ways that barcoding tools are paving the way for new mechanistic insights into tumor heterogeneity, progression, treatment response and relapse [14].

Overview of genomic barcoding platforms

While numerous types of genomic barcoding technologies are now in use, they share an overall conceptual strategy. A unique random or semi-random nucleic acid sequence is introduced into each founder cell in the population, usually by viral delivery. This label becomes integrated into the genome and is heritable to daughter cells, serving to tag the clone and all of its descendants over many expansions and generations [15], [16] **(Figure 1A).** In general, barcode labeling occurs by infection with lentivirus at a low MOI, to optimize the delivery of a single barcode per cell. Positively-transduced cells are then selected by antibiotic resistance or a fluorescent reporter to ensure that the starting cell population consists entirely of barcode-labeled cells **(Figure 1A)**. The number of initial founder cells in this labeled population determines the barcode diversity; a starting population of 1000 cells will include 1000 unique barcode labels. With cell subculturing and expansion, the relative frequency of these 1000 barcodes may change, as some cells proliferate more rapidly than others and some cells are lost from the population. Importantly, it is straightforward to determine the frequency of these 1000 clones in the population by high-throughput sequencing of the ensemble of barcodes. The number of reads of each barcode sequence corresponds to the frequency of that clone in the cell population. In some variations of barcoding technologies, the integrated sequence is also expressed as a transcript that can be detected by standard scRNA-Seq methods allowing detailed transcriptomic analysis of the clonal population [17]–[22] **(Figure 1B**). Measuring barcode frequencies and barcode-resolved single cell transcriptomes over longitudinal time points enables the study of trajectories of clonal behaviors within an overall heterogeneous tumor cell population. In another extension of the technology, the ClonMapper and CaTCH technologies add additional functionalities to clone

tracking studies by enabling the clone-specific expression of a reporter for subpopulation isolation [21], [23], [24] (**Figure 1C**). In this application, expressed barcodes consist of random sgRNA sequences and dCas9-TA drives the activation of synthetic gene circuits to induce clone-specific fluorescent reporter expression. Transcriptional activation of functional genes has also been demonstrated, enabling clone-specific control of cell fate decisions, such as activation of the apoptosis program by BAX [23].

Cancer heterogeneity and the contributions of selection and plasticity

Intratumor heterogeneity contributes to numerous aspects of tumor progression, metastasis, therapeutic response, and disease relapse. The quantitation of clonal dynamics by nucleic acid barcodes is particularly well-suited to studies of large, heterogeneous, evolving cell populations.

By enabling the quantitation of clonal frequency in parallel with clonal transcriptomic state, barcoded cancer cells shed light on the contributions of selection and non-genetic plasticity to resistance mechanisms [25], [26]. While much work has sought to explain the genetic underpinnings of tumor resistance and relapse, in recent years non-genetic mechanisms have also been widely studied. The relative contributions of these processes vary in different tumor settings, drug treatments and drug schedules [27]–[30]. With the advantages of replicate cell libraries and longitudinal quantitative observation, studies of barcoded cells are beginning to reveal the complexities of cancer cell selection and plasticity.

A unique feature of working with barcoded cell libraries is ability to compare clonal behaviors across experimental replicates (*in vitro* replicate cultures or *in vivo* replicate tumors). The parental barcoded cell population is split into replicates, each consisting of a comparable

collection of clones. Clones that display similar dynamics across multiple replicates may have a pre-existing genetic mutation that contributes to fitness. In contrast, a clone that exhibits differences in abundance (reflecting net survival and proliferation rate) across replicates may have acquired a *de novo* alteration, which may be genetic or epigenetic [9], [31]. Barcoded cell studies using this experimental design have identified examples of both types of outcomes.

Drug resistance is linked to heritable genomic alterations of the cells when a minority of clones in the initial cell population undergo selection and become highly enriched in the population after drug treatment [21], [32]–[36]. One of the first high-complexity cancer cell barcoding systems, the ClonTracer technology, was used to track more than 1 million unique barcode labels in the non–small cell lung cancer cell line HCC827 which harbors an activating epidermal growth factor receptor (EGFR) mutation that confers sensitivity to the EGFR inhibitor erlotinib [32], [33]. Due to the large number of barcodes in this system, the contributions of rare clones could be detected. Treatment of replicate barcoded cultures revealed that a small pre-existing subpopulation, representing 0.05% of the initial population, was selected and expanded under erlotinib treatment. Other studies used ClonTracer to pinpoint rare resistant clones in the KCL-22 model of chronic myeloid leukemia that were enriched and swept the population following nilotinib or imatinib treatment [33].

In other studies, clonal selection is not associated with therapeutic resistance, as the barcode diversity is not significantly reduced. Instead, cells from many clones may have the potential to evade the drug through the activation of various non-genetic resistance mechanisms [19], [37]–[39]. An *in vivo* study of colorectal cancer (CRC) patient-derived cells engrafted

barcoded tumor cells into NOD/SCID mice and found no loss of clonal diversity upon chemotherapy treatment. Additionally, the barcode composition of individual tumors was unique, and the survivor cells appeared to escape treatment by entering a drug-tolerant persister state that resembles diapause [39]. Cell fates and clonal dynamics of acute myeloid leukemia have been elucidated with the single-cell profiling and lineage tracing technology (SPLINTR), which utilizes expressed barcodes [19]. While malignant clonal dominance is a clone-intrinsic property of leukemia cells, increased transcriptional heterogeneity was also a consistent feature of clonal fitness [19].

Even within individual studies, selection and plasticity may both be observed. A conditional mouse model of mammary tumorigenesis MMTV-rtTA;TetO-neu has been utilized with lentiviral-mediated cell barcoding to study the clonal dynamics of disease relapse[8]. Approximately half of the recurrent tumors in this model were marked by clonal dominance, with one or two clones demonstrating de novo acquisition of Met amplification and sweeping the population. The other tumors were polyclonal, with no obvious reduction in clonal diversity compared to the primary tumors and characterized by alterations in the IL-6—Jak/Stat3 pathway. In this model then, recurrence can proceed through several distinct routes, and it is not yet clear how tumor-intrinsic factors, microenvironmental conditions, or stochasticity may contribute to the particular clonal dynamics of any individual tumor.

Understanding the specific relative contributions of selection and non-genetic plasticity in different tumor types and patients may be critical for achieving a durable response to treatment. Barcoded cell populations provide a new tool for dissecting these cancer cell responses and

studies tracking clonal dynamics are already highlighting the range of diverse interactions that may be observed [40]–[42].

Barcoding examines variation in metastatic potential

Sites of metastasis and mechanisms vary depending on the type of cancer, but even within a lesion there is considerable variation in invasion potential and metastatic-initiating potential among cells. Only a small fraction of malignant cells have the capacity to seed secondary tumors sites [12], [37], [40], [43]. Therefore, there is a notable interest in characterization of the qualities and clonal dynamics of cells comprising successful metastases [44], [45]. Barcoding technology provides an opportunity to study these characteristics of metastasis *in vivo* [43], [46]–[48].

One recent study involving the use of barcoding technology for investigation of metastasis utilized a Cas9-based, single-cell lineage tracer in a lung cancer xenograft mouse model to study factors related to metastasis [49]. A Cas9-based barcoding method was utilized in which a target site was cleaved by Cas9 and an inheritable allele inserted which could then be tracked in future generations. This allowed for single-cell sequencing of individual cells and their progeny in addition to rates of metastasis for tumor populations, the heritability of these metastatic phenotypes, and tissue routes of metastasis [49].

It is commonly accepted that the tumor microenvironment is highly heterogeneous with some populations potentially possessing characteristics resulting in treatment resistance or increased

metastasis. However, it is unknown whether these characteristics are pre-existing within cancer population or if they arise evolutionarily through disease progression. DNA barcoding and gene expression recording have been employed to this end, to reveal the presence of cancer cells with unique properties during tumor progression [50]. In this investigation, mouse melanoma cells B16F10 and BL6, and mouse mammary carcinoma cells, 4T1 were transfected using a lentivirus to generate clones with barcodes 30 base pairs each. Results indicated negative impacts on tumor growth throughout treatment without a notable decrease in barcode expression.

Ultimately, an understanding of the genetic or epigenetic basis for tumor heterogeneity could inform our understanding of cancer progression and dispersion within the body [49]. Barcoding techniques provide a unique platform for investigation of these phenomena, allowing for a clearer understanding of the interactions taking place and the characteristics of specific cells which can be isolated for investigation. Future work could investigate timing or order of genetic alterations resulting in malignancy, phenotypic changes as a result of tumor microenvironments, or the acquisition of resistance to treatments.

Harnessing barcoding in therapy development and clinical applications

Preclinical drug evaluation has often measured the sensitivity of cell lines and primary human cells, without considering the heterogeneity of cellular responses. While advances in organoid culture and biomimetic models now incorporate key features of the tissue environment to increase cell diversity, traditional cell culture models may also harbor heterogeneous subpopulations. Designing *in vitro* studies to capture this subpopulation information may provide new insights. A large-scale drug screening effort of 578 human cancer cell lines was performed

using the PRISM platform of nucleic acid barcoding with pooled screening, to assess the potential for drug-repurposing of non-oncological agents. [51]. Researchers screened 4,518 drugs from the Drug Repurposing Hub in a 2-stage dosing protocol. Six non-oncological treatments demonstrated cytotoxic effects on cancer cells; however, the genomic characteristics of the cell lines suggests these cytotoxic effects are likely variable. In some cases, CRISPR/Cas9 loss-of-function and gain-of-function screening shed additional light on molecular pathways of some drug targets and with further contributions this type of dataset may become a valuable tool in preclinical testing and therapeutic development.

Other studies have employed barcoded cells to explore the effects of treatment schedules on the evolution of drug resistance in heterogeneous cancer cell populations [52]. An extensive characterization of the spectrum of cell responses in the triple negative breast cancer MDA-MB-231 cell line compared 696 treatment conditions, all administering concurrent or sequential crizotinib and navitoclax with different intervals in exposure window and recovery time. Quantitation of barcode dynamics and gene expression states with expressed barcodes revealed that navitoclax may selects for pre-existing resistant clones and that the 2 drugs in combination at low doses resulted in a similar distribution of survivor barcodes to a single high dose of crizotinib. Importantly, the response of subpopulations depends on the history of prior drug exposure, a phenomenon which has been highlighted in numerous studies and can now be directly measured using barcode dynamics [47], [48]. Future studies in the field will likely focus on the molecular mechanisms by which clonal subpopulations interact and may mutually impact therapeutic response and treatment resistance.

Challenges and future considerations

Despite the rapid growth of new platforms for genomic barcoding, there remain some technological considerations and limitations. As clones are identified by relatively short nucleic acid sequences, mutations to the barcode sequence may be introduced in the sequencing step or earlier during PCR amplification of the barcode ensemble. This presents a challenge in distinguishing whether two highly similar barcodes in a dataset represent distinct clones or are derived from the same clonal population before and after a mutation event. Deep sequencing of the initial barcode-labeled cell population can generate a "whitelist" of verified barcode labels, although it may not be possible to completely exclude the possibility of rare clones with highly similar sequence. Integrating transcribed barcode labels with single-cell transcriptomic measurements introduces a new technical consideration, as only partial barcode data may be captured in an individual cell. In addition, single-cell experimental design needs to consider the starting diversity of the barcoded population and the expected frequency of clones in the dataset, in order to obtain sufficient individual cell measurements of each.

The genotypic and phenotypic heterogeneity among barcoded cells reflects the population status at the instantiation of labeling. A cell population that is genetically homogenous at the time of barcode introduction will include many clones with similar genotype, although they are tagged with different barcodes. In a population with high genetic heterogeneity, it is possible that low frequency genotypes may be lost. All protocols for barcode delivery and integration introduce the potential for clonal selection, as some clones may be less amenable or less robust to viral delivery or transfection. One practical biological challenge is to monitor the potential loss of barcode diversity with long-term passaging of cell lines *in vitro* and to understand how various

culture conditions impact this process. Similarly, in xenograft studies it is necessary to measure the variation among clones in rates of engraftment, immune escape, and proliferation rates *in vivo*. An absence of concordance among clonal-specific behaviors *in vitro* and *in vivo* may identify new considerations for improving the physiological relevance of cell culture models.

The growing adoption of barcoding methodologies presents a number of unique opportunities for the cancer biology field. Once a barcoded cell population has been generated for a tumor cell line or primary cell population, it becomes possible to compare the behavior of specific clones across different experimental settings, in different culture conditions or microenvironments, and under different therapeutic pressures. Barcoded cell models may therefore be a key to the facilitation of sharing and synergy across laboratories, as the barcodes serve as a stable index for identifying a particular clone over time. We envision the generation of databases that compile clonal cell behaviors and subpopulation behaviors within each cancer cell model. These resources could serve as tools to integrate multidimensional characterization of individual clones, encompassing transcriptomic and genomic information about specific clonal subpopulations, and also drug sensitivity, migration rate, invasion rate, measurements of physical force and more. The enormous potential to build cellular barcoding into existing experimental workflows is therefore catalyzing a shift towards a deeper understanding of heterogeneous cancer cells.

Acknowledgments

The authors thank Sui Huang, Arja Kaipainen and Aziz Al'Khafaji for ongoing discussions and are grateful for funding support through the NIH (1U01CA25354 to AB) and DOD (W81XWH-

18-1-0420). KH is supported by an NSF Graduate Fellowship. The authors report no financial conflicts of interest.

Bibliography

[1] W. M. Elsasser, "Outline of a theory of cellular heterogeneity.," *Proc Natl Acad Sci U S A*, vol. 81, no. 16, pp. 5126–5129, Aug. 1984.

[2] "The Origin of Phenotypic Heterogeneity in a Clonal Cell Population In Vitro," *PLOS ONE*, vol. 2, no. 4, p. e394, Apr. 2007, doi: 10.1371/journal.pone.0000394.

[3] "Role of network-mediated stochasticity in mammalian drug resistance | Nature Communications." https://www.nature.com/articles/s41467-019-10330-w (accessed Jan. 23, 2023).

[4] E. Kim, J.-Y. Kim, M. A. Smith, E. B. Haura, and A. R. A. Anderson, "Cell signaling heterogeneity is modulated by both cell-intrinsic and -extrinsic mechanisms: An integrated approach to understanding targeted therapy," *PLOS Biology*, vol. 16, no. 3, p. e2002930, Mar. 2018, doi: 10.1371/journal.pbio.2002930.

[5] A. Brock, H. Chang, and S. Huang, "Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours," *Nat Rev Genet*, vol. 10, no. 5, pp. 336–342, May 2009, doi: 10.1038/nrg2556.

[6] T. M. A. Neildez-Nguyen *et al.*, "Epigenetic gene expression noise and phenotypic diversification of clonal cell populations," *Differentiation*, vol. 76, no. 1, pp. 33–40, Jan. 2008, doi: 10.1111/j.1432-0436.2007.00219.x.

[7] C. Sonnenschein and A. M. Soto, "Over a century of cancer research: Inconvenient truths and promising leads," *PLoS Biol*, vol. 18, no. 4, p. e3000670, Apr. 2020, doi: 10.1371/journal.pbio.3000670.

[8] P. Kulkarni *et al.*, "Addressing the genetic/nongenetic duality in cancer with systems biology," *Trends Cancer*, pp. S2405-8033(22)00265–5, Jan. 2023, doi: 10.1016/j.trecan.2022.12.004.

[9] M. Greaves and C. C. Maley, "Clonal evolution in cancer," *Nature*, vol. 481, no. 7381, Art. no. 7381, Jan. 2012, doi: 10.1038/nature10762.

[10] N. McGranahan and C. Swanton, "Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution," *Cancer Cell*, vol. 27, no. 1, pp. 15–26, Jan. 2015, doi: 10.1016/j.ccell.2014.12.001.

[11] P. C. Nowell, "The clonal evolution of tumor cell populations," *Science*, vol. 194, no. 4260, pp. 23–28, Oct. 1976, doi: 10.1126/science.959840.

[12] A. Marusyk and K. Polyak, "Tumor heterogeneity: causes and consequences," *Biochim Biophys Acta*, vol. 1805, no. 1, p. 105, Jan. 2010, doi: 10.1016/j.bbcan.2009.11.002.

[13] A. G. X. Zeng *et al.*, "A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia," *Nat Med*, vol. 28, no. 6, pp. 1212–1223, Jun. 2022, doi: 10.1038/s41591-022-01819-x.

[14] A. Serrano, J. Berthelet, S. H. Naik, and D. Merino, "Mastering the use of cellular barcoding to explore cancer heterogeneity," *Nat Rev Cancer*, vol. 22, no. 11, Art. no. 11, Nov. 2022, doi: 10.1038/s41568-022-00500-2.

[15] N. Navin *et al.*, "Inferring tumor progression from genomic heterogeneity," *Genome Res.*, vol. 20, no. 1, pp. 68–80, Jan. 2010, doi: 10.1101/gr.099622.109.

[16] Y. Wang *et al.*, "Clonal evolution in breast cancer revealed by single nucleus genome sequencing," *Nature*, vol. 512, no. 7513, Art. no. 7513, Aug. 2014, doi: 10.1038/nature13600.

[17] B. A. Biddy *et al.*, "Single-cell mapping of lineage and identity in direct reprogramming," *Nature*, vol. 564, no. 7735, pp. 219–224, Dec. 2018, doi: 10.1038/s41586-018-0744-4.

[18] C. Weinreb, A. Rodriguez-Fraticelli, F. Camargo, and A. M. Klein, "Lineage tracing on transcriptional landscapes links state to fate during differentiation," *Science*, vol. 367, no. 6479, p. eaaw3381, Feb. 2020, doi: 10.1126/science.aaw3381.

[19] K. A. Fennell *et al.*, “Non-genetic determinants of malignant clonal fitness at single-cell resolution,” *Nature*, vol. 601, no. 7891, Art. no. 7891, Jan. 2022, doi: 10.1038/s41586-021-04206-7.
[20] Y. Oren *et al.*, “Cycling cancer persister cells arise from lineages with distinct programs,” *Nature*, vol. 596, no. 7873, pp. 576–582, Aug. 2021, doi: 10.1038/s41586-021-03796-6.
[21] C. Gutierrez *et al.*, “Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment,” *Nat Cancer*, vol. 2, no. 7, pp. 758–772, Jul. 2021, doi: 10.1038/s43018-021-00222-8.
[22] K. E. Johnson *et al.*, “Integrating transcriptomics and bulk time course data into a mathematical framework to describe and predict therapeutic resistance in cancer,” *Phys Biol*, vol. 18, no. 1, p. 016001, Nov. 2020, doi: 10.1088/1478-3975/abb09c.
[23] A. M. Al’Khafaji, D. Deatherage, and A. Brock, “Control of Lineage-Specific Gene Expression by Functionalized gRNA Barcodes,” *ACS Synth Biol*, vol. 7, no. 10, pp. 2468–2474, Oct. 2018, doi: 10.1021/acssynbio.8b00105.
[24] C. Umkehrer *et al.*, “Isolating live cell clones from barcoded populations using CRISPRa-inducible reporters,” *Nat Biotechnol*, vol. 39, no. 2, pp. 174–178, Feb. 2021, doi: 10.1038/s41587-020-0614-0.
[25] S. Huang, “Reconciling Non-Genetic Plasticity with Somatic Evolution in Cancer,” *Trends Cancer*, vol. 7, no. 4, pp. 309–322, Apr. 2021, doi: 10.1016/j.trecan.2020.12.007.
[26] M. Pillai and M. K. Jolly, “Systems-level network modeling deciphers the master regulators of phenotypic plasticity and heterogeneity in melanoma,” *iScience*, vol. 24, no. 10, p. 103111, Sep. 2021, doi: 10.1016/j.isci.2021.103111.
[27] C. Neftel *et al.*, “An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma,” *Cell*, vol. 178, no. 4, pp. 835-849.e21, Aug. 2019, doi: 10.1016/j.cell.2019.06.024.
[28] S. Yuan, R. J. Norgard, and B. Z. Stanger, “Cellular Plasticity in Cancer,” *Cancer Discov*, vol. 9, no. 7, pp. 837–851, Jul. 2019, doi: 10.1158/2159-8290.CD-19-0015.
[29] N. D. Marjanovic *et al.*, “Emergence of a High-Plasticity Cell State during Lung Cancer Evolution,” *Cancer Cell*, vol. 38, no. 2, pp. 229-246.e13, Aug. 2020, doi: 10.1016/j.ccell.2020.06.012.
[30] J. Foo *et al.*, “Roadmap on plasticity and epigenetics in cancer,” *Phys Biol*, vol. 19, no. 3, Apr. 2022, doi: 10.1088/1478-3975/ac4ee2.
[31] P. A. Jones and M. L. Gonzalgo, “Altered DNA methylation and genome instability: A new pathway to cancer?,” *Proc Natl Acad Sci U S A*, vol. 94, no. 6, pp. 2103–2105, Mar. 1997.
[32] A. N. Hata *et al.*, “Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition,” *Nat. Med.*, vol. 22, no. 3, pp. 262–269, Mar. 2016, doi: 10.1038/nm.4040.
[33] H. C. Bhang *et al.*, “Studying clonal dynamics in response to cancer therapy using high-complexity barcoding,” *Nat Med*, vol. 21, no. 5, pp. 440–448, May 2015, doi: 10.1038/nm.3841.
[34] K. Hinohara *et al.*, “KDM5 histone demethylase activity links cellular transcriptomic heterogeneity to therapeutic resistance,” *Cancer Cell*, vol. 34, no. 6, pp. 939-953.e9, Dec. 2018, doi: 10.1016/j.ccell.2018.10.014.
[35] F. Caiado *et al.*, “Lineage tracing of acute myeloid leukemia reveals the impact of hypomethylating agents on chemoresistance selection,” *Nat Commun*, vol. 10, p. 4986, Nov. 2019, doi: 10.1038/s41467-019-12983-z.
[36] C. Yu *et al.*, “High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines,” *Nat Biotechnol*, vol. 34, no. 4, pp. 419–423, Apr. 2016, doi: 10.1038/nbt.3460.
[37] G. V. Echeverria *et al.*, “Resistance to neoadjuvant chemotherapy in triple negative breast cancer mediated by a reversible drug-tolerant state,” *Sci Transl Med*, vol. 11, no. 488, p. eaav0936, Apr. 2019, doi: 10.1126/scitranslmed.aav0936.

[38] D. Merino *et al.*, "Barcoding reveals complex clonal behavior in patient-derived xenografts of metastatic triple negative breast cancer," *Nat Commun*, vol. 10, no. 1, Art. no. 1, Feb. 2019, doi: 10.1038/s41467-019-08595-2.
[39] S. K. Rehman *et al.*, "Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy," *Cell*, vol. 184, no. 1, pp. 226-242.e21, Jan. 2021, doi: 10.1016/j.cell.2020.11.018.
[40] C. A. Klein, "Parallel progression of primary tumours and metastases," *Nat Rev Cancer*, vol. 9, no. 4, Art. no. 4, Apr. 2009, doi: 10.1038/nrc2627.
[41] N. Navin *et al.*, "Tumour evolution inferred by single-cell sequencing," *Nature*, vol. 472, no. 7341, pp. 90–94, Apr. 2011, doi: 10.1038/nature09807.
[42] K. Anderson *et al.*, "Genetic variegation of clonal architecture and propagating cells in leukaemia," *Nature*, vol. 469, no. 7330, pp. 356–361, Jan. 2011, doi: 10.1038/nature09650.
[43] L. V. Nguyen *et al.*, "DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts," *Nat Commun*, vol. 5, p. 5871, Dec. 2014, doi: 10.1038/ncomms6871.
[44] S. K. Singh *et al.*, "Identification of human brain tumour initiating cells," *Nature*, vol. 432, no. 7015, pp. 396–401, Nov. 2004, doi: 10.1038/nature03128.
[45] M. Zhang *et al.*, "Identification of Tumor-initiating Cells in a p53 Null Mouse Model of Breast Cancer," *Cancer Res*, vol. 68, no. 12, pp. 4674–4682, Jun. 2008, doi: 10.1158/0008-5472.CAN-07-6353.
[46] S. Seth *et al.*, "Pre-existing Functional Heterogeneity of Tumorigenic Compartment as the Origin of Chemoresistance in Pancreatic Tumors," *Cell Rep*, vol. 26, no. 6, pp. 1518-1532.e9, Feb. 2019, doi: 10.1016/j.celrep.2019.01.048.
[47] G. R. Howard, T. A. Jost, T. E. Yankeelov, and A. Brock, "Quantification of long-term doxorubicin response dynamics in breast cancer cell lines to direct treatment schedules," *PLoS Comput Biol*, vol. 18, no. 3, p. e1009104, Mar. 2022, doi: 10.1371/journal.pcbi.1009104.
[48] J. B. West, M. N. Dinh, J. S. Brown, J. Zhang, A. R. Anderson, and R. A. Gatenby, "Multidrug Cancer Therapy in Metastatic Castrate-Resistant Prostate Cancer: An Evolution-Based Strategy," *Clin Cancer Res*, vol. 25, no. 14, pp. 4413–4421, Jul. 2019, doi: 10.1158/1078-0432.CCR-19-0006.
[49] J. J. Quinn *et al.*, "Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts," *Science*, vol. 371, no. 6532, p. eabc1944, Feb. 2021, doi: 10.1126/science.abc1944.
[50] Y. Umeki, N. Ogawa, Y. Uegaki, K. Saga, Y. Kaneda, and K. Nimura, "DNA barcoding and gene expression recording reveal the presence of cancer cells with unique properties during tumor progression," *Cell. Mol. Life Sci.*, vol. 80, no. 1, p. 17, Dec. 2022, doi: 10.1007/s00018-022-04640-4.
[51] S. M. Corsello *et al.*, "Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling," *Nat Cancer*, vol. 1, no. 2, pp. 235–248, Feb. 2020, doi: 10.1038/s43018-019-0018-6.
[52] G. A. Patwardhan, M. Marczyk, V. B. Wali, D. F. Stern, L. Pusztai, and C. Hatzis, "Treatment scheduling effects on the evolution of drug resistance in heterogeneous cancer cell populations," *NPJ Breast Cancer*, vol. 7, p. 60, May 2021, doi: 10.1038/s41523-021-00270-4.

Figure Legends

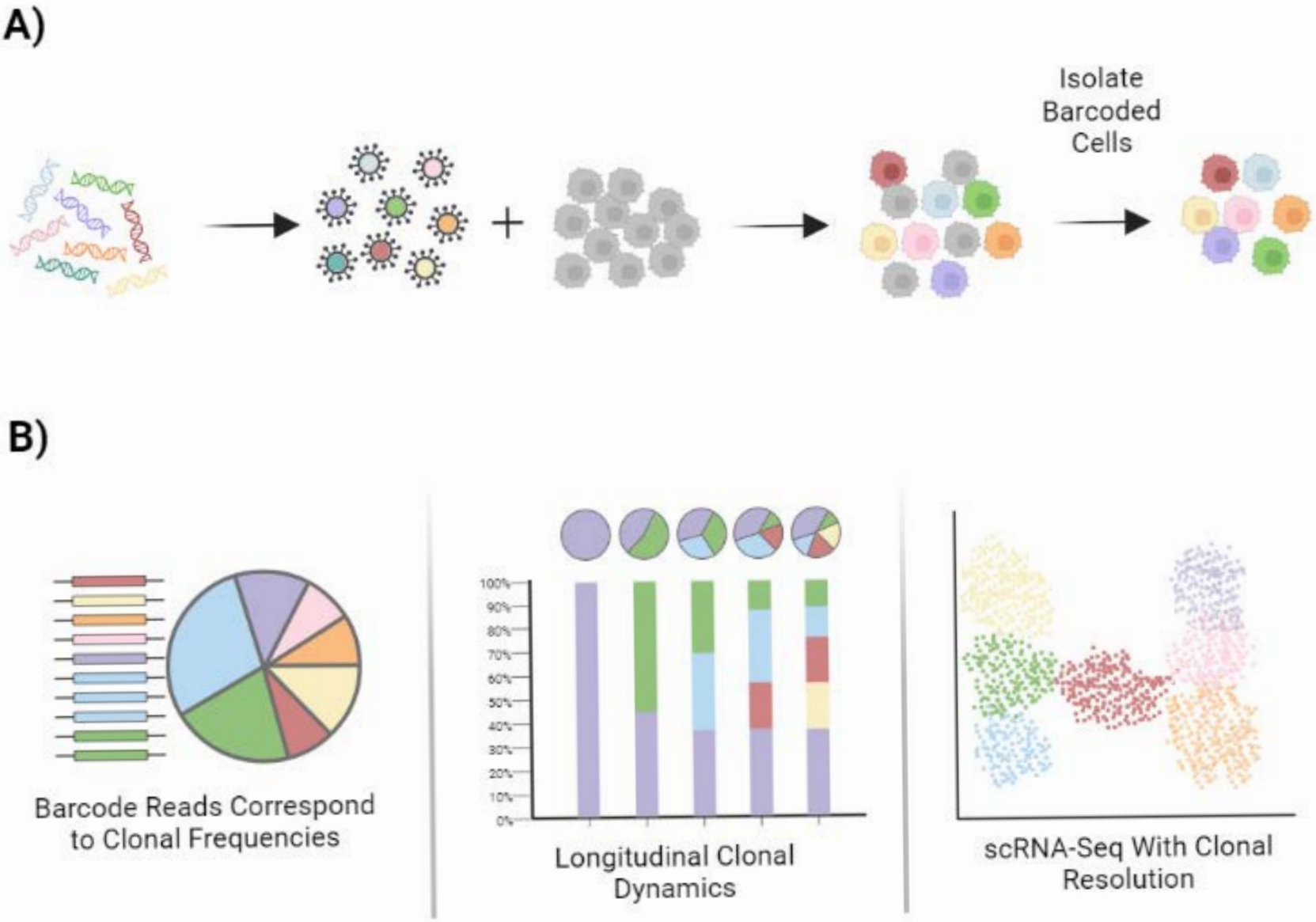
Figure 1 Genomic Barcoding. The general process and outcomes of genomic barcoding are similar among technologies. **A)** Barcodes are delivered into cells by viral transduction and cells which successfully integrate a barcode are isolated via FACS or antibiotic selection. **B)** Investigations using barcoded cells measure clonal frequencies and longitudinal clonal dynamics by high-resolution genomic sequencing of the barcode ensemble. Clonally-resolved transcriptome data is measured by scRNA-Seq. **C)** Some barcoding techniques allow for clone-specific activation of fluorescent reporter genes, allowing for isolation of particular clones and further live cell or molecular characterization.

Figure 2. DNA barcodes enable population replicates. Quantitation of clonal population structures in replicate studies is useful for the categorization of underlying biological processes that impact changes in population structure. Analysis of barcoded replicates may identify pre-existing variants or *de novo* changes that arise in many different clones upon treatment or perturbation. **A) H**igh resolution sequencing of the barcode ensemble may reveal diverse clonal population structures across replicates ,which may result from de novo alterations following drug treatment. **B)** In contrast, barcode quantitation may find that consistent subsets of clones dominate, suggesting the presence of pre-existing resistant clones in the initial cancer cell population.

Figure 3. Uncovering the dynamics of metastasis with cell barcoding. Barcoding techniques provide a platform for investigation of tumor phenotypes such as metastasis *in vivo*. **A)**

Barcoded cells may be implanted into mice to form primary tumors and secondary tumors can be isolated and assessed for clonal frequency and cellular characteristics. **B)** Complementary *in vitro* studies may also be performed, such as transwell migration of barcoded cells with or without chemoattractant or ECM changes. Cells which migrate through the transwell can be identified by their barcode and the relationship to clonal identity and gene expression state can assessed.

Figure 1


A)
Isolate
Barcoded
Cells
B)
Barcode Reads Correspond
to Clonal Frequencies
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Longitudinal Clonal
Dynamics
scRNA-Seq With Clonal
Resolution
C)
Further Characterization
of Isolated Clones
Genomics
Transcriptomics
Epigenomics
Proteomics
Functional Studies

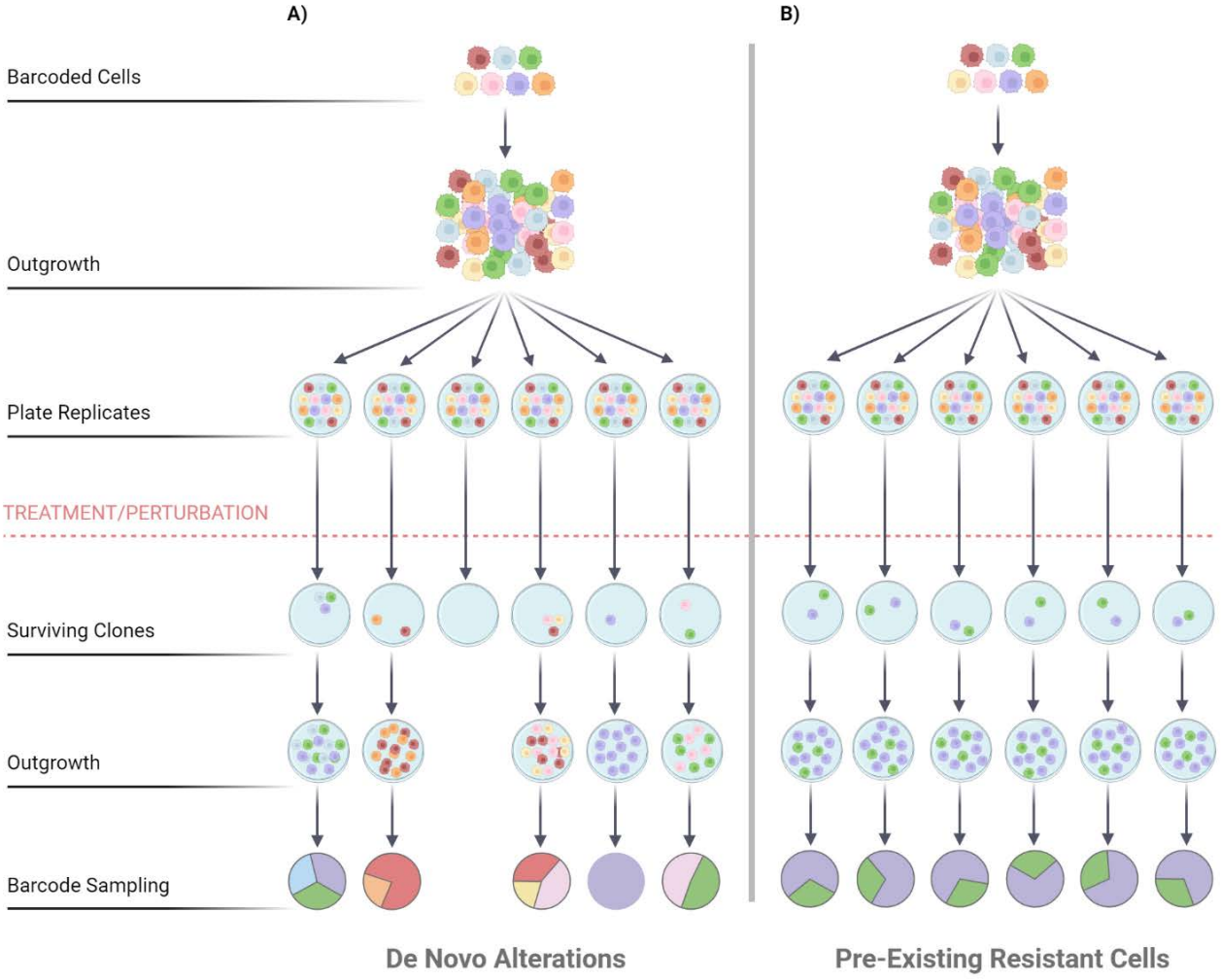
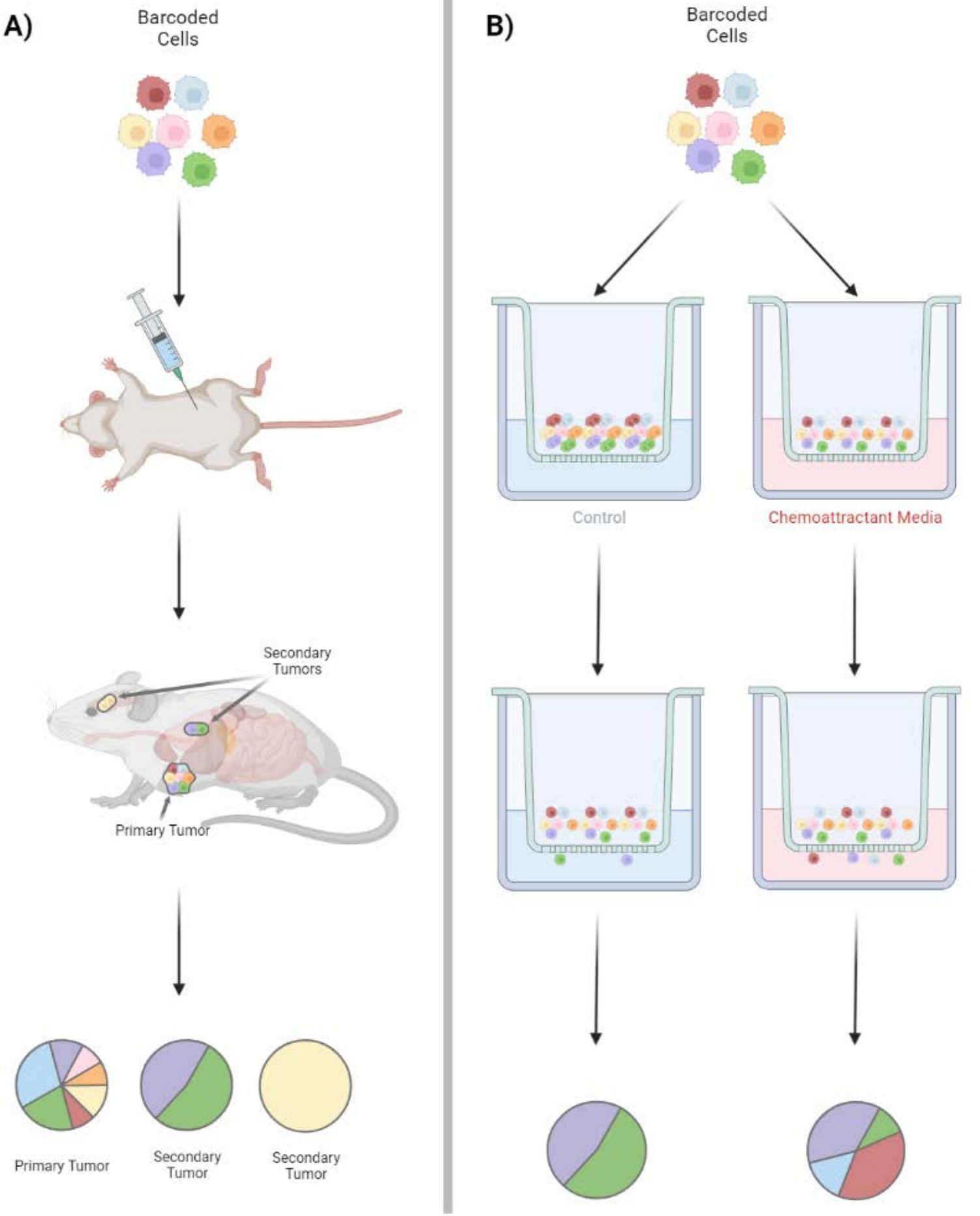

A)
B)
Barcoded Cells
Outgrowth
Plate Replicates
TREATMENT/PERTURBATION
Surviving Clones
Outgrowth
Barcode Sampling
De Novo Alterations
Pre-Existing Resistant Cells

Figure 3


A)
Barcoded Cells
Secondary Tumors
Primary Tumor
Primary Tumor
Secondary Tumor
Secondary Tumor
B)
Barcoded Cells
Control
Chemoattractant Media