Disentangling Sources of Phenotypic Variability in Oncogene-addicted Cancers

By

Corey Eduardo Hayford

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in

Biomedical Sciences
March 31, 2021
Nashville, Tennessee

Approved:

John A. Capra, Ph.D.

Bruce M. Damon, Ph.D.

Ken S. Lau, Ph.D.

Carlos F. Lopez, Ph.D.

John A. McLean, Ph.D.

Vito Quaranta, M.D.

To my friends and family, who always assured me along the way.

Acknowledgements

This would have not been possible without financial support from an NCI NRSA predoctoral fellowship award and NIGMS training grant. I would like to thank my mentor, Dr. Vito Quaranta, for providing an environment in which students can freely pursue ideas and projects. Your mentoring style allowed me to determine my own personal goals as a scientist, and for that independence I will always be grateful. I would also like to thank my committee members for their clear-sighted support and scientific expertise. Your insights have truly helped in my development as a scientist. I am also eternally grateful to Dr. Bruce Damon and Patty Mueller for being continuous student advocates.

I truly enjoyed the opportunity to work with peer scientists from various interdisciplinary backgrounds. To all of the Quaranta and Lopez lab members, I am grateful for your support and feedback. I would specifically like to thank Drs. Leonard Harris and Darren Tyson, who served as my secondary mentors and helped get projects across the finish line. To Sarah Maddox Groves and Geena Ildefonso, thank you for being friends, sounding boards, and a source of fun along the way. To Drs. Bishal Paudel and Keisha Hardeman, I appreciate your warm welcome and friendly scientific advice. To Drs. David Wooten and Christian Meyer, thank you for being colleagues who I admired. To Melaine Sebastian, you are missed. To other lab members and mentees, I am better for your knowledge and friendship.

I also want to extend a special thank you to my friends from the QCB program, specifically Ian Setliff, Mary Lauren Benton, and Andy Perreault. I truly enjoyed all of our shenanigans - you are friends who I appreciate both personally and professionally. To my family, I am thankful for all of your encouraging words and actions along the way. And finally, to my girlfriend Caroline Stuart, thank you for your love and providing a safe place for me, without judgement or scrutiny. Your support during these tough years have meant the world to me. Thanks to all my friends and family for believing in me, listening to my concerns, and providing helpful advice.

Table of Contents

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Introduction

Cancer Oncogene Addiction and Targeted Therapies

Cancer is among the leading causes of death worldwide[1]. For decades, traditional cancer treatments, such as chemotherapy and radiation, sometimes increased patient life expectancy while sacrificing quality of life[2]. The outcomes of these traditional therapies have improved slowly over the last several decades through minor modifications to dosing and timing, but patients still succumb to the disease. Over this time, accumulating research identified a subset of cancers that were reliant on specific genes for growth and survival, which became known as "oncogene addiction"[3,4]. Because these cancer types signaled strongly through specific genes, direct inhibition of these genes were believed to quell the disease. These observations served as a paradigm shift for how the community viewed these cancer types, and sparked a revolution of small molecule generation to block oncogenes[5,6]. These cancer types, which included EGFR- and ALK-mutant NSCLC, BRAF-mutant melanoma, HER2-mutant breast cancer, and BCR-ABL mutant chronic myeloid leukemia, became the diseases we find somewhat treatable today.

Although the mechanism of oncogene addiction is not entirely clear, targeting oncogenes has proven effective in the clinic. In many clinical studies, tumor size is reduced greatly (>30%), as defined by the RECIST criteria[7]. These targeted therapies were proved significantly more efficacious than traditional therapies for oncogene-addicted cancers in short-term studies[8–10]. However, long-term studies show that while many of these patients showed increased life expectancy with fewer side effects than traditional chemotherapies and radiation, the disease largely returned[11]. The drug evasion mechanism for many of these tumors were attributed to pre-existing or acquired mutations[12,13], some of which had greater fitness in treatment and sparked a resistant tumor. This theory of tumor clonal selection[14] (FIG. 1), where cells have an initial decline in response to targeted therapy followed by a rebound attributed to drug

resistance, dominated the field, arguably still to this day. Under this view, identification of all resistance-conferring mutations would solve the problem of tumor recurrence, as the community could develop new drugs specific for these mutations.

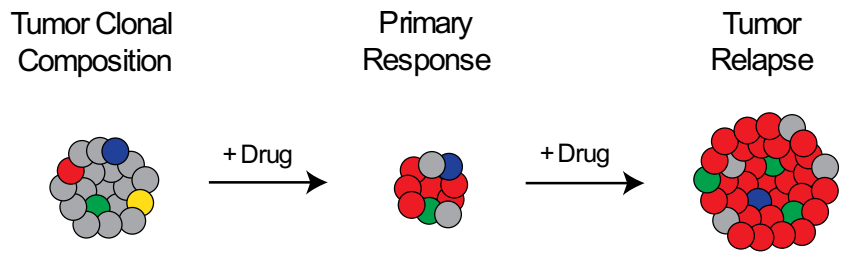

Figure 1 | *Clonal selection in cancer.*
Prior to drug treatment, the tumor is composed of heterogeneous clonal populations with differential fitness in drug treatment (left). After initial treatment, the tumor shrinks but clones with better in-drug fitness make up a larger proportion of the population (middle). In continued treatment, the most fit clones (red) overtake the tumor, leading to relapse.

An interesting case study for these ideas is EGFR-mutant NSCLC, which is driven by an exon 19 deletion in the EGFR protein that confers resistance to 1st and 2nd generation EGFR inhibitors (EGFRi). A specific point mutation (EGFR T790M) was identified in multiple resistant tumors, and led medicinal chemists to develop 3rd generation inhibitor Osimertinib[15], an EGFRi that specifically targets the point mutation. However, patient tumors eventually acquire resistance to Osimertinib, and the cycle continues. Other cancer types have moved towards combinations of targeted therapies to combat tumor resistance. In BRAF-mutant melanoma, a combination treatment of BRAF inhibitor (BRAFi) Vemurafenib and MEK inhibitor (MEKi) Trametinib was used to treat the disease[16]. Although few patients exhibited increased survival with this combination, tumors still invariably recurred[17].

Here inlies the problem: tumors have seem to have unrestrained ways to evade treatment. So, how can the disease be treated? Part of the answer lies in understanding the types of heterogeneity and

plasticity that plague treatment. Another aspect is developing tools to identify and separate variabilities. Finally, a method is necessary to quantify the outcomes targeting variability. Each of these aspects will be addressed in the sections below.

Tumor Variability

Tumor variability can be broadly separated into two categories: heterogeneity and plasticity[18]. Heterogeneity corresponds to variability at a single point in time, while plasticity refers to variability over time. Importantly, understanding both types of variability is key to understanding cancer drug sensitivity and potential escape mechanisms. Clinical responses are variable in response of oncogene addicted cancers to targeted therapies, with some studies showing early signs of resistance to treatments[19,20], while others show a form of drug tolerance after therapy that leads to eventual recurrence[12,21]. Analyses of post-resistant tumors or cells provide most of our current knowledge of tumor recurrence, usually attributed to rare, resistance conferring genetic alterations that either preexist or develop during therapy[22]. Although genetic alterations are the focus of a traditional view of tumor variability, variability has been derived from a variety of sources, and are discussed in a later section.

Heterogeneity can be further subdivided into intertumoral and intratumoral heterogeneity. Intertumoral heterogeneity, or variation between tumors, is often attributed to different genetic backgrounds, but tumors with similar genetic backgrounds also display a wide range of responsiveness to targeted therapies[23,24]. While effective targeted therapies induce clinical responses in the majority of patients, responses can vary greatly depending on the study[25]. Furthermore, cancer cell lines, derived from different patients, show remarkable variability in drug response phenotypes. Even different tumors derived from the same patient can respond very differently to therapy[12,26]. Intratumoral heterogeneity, or variation within a single tumor, is slightly more complex. While genetic variation has been identified within tumors across long timescales, non-genetic variation has become more appreciated as a

contributor to intratumoral heterogeneity at shorter timescales. Single-cell approaches have shown substantial variation within individual cell lines and tumors, which have been tied to variable outcomes[27,28].

Tumor plasticity has also been tied to poor patient outcomes[29,30]. As with heterogeneity, plasticity can arise from a variety of sources, but is largely shown in response to treatments. In studies of post-resistant tumors, it is difficult to discern between heterogeneity and plasticity. Therefore, much of the community knowledge on plasticity has come from time-series studies, specifically using single-cell approaches and lineage tracing technologies[31–33]. Acquired resistance mutations during drug treatment are the most prevalent examples of plasticity[12,13]. However, recent studies have shown epigenetic treatment evasion strategies where cells dynamically transition to more fit phenotypic states[34–36]. Therefore, it is likely that cancer cells employ a variety of these strategies to evade treatment, and while many fail, some are successful and lead to tumor recurrence[37–40].

Implicit in our understanding of tumor variability is that external influences will have an effect on drug-response phenotypes. For example, the tumor microenvironment has a large effect on treatment response and durability[41,42]. The heterogeneous population of cancer cells, immune infiltrate, secreted factors, and extracellular matrix is known to modify both treatment responses and metastatic resistance[43]. Furthermore, there are physiological influences on tumors that impact phenotype. Nutrient deprivation often leads to hypoxic conditions in the center of solid tumors, resulting in an expanding outer layer and necrotic core[44]. Adjacently, vascularity and angiogenesis create differential blood (i.e. nutrient) flow into tumors that further complicates our understanding of variability[45]. Some treatments have been developed to counteract these external forces[46,47]. While these forces are important, our cell line systems *do not take them into account* and therefore we do not consider their influences on our view directly. However, in the context of this framework, we can consider these external influences as boundary

conditions on the epigenetic landscape, potentially restricting or guiding cells into a region of the landscape.

Genetic, Epigenetic, and Stochastic Sources of Tumor Heterogeneity

Genetic differences among cancer cells within and across tumors have long been appreciated[48–52]. Indeed, genomic instability is a hallmark of cancer[53,54] and is considered to be the primary source of this genetic heterogeneity. However, it is becoming increasingly apparent that genetics alone cannot fully explain the wide ranges of responses observed in patient populations to anticancer therapies[55,56]. Epidermal growth factor receptor (EGFR) inhibitors, for instance, are not equally effective across EGFR-mutant lung cancer patients and in almost all cases tumors eventually acquire resistance and recur[22,57].Researchers, therefore, are increasingly looking to *non-genetic* sources of tumor heterogeneity for explanations. Broadly speaking, non-genetic heterogeneity comes in two forms: *epigenetic*[23,27,29,58–60], which is heritable (at least for a few generations), and *stochastic*[61–65], which is not heritable and arises due to intrinsic factors such as gene expression noise and asymmetric cell division[66,67]. Non-genetic heterogeneity has been linked to drug tolerance and decreased drug sensitivity *in vitro*[12,13,36,68,69] , *in vivo*[13,68,70], and clinically[71,72].

Prior work has shown heterogeneity as comprising genetic, epigenetic, and stochastic components that are broadly distinguished based on characteristic timescale of change: functional genetic mutations are acquired on the order of weeks to months[73,74], transitions between epigenetic states occur on the order of hours to weeks[75], and stochastic fluctuations in protein concentrations, and other sources of intracellular noise, operate on the order of seconds to minutes[76]. Separating processes based on characteristic timescale is a common approach in physics[77], and has a long history in biology[78]. Below, a detailed description of each level is provided, with precise definitions of terms that will be important for

later analyses. A simple visual illustration of these levels, and connections between them, are provided (FIG. 2).

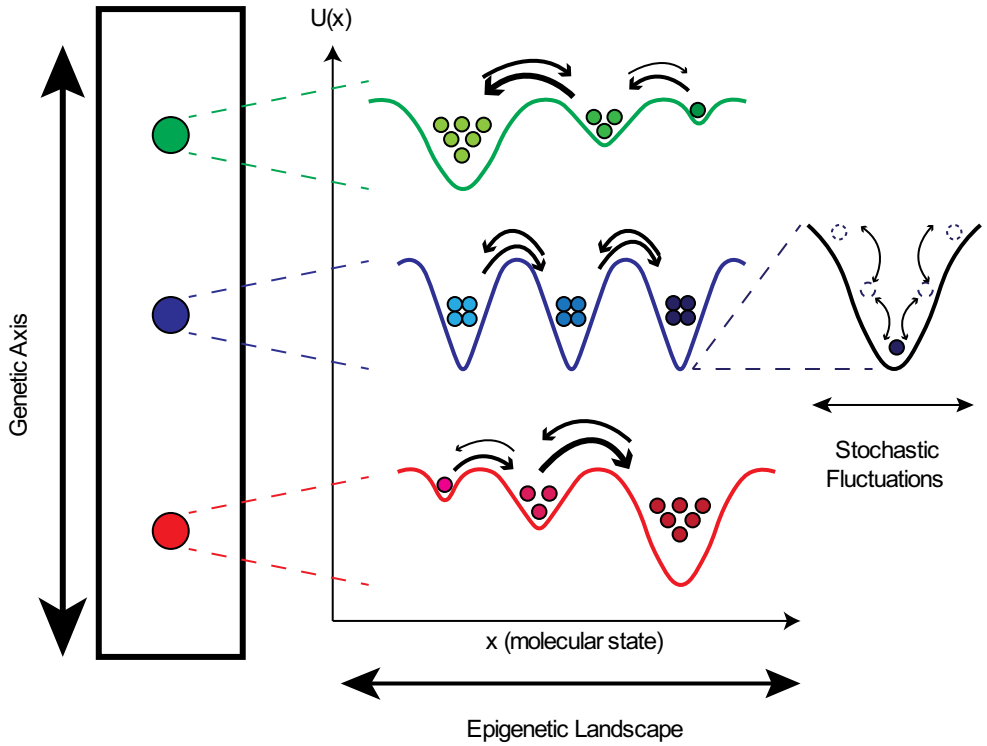

Figure 2 | *Multiple levels of heterogeneity are believed to operate within tumors.*
(*left*) The genetic "axis" defines mutational differences that have an effect on phenotype, e.g., drug sensitivity. (*middle*) Each genetic clone has an associated epigenetic landscape, where cells are distributed across basins, known as "attractors". The topography of the epigenetic landscape is defined by the dynamical biochemical network that controls cell fate and function. Molecular state (labeled as *x*) lies along the x-axis and the quasi-potential energy (labeled as *U(x)*) lies along the y-axis. (*right*) Cell states fluctuate within epigenetic basins due to intrinsic (e.g., gene expression) and extrinsic sources of noise. Most fluctuations are minor and do not significantly change the cell state but occasionally a large fluctuation results in a barrier crossing, i.e., a phenotypic state change.

Genetic (Level 1):

Genomic instability is a hallmark of cancer[53,54]. For cancer cells in culture, it has been estimated that approximately five million cell divisions are necessary to acquire at least one mutation per gene[73]. With more than 20,000 genes in the genome, this amounts to one mutation every two weeks on average. While mutations in the genome are relatively frequent, many do not have functional consequence, so-

called “passenger” mutations[58]. Therefore, we make a distinction between the *genomic state* of a cell and the *genetic state*: the genomic state is the full sequence of nucleotides that make up the DNA whereas the genetic state is the subset of this sequence that contributes to a specific cellular phenotype[79,80]. In other words, *two cells may differ genomically but be genetically identical if the mutations that differentiate them occur in portions of the genome that have no functional consequence on the phenotype of interest* (FIG. 3). As such, the timescale for generating genetic heterogeneity within a population of cancer cells can be quite slow (much slower than the acquisition of genomic heterogeneity), on the order of weeks to months[73,74]. Note that enumerating the subset of the genome that defines the genetic state can be difficult (if not impossible) in practice and depends on the phenotypic context[81]. As mentioned in previous sections, drug response is the phenotype of interest. Thus, the genetic state is, in principle, defined over all genes associated with drug response, which includes those involved in stress response, metabolism, cell cycle regulation, and many others.

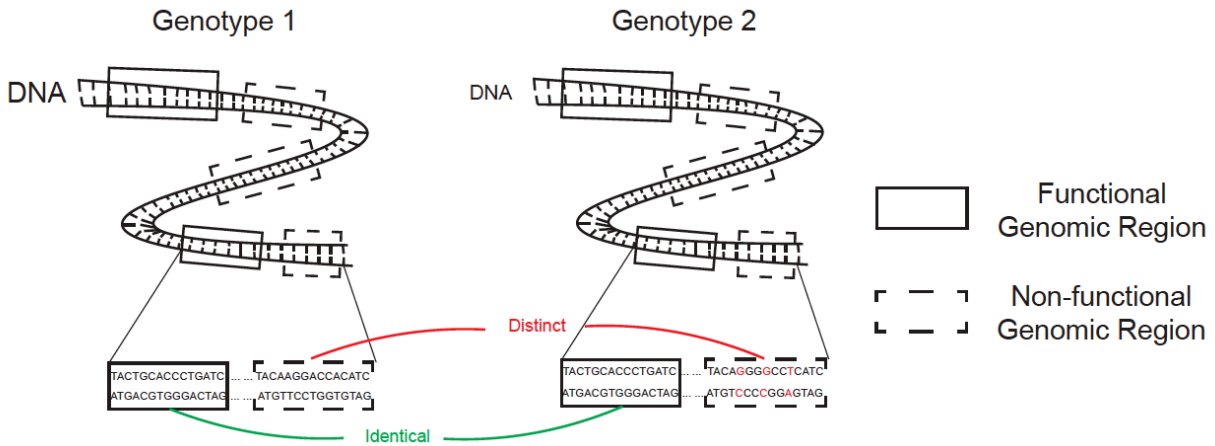

Figure 3 | *Difference between genomic and genetic variability.*
Cartoon illustration of two genomes that are genomically distinct but genetically identical with respect to phenotype (e.g., drug-response sensitivity). Functional genomic regions (solid black box) are those that have an impact on phenotype, while non-functional regions (dashed black box) do not.

Mutational events, by definition, are changes in the structure of a gene. However, these changes can take many different forms. The first scheme to categorize mutations is *variant classification*. Variant classification falls into two primary forms, sequence variants and structural variants. Sequence variants pertain to mutations of one or a few nucleotides, while structural variants deal with modification of larger genomic regions. Within the sequence variant category, there is a further breakdown into substitutions and indels. Substitutions are alterations where the length of the variant change is the same as the reference region. Substitutions can occur in a range of nucleotide lengths, but most occur as single nucleotide substitutions, called single nucleotide polymorphisms (SNPs). SNPs are the most common form of sequence variation, and heavily outweigh other length substitutions. The other type of sequence variation, indels, are broken down into insertions, deletions, and combinations (also referred to as indels). As the name suggests, insertions are an insertion of one or many nucleotides, while deletions are a deletion of one or several nucleotides. In the structural variant category, there are inversions, translocations, and copy number variants (CNVs). Inversions are nucleotide sequences that have been inverted but in the same position as prior to the mutation, while translocations are regions of the nucleotide sequence that has moved to a new position. By far, CNVs are the most measured types of structural variants, and correspond to duplications (also called amplifications; copy number gain compared to reference) and deletions (copy number loss where a contiguous region was removed).

A second categorization scheme is *variant consequence*. Variants of any classification can encode for any consequence. Variant consequence can initially be separated into synonymous and nonsynonymous mutations. Synonymous mutations, otherwise known as silent mutations, are changes in the DNA structure that encode for the same amino acid as the reference versions of the gene. Therefore, these genes are considered to have *no impact on phenotype*. Nonsynonymous mutations, conversely, are the class of all other mutations that encode for a different amino acid than the reference gene. These mutations *can have an impact on phenotype*, depending on the type of mutation and aforementioned

phenotypic context. Nonsynonymous mutations can be further broken down into missense and nonsense mutations. Missense mutations encode for a different amino acid compared to the reference, while nonsense mutations encode specifically for a start or stop codon where one was not encoded in the reference. Missense mutations can be further subdivided into conservative (same amino acid type – e.g. polar) and non-conservative (different amino acid type – e.g. polar changing to non-polar). In terms of impact, the general rankings of variant consequence (from low to high) is: synonymous/silent, conservative missense, non-conservative missense, and nonsense mutations. But, depending on the phenotypic context, these rankings *can* be variable. For example, the canonical EGFR T790M mutation that encodes for resistance to EGFRi, is a missense mutation that has profound impact.

Tools such as Ensembl Variant Effect Predictor[82] (VEP) and SnpEff[83] utilize variant annotations, such as classification and consequence, to predict the severity of variant consequence. In this work, we quantify the mutational landscape using whole exome sequencing (WES) to identify sequence variants and single-cell RNA sequencing (scRNA-seq) to predict one type of structural variants, CNVs. Sequence variants are annotated using the classification and consequence categorizations, and predicted effect is calculated to determine the overall impact of each mutation. CNVs, since they correspond to genomic regions rather than specific genes, were not subjected to predicted effect calculations. Together, these analyses provide sufficient coverage of the genomic landscape, insofar as their relationship to phenotype (i.e. DIP rate).

Epigenetic (Level 2):

Conceptualized as a quasi-potential energy surface where local minima, or "basins of attraction," correspond to cellular phenotypes, epigenetic landscapes were first proposed by Waddington as an abstract tool for understanding phenotypic plasticity and cellular differentiation during development[84]. The genetic state of a cell sets the topography of the landscape[24,85] and intrinsic (e.g., gene expression[61])

or extrinsic[28,86] sources of noise drive transitions between phenotypes. Phenotypic state transitions (between epigenetic basins) have been observed to occur on the order of hours to weeks[21,36], meaning epigenetic diversification occurs on a much faster timescale than genetic diversification. From a molecular perspective, the epigenetic landscape is the consequence of the complex biochemical interaction networks underlying cellular function[87,88]. Complex dynamical networks can harbor multiple stable states, termed "attractors", towards which a system will tend to return (relax back) in response to small perturbations[89,90]. This property underlies epigenetic heritability: daughter cells inherit similar molecular contents to their parent[91]; hence, they tend to remain within the region of influence of the parental attractor[92]. Larger perturbations, however, can move a system into the region of influence of a neighboring attractor, resulting in a spontaneous phenotypic state change. This explains why epigenetic inheritance is often short lived[93,94].

From a molecular standpoint, the Waddington epigenetic landscape is a consequence of the complex biochemical interaction networks underlying cellular function[87,88]. An "epigenetic state" within this definition is thus simply the molecular state of the associated attractor. Here, we use the transcriptional state of a cell, revealed through single-cell transcriptomics, as a proxy for the epigenetic state, with the understanding that mRNA is only one lens through which to view epigenetics. Indeed, this definition of an epigenetic state, i.e., as a stable state of a complex biochemical network, differs from the traditional molecular biology definition in terms of "epigenetic marks"[95] (e.g., DNA methylation, regions of open chromatin). The two are related, however, in the sense that the same biochemical network that defines the epigenetic landscape also sets the epigenetic marks on the DNA[60]. It is important to recognize that each modality serves only as one possible view, i.e., a proxy, for the epigenetic state. In molecular biology, for example, epigenetic states are usually discussed in terms of chromosomal changes[96], e.g., DNA methylation and histone modification. Systems biologists have often utilized single-cell transcriptomics to identify epigenetic states, where clusters correspond to stable network states. Both

epigenomics and transcriptomics views of the epigenetic landscape are utilized in this work, and can simply be seen as different views of the same epigenetic states since the biochemical network that defines the Waddington landscape also regulates the epigenetic marks on the DNA[60].

Stochastic (Level 3):

Fluctuations, or noise, in intracellular species concentrations have long been recognized as a source of non-genetic heterogeneity in isogenic cell populations, first in bacteria[38,76] and then in yeast[97,98] and mammalian cells[21,99,100]. Intracellular noise can be "intrinsic", i.e., due to the probabilistic nature of biochemical interactions[101], or "extrinsic"[102], affecting the rates of interactions, synthesis, and degradation within a biochemical network. Intrinsic sources of noise include transcriptional bursting[103,104], translational bursting[105], and randomness in mRNA/protein degradation[106], oligomerization[107], and post-translational modification[108]. Extrinsic noise includes randomness in the distribution of molecular contents upon cell division[67,109], environmental factors such as inhomogeneities in cell culture media[110], fluctuations in temperature[111] and pH[112], and spatial variations in the microenvironment[113]. Importantly, fluctuations at the molecular level can drive probabilistic cell fate decisions[61,62,114–116], including division and death[117,118], at the single-cell level and phenotypic diversification at the epigenetic (population) level[99,119]. Experimentally, at the intracellular level, intrinsic and extrinsic noise are difficult to distinguish, generally requiring multiple fluorescent reporters and the ability to fine-tune the external environment[102]. Theoretically, the chemical Master Equation[120,121] (CME) is the construct upon which stochastic dynamical analyses are based. Since the CME is difficult to solve in general, a number of stochastic algorithms have been developed for simulating fluctuations at both the single cell and cell population levels[101,122].

In this work, SSA simulations of population dynamics are utilized to quantify stochastic variability in epigenetic states, with stochastic cell division and death used to model intrinsic noise associated with cell fate decisions. This stochasticity is analogous to *transcriptional heterogeneity* noted in various

systems, where cells exhibit small variations (shown by techniques such as single-molecule RNA-FISH) that do not correspond to a new steady state of the system (i.e. epigenetic state)[93]. Importantly, intrinsic stochasticity is *not* thought to be inherited between generations, which further distinguishes it from epigenetic heterogeneity. However, intrinsic stochasticity is thought to play a role in cellular plasticity, reversibly transitioning cells between basins in the epigenetic landscape.

Together, we utilize this *three-tiered view* of tumor heterogeneity (FIG. 2) to disentangle genetic, epigenetic, and stochastic sources of variability in cancer. Typical tumors comprise numerous genetic states and are thus expected to harbor numerous overlapping epigenetic landscapes, each of which is subject to noise-induced phenotypic transitions. Note that the epigenetic landscape was originally devised as a way to explain cellular differentiation during development. In fact, a developmental hierarchy is a special type of epigenetic landscape, where successive basins are lower in quasi-potential energy, causing cells to descend from one state to the next. In cancer, rather than a well-defined hierarchy, multiple basins of comparable depth coexist. A population of cancer cells is thus expected to spread out across these basins, resulting in a highly heterogeneous population. This may confer a survival benefit to the population in the face of future stressors, such as drug treatments.

An obvious complication to this view is the overlap between levels of heterogeneity. Cellular variation does not happen in independent level vacuums, and variations at each level may have effects on other levels. For instance, genetic mutations often have an effect on transcription[123], which is reflected in modified epigenetic landscapes. The next section attempts to better understand the genetic-to-epigenetic connection. However, there are also cases where epigenetic variation, such as open regions of chromatin, can have an effect on the likelihood of a mutation[124]. In these cases, there would be feedback and dependencies between the genetic and epigenetic levels that complicate our view of cancer heterogeneity. Stochasticity, as a property of biological systems, does not inherently have an effect on

other levels, but can induce different influences on specific epigenetic landscapes, which are defined by a single genetic state. For example, one genetic state may emanate an epigenetic landscape with low barrier heights between basins, while another may have deep basins. In the landscape with low barrier heights, stochasticity will have a greater effect because transitions are more likely to traverse cells between basins, a property known as epigenetic plasticity. A landscape with deeper basins will take longer for cells to transition between basins and be reflected as a more stable landscape. Regardless of how one views these levels, it is clear that we cannot currently quantify all of the complexity associated with the overlap between levels. From a theoretical perspective, we can expect that most tumors are composed of multiple genetic states, and therefore many overlapping epigenetic landscapes. This work provides a first step to disentangle the web of complexity found in heterogeneous tumors.

The Genetic-to-Epigenetic Connection

In the genetic/epigenetic/stochastic view of tumor heterogeneity that we utilize in this work (FIG. 2; see also *"Genetic, epigenetic, and stochastic sources of tumor heterogeneity"* above), genetics are fundamentally tied to epigenetics. To understand how, recall that an epigenetic landscape quantifies, in terms of a quasi-potential energy, the accessible molecular states that cells can occupy. The topography of the landscape depends on the molecular species present (the axes of the state-space), the biochemical processes that can occur, and the *parameters* that quantify the rates at which these processes proceed. The latter means that changes to the rate parameters can change the topography of the landscape, i.e., the depths of basins, heights of barriers, etc. For processes such as protein-protein interactions, the rate parameters are directly dependent on protein structure, which is encoded in the DNA. Thus, mutations to the DNA that, e.g., alter the accessibility of a binding domain, change the values of rate parameters and, hence, the epigenetic landscape as a whole. In this way, each genetic state can be thought of as having an associated epigenetic landscape, which may have multiple basins, or cell phenotypes.

A consequence of this final point is that phenotypic differences at the transcriptomic level do not always indicate differences at the genetic level. In other words, cells can be genetically identical yet differ transcriptomically if they occupy different basins within a common epigenetic landscape. In theory, we should be able to distinguish between cases in which transcriptomic differences are tied to genetic variations and those in which they are not by simply comparing across samples parts of the genome that are relevant to the phenotype of interest (e.g., drug response). However, in practice we do not have a full accounting of all phenotypically-relevant parts of the genome, we only have the full genomic sequences, which will invariably have differences. The question is whether those differences are functionally relevant to the phenotype of interest (FIG. 3). If not, we can consider the samples to be genetically identical, even though they are genomically distinct (see *"Genetic, epigenetic, and stochastic sources of tumor heterogeneity: Genetic"* above). The functional relevance of genomic mutations makes comparison across the levels of heterogeneity more difficult, and we utilize approaches in this work in an attempt to identify functional relevance.

Understanding Drug Response Phenotypes with the DIP Rate

Cellular phenotypes are directly tied to context. In the case of cancer therapies, the most apt measure of phenotype is drug sensitivity. Traditional metrics of drug sensitivity, such as cell viability[125] and clonogenic assays[126], have minimal resolution and have a variety of problems and biases that make them poor prognosticators of phenotype. To combat these insufficiencies, the laboratory previously created the drug-induced proliferation (DIP) rate[117,118], an unbiased metric to quantify drug effect *in vitro*. This metric measures the stable cell proliferation rate of a cell population achieved during prolonged treatment, similar to traditional cell viability assays but designed to capture data across the drug response time series, and thus, more accurate data. However, the metric only captures the stable proliferation rate of the *population*, which corresponds to the most fit clones in the treated population. Therefore, to

supplement the standard DIP rate assay (indicated hereafter as population-level), we developed a single-cell DIP rate assay. Originally published as the clonal fractional proliferation (cFP) assay[127], this approach tracks the DIP rates of many single-cell derived colonies of a cell population to generate a distribution of rates. This method captures colony proliferation in a much more quantitative way than the clonogenic assay, and removes some of the subjective aspects such as influence of other cell types (microenvironment, signaling, etc.) on clonal proliferation. The resulting single-cell distribution of DIP rates, specifically the variance of the distribution, shows the extent of drug-response heterogeneity present in a population. Both the population-level and single-cell assays quantify cellular fitness in a way that can be tied to phenotype.

The DIP rate is rooted in theory, derived from the probabilities of cell division and death (DIP rate =~ division rate – death rate). As such, population dynamics models of cell proliferation can provide further insight into cell division and death in drug treatment (see "*Genetic, Epigenetic, and Stochastic Sources of Tumor Heterogeneity: Stochastic"*), key components of phenotype, as defined by the DIP rate. As we expand to multiple cell phenotypes, the DIP rate can be used to identify phenotypic heterogeneity, corresponding to states with distinguishable DIP rates, each with distinct division and death rate probabilities. Transitions between cellular states can be measured by changing DIP rates over time, an important aspect of cellular plasticity. Together, the DIP rate provides a metric to understand drug response dynamics reflective of various biological processes.

Building *in vitro* Models to Mimic Tumor Variability

To provide experimental evidence for heterogeneity, an experimental system was needed that could explain both an independent level of heterogeneity and other levels dependent upon it (if necessary). Conversely, another experimental system was required to explain plasticity, both in the

absence or presence of perturbations (i.e. drugs). With this in mind, two separate systems were established for heterogeneity and plasticity.

For the heterogeneity model, a *family* of cell line 'versions' and 'sublines' that could capture different levels of heterogeneity were generated (FIG. 4). This cell line family can be viewed as a proxy for a heterogeneous tumor composed of multiple genetic states, each of which has multiple epigenetic states, within which stochastic fluctuations occur and occasionally drive phenotypic state changes. Based on these needs, NSCLC cell line PC9[12,13,21] was identified as a promising candidate to quantify heterogeneity. In addition to fulfilling the requirements above, PC9 has a mutation (exon 19 deletion) in the EGFR gene that makes it sensitive to EGFRi[128,129], a common treatment for NSCLC. Consequently, response to EGFR inhibition (i.e. DIP rate in EGFRi) was a natural way to identify phenotypic changes, which we measure in terms of drug response to EGFRi erlotinib. We first identified two versions of the PC9 cell line cultivated at different academic institutions (Massachusetts General Hospital–MGH[13,21,129]; Vanderbilt University–VU[118,130]) in an effort to capture genetic heterogeneity. The extended period of separate evolution in cell culture likely allowed for extensive genomic evolution. PC9-BR1, a resistant cell line derived from PC9-VU through dose-escalation therapy in EGFRi afatinib[130], would serve as a positive control for genetic heterogeneity. Multiple single-cell derived clones of cell line version PC9-VU were then derived to avoid genetic heterogeneity completely, designed to identify individual epigenetic basins in the underlying PC9-VU epigenetic landscape. Variability within the sublines themselves would be utilized to understand stochastic variability. In this way, a system was built to capture multiple levels of heterogeneity: genetic, epigenetic, and stochastic.

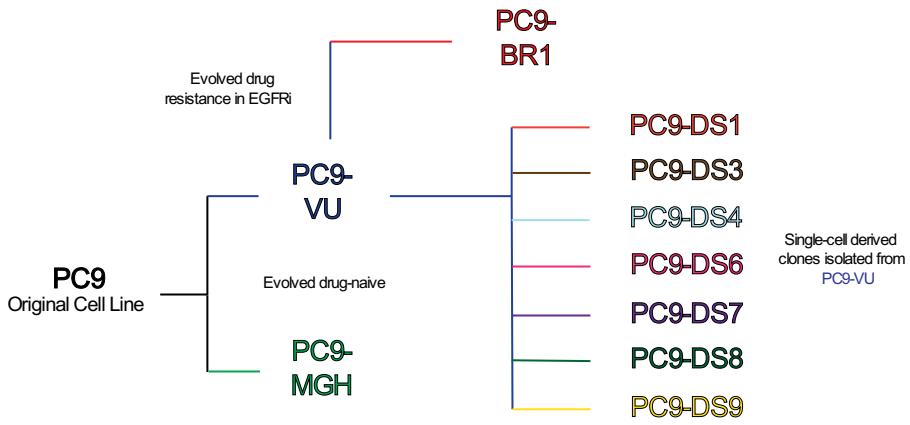

Figure 4 | *The PC9 cell line family tree.*
Two versions the PC9 cell line were maintained separately in culture at two different institutions (VU – Vanderbilt University; MGH – Massachusetts General Hospital). A resistant cell line (PC9-BR1) was derived from PC9-VU by dose escalation in the EGFR inhibitor afatinib. Several discrete sublines (DS) were also single-cell isolated from PC9-VU. Colors are consistent with data visualizations.

In the case of plasticity, a model system was needed that would not acquire functional genetic mutations across short time scales (i.e. weeks), but still exhibit epigenetic heterogeneity and plasticity. BRAF-mutant melanoma became an interesting option, as it is characterized by multiple epigenetic states, and known to show epigenetic adaptation to treatments[35,131,132] (i.e. multiple shallow basins in a common epigenetic landscape). BRAF-mutant melanoma cell line SKMEL5 was chosen as a model system of plasticity[36,133,134]. Like PC9, SKMEL5 harbors a mutation in an oncogene, $BRAF^{V600E}$, that makes it susceptible to BRAF inhibition (BRAFi). As such, BRAFi could be used to quantify phenotypic differences among single-cell derived clonal derivatives (i.e. epigenetic states; distinct short-term DIP rates) and plasticity in the population itself (i.e., changing DIP rates). Therefore, this system was designed to complement the processes that PC9 could not, namely epigenetic plasticity in the presence or absence of drug treatment.

Chapter Motivation

Variability plays a key role in cancer treatment insensitivity and eventual resistance[9,11]. The vast majority of studies consider cancer variability as a monolith, chalking up poor patient outcomes to pre-existing or acquired genetic mutations. But, cancer variability is itself variable. Broadly speaking, cancer variability can be further segmented into heterogeneity, a snapshot of variability at a single point in time, and plasticity, variability as it evolves over time[18,131]. Additionally, variability can be derived from different sources, including (but not limited to) genetics[48,49], epigenetics[58,135,136], stochastic fluctuations[38,99,137], environmental[44] and pharmacologic effects[29], and many more[24]. Given these sources, identifying and tracking cancer variability is a difficult task. Currently, seminal studies aim to perform as many types of experiments as possible on patient-derived material in hopes of finding the magic bullet to cure cancer[68,132,138]. However, we must be open to the idea that there are multiple needles in the proverbial haystack, with different levels of importance at different times. Established methods to find these needles and interpret how they work together are only in the infancy stage[92,139]. With such a wide range of data modalities and potential findings, a clarifying lens is needed to help researchers sort through the haystack. This work uses *in vitro* model systems of heterogeneity and plasticity to study cancer variability across multiple sources. By combining quantitative experimentation, bioinformatic analyses, and mathematical modeling, it shows how the iterative loop of systems biology can create new insights for old problems. It highlights the staggering amount of complexity present in these systems, and creates a high-level view for how variability contributes to numerous treatment outcomes in cancer. Although we do not fully understand the implications of all types of variability on cancer disease progression, this work is a step to quantify the disease in nontraditional ways, and use that knowledge to create novel treatments.

Differentiation between types of tumor variability is crucial to develop novel therapy regimens to combat tumor recurrence. While many studies have aimed to quantify individual aspects of tumor variability, we hypothesized that a common heterogeneity framework, paired with the right experimental

systems, could shed light on how current treatment paradigms are failing to eradicate the disease. Chapter II uses an *in vitro* model of tumor heterogeneity to disentangle various sources of tumor heterogeneity: genetic, epigenetic, and stochastic. Using extensive drug response profiling, followed by genomic and transcriptomic characterization and mathematical population dynamics modeling, we verify that the versions a commonly used NSCLC cell line are genetically distinct and argue that all but one of recently single-cell derived sublines are genetically indistinct but differ epigenetically, i.e., they correspond to basins within a common epigenetic landscape. We detail one case where our analyses suggest that a subline harbors a distinct genetic state that appears to have emerged in the absence of selective pressures. Chapter III identifies the existence of a novel population 'idling' state in BRAF-mutant melanoma, achieved through phenotypic plasticity in response to drug treatment. We show that the idling state is characterized by a net-zero proliferation state in BRAF inhibition (BRAFi), achieved by various clones with differential short-term (<100 hours) proliferation rates but eventually adopt the net-zero rate. A mathematical population dynamics model, fit to drug-response data, suggests that idling is a *population state*, meaning that multiple states co-exist to reflect the net-zero proliferation rate in BRAFi. Furthermore, we show that epigenetic plasticity is not limited to drug treatment, that cell populations diversify in the absence of selective pressure that change drug response phenotypes. In Chapter IV, we characterize the idling cells using various single-cell technologies to determine exactly how BRAF-mutant melanoma cells achieve the population state. We observe that idling is a property of nearly all BRAFi treated lineages, which adopt a new convergent epigenetic landscape over which the population equilibrates after treatment with BRAFi. This shift is tied to ion channel signaling and membrane potential imbalance, which can be targeted to enhance idling cell death. Ultimately, our results highlight the need to understand tumor heterogeneity and plasticity, which both contribute to poor patient outcomes and the inability to win the war on cancer.

CHAPTER 2

Disentangling Genetic, Epigenetic, and Stochastic Sources of Cell State Variability in an *in vitro* Model of Tumor Heterogeneity[1]

Introduction

Tumor heterogeneity is a primary cause of treatment failure and acquired resistance in cancer patients. Even in cancers driven by a single mutated oncogene, variability in response to targeted therapies is well known. The existence of additional genomic alterations among tumor cells can only partially explain this variability. As such, non-genetic factors are increasingly seen as critical contributors to tumor relapse and acquired resistance in cancer. Here, we show that both genetic and non-genetic factors contribute to targeted drug-response variability in an experimental model of tumor heterogeneity. We observe significant variability to EGFR inhibition among and within multiple versions and clonal sublines of PC9, a commonly used EGFR-mutant non-small cell lung cancer cell line. We disentangle genetic, epigenetic, and stochastic components of this variability using a theoretical framework in which distinct genetic states give rise to multiple epigenetic "basins of attraction", across which cells can transition driven by stochastic noise. Using mutational impact analysis, single-cell differential gene expression, and correlations among gene ontology terms to connect genomics to transcriptomics, we establish a baseline for genetic differences driving drug-response variability among PC9 cell line versions. Applying the same approach to clonal sublines, we conclude that drug-response variability in all but one of the sublines is due to epigenetic differences; in the other, it is due to genetic alterations. Finally, using a clonal drug-response assay together with stochastic simulations, we attribute subclonal drug-response

[1]Adapted with permission from Hayford *et al.* Disentangling genetic, epigenetic, and stochastic sources of cell state variability in an *in vitro* model of tumor heterogeneity *In review* (2021).

variability within sublines to stochastic cell fate decisions and confirm that one subline likely contains genetic resistance mutations that emerged in the absence of selective pressures.

Results

Cell line versions and single cell-derived sublines exhibit drug-response variability at the cell population level

We chose commonly-used NSCLC cell line PC9[128] as a model system for tumor heterogeneity. The PC9 cell line is characterized by an EGFR-ex19del mutation, making it sensitive to inhibition of the mutant EGFR protein. We utilize three versions of the cell line: PC9-VU, originating from Vanderbilt University[130]; PC9-MGH, maintained at Massachusetts General Hospital[21,129]; and PC9-BR1, derived from PC9-VU and containing a known secondary resistance mutation (EGFR-T790M) obtained through dose escalation therapy in the EGFR inhibitor (EGFRi) afatinib[130]. Although it is unclear when the PC9-VU and PC9-MGH versions (originating from a common founder cell population) were independently established (FIG. 4), both maintain the oncogenic mutation in the EGFR gene and display sensitivity to EGFR inhibition[21,118]. In the absence of drug, PC9-VU and PC9-BR1 have essentially identical proliferation rates, while PC9-MGH grows at a slightly lower rate (FIG. 5A). However, in response to the EGFRi erlotinib, the three cell line versions display drastically different drug sensitivities (FIG. 5A): PC9-MGH exhibits substantial cell death after an initial equilibration phase (~72h), PC9-VU settles into a near-zero rate of growth, and PC9-BR1 displays insensitivity to EGFRi (as expected). These observations are consistent with the high sensitivity of PC9-MGH to erlotinib reported in Sharma et al.[21] and the lower sensitivity of PC9-VU that we reported previously[117,118].

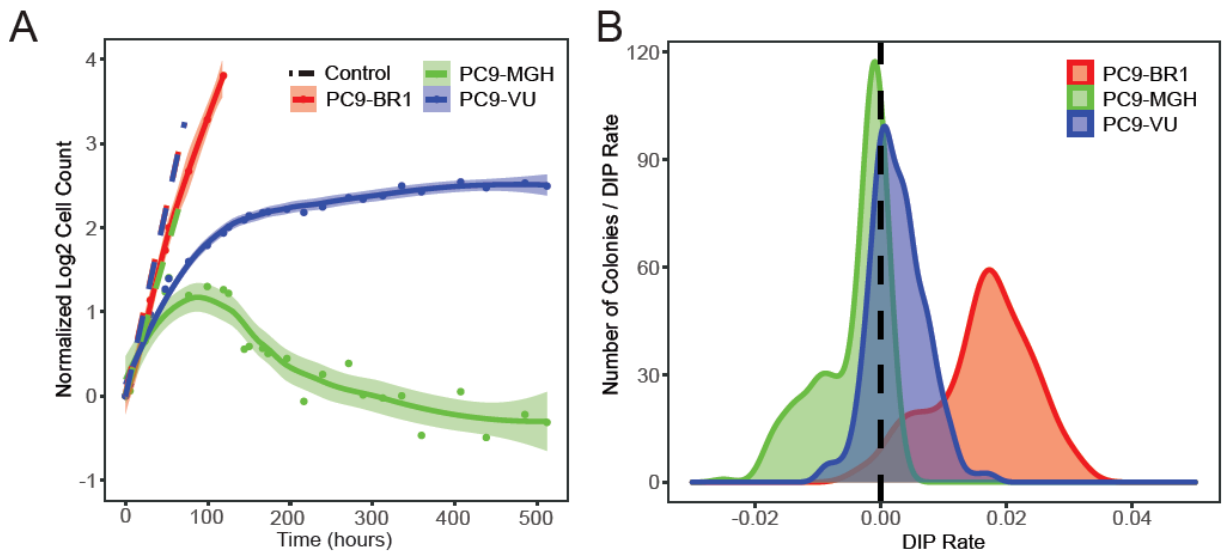

Figure 5 | *Phenotypic differences among PC9 cell line versions quantified in terms of drug response.*
(A) Population growth curves for three cell line versions treated with 3μM erlotinib for approximately three weeks, plus vehicle (DMSO) control. (B) Drug-induced proliferation (DIP) rate distributions compiled from single-colony growth trajectories under erlotinib treatment (3uM) in a clonal Fractional Proliferation (cFP) assay. DIP rates are calculated from the growth curves 48h post-drug addition to the end of the experiment. Dashed black lines signify zero DIP rate, for visual orientation. In *A*, dots are the means of six experimental replicates at each time point; solid lines are best fits to the drug response trajectories with point-wise 95% confidence intervals. In *B*, DIP rate distributions are plotted as kernel density estimates.

We also quantified clonal drug response variability within the cell line versions using clonal Fractional Proliferation[127] (cFP), an assay that tracks the growth of many single cell-derived colonies over time and quantifies drug sensitivity for each colony in terms of the drug-induced proliferation (DIP) rate[117,118], defined as the stable rate of proliferation achieved after extended drug exposure (see Methods). We performed cFP on each cell line version under erlotinib treatment and observed wide ranges of drug responses across colonies and substantial differences in the response distributions across versions (FIG. 5B). The distribution of DIP rates for PC9-BR1 lies almost entirely in the positive DIP rate range and is clearly distinct from the others. The PC9-VU and PC9-MGH distributions have significant overlap but the PC9-MGH distribution has a marked shoulder in the negative DIP rate region, while the PC9-VU distribution extends further into the positive range. These distributions are consistent with and

explain the differential drug responses observed among the cell line versions (FIG. 5A): PC9-BR1 is resistant to EGFRi because its DIP rate distribution is entirely in the positive range, PC9-VU goes into a near-zero (slightly positive) growth phase because its DIP rate distribution is centered near zero, and the large shoulder in the PC9-MGH distribution explains why it exhibits significant cell death in the period immediately following drug treatment.

In addition, several single cell-derived discrete sublines (DS1, DS3, DS4, DS6, DS7, DS8, DS9) were isolated from PC9-VU and subjected to the same analyses as above. In the absence of drug, all sublines grow at almost equal rates in culture (FIG. 6A). However, in the presence of EGFRi, the sublines exhibit a wide range of responses, from positive to negative growth (FIG. 6A). When overlaid with the cFP results for PC9-VU, the subline responses broadly recapitulate the observed variability seen in the parental line. A notable exception is DS8, which is essentially resistant to EGFRi, having only a slightly lower proliferation rate than the fully resistant PC9-BR1 (cf. FIG. 5A). We also performed cFP assays on the sublines under erlotinib treatment (FIG. 6B) to quantify subclonal drug response variability. Interestingly, similar to the cell line versions, we found that the sublines also exhibit distributions of DIP rates, albeit narrower than those for the cell line versions. The subline distributions have a large degree of overlap with one another but the medians of the distributions are statistically distinct ($p < 0.001$, Mood's median test). DS8 is again an exception, exhibiting a bimodal DIP rate distribution with a major mode centered close to zero and a large shoulder in the positive DIP rate range.

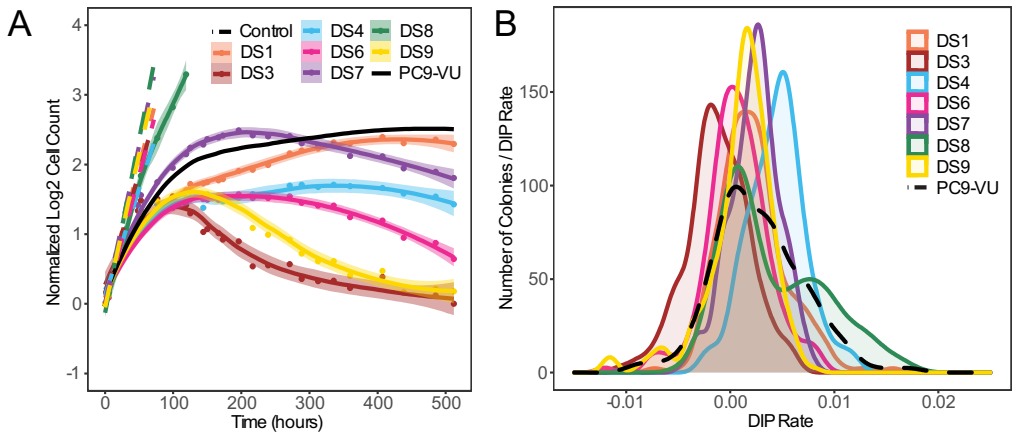

Figure 6 | ***Phenotypic differences among PC9 discrete sublines quantified in terms of drug response.***
(A) Seven discrete sublines (DS) derived from PC9-VU were treated with 3μM erlotinib for three weeks, along with vehicle control. Parental PC9-VU is included for reference. (D) DIP rate distributions from a cFP assay of the sublines in 3μM erlotinib. Parental PC9-VU is included for reference. In *A*, dots are the means of six experimental replicates at each time point; solid lines are best fits to the drug response trajectories with point-wise 95% confidence intervals. In *B*, DIP rate distributions are plotted as kernel density estimates.

Cell line versions differ significantly at the genetic and transcriptomic levels

Given that the PC9-VU and PC9-MGH cell line versions have been maintained separately for many years, it is virtually certain that they differ genetically[26]. We also know that PC9-BR1 contains a known genetic resistance mutation and likely numerous additional mutations acquired during dose escalation. Thus, we perform bulk whole exome sequencing (WES) and single-cell RNA sequencing (scRNA-seq) on the cell line versions in order to establish a benchmark for genetic variation against which we can compare the sublines.

From WES, we identify mutations (SNPs and InDels) in each cell line version that pass a specified threshold for variant detection (see Methods) and calculate the number of mutations per chromosome (FIG. 7A). We see a large amount of variability in the number of called variants between the cell line versions (average coefficient of variation (CV) per chromosome = 12.84). We also identify mutations unique to each cell line version and calculate the proportionality of unique mutations compared to the total number of mutations (FIG. 7B). Although a majority of the mutations are shared between versions,

a significant number are unique: PC9-BR1 has the largest proportional representation of unique mutations, followed by PC9-MGH and then PC9-VU. Furthermore, we annotate unique mutations within each cell line version with an IMPACT score, a variant severity classifier calculated by the Ensembl Variant Effect Predictor[82] (VEP). The IMPACT score differentiates mutations based on a variety of factors that predicts whether a mutation is likely to have a phenotypic effect (see Methods). Categorizing mutations into *low*, *moderate*, and *high* IMPACT score reveals that PC9-BR1 has many more potentially impactful mutations than PC9-MGH and PC9-VU, which have similar numbers to each other (FIG. 7C). However, *as a percentage*, only 1% of PC9-MGH unique mutations are predicted to be impactful, compared to 11% in PC9-VU, suggesting that PC9-MGH harbors a large number of passenger mutations.

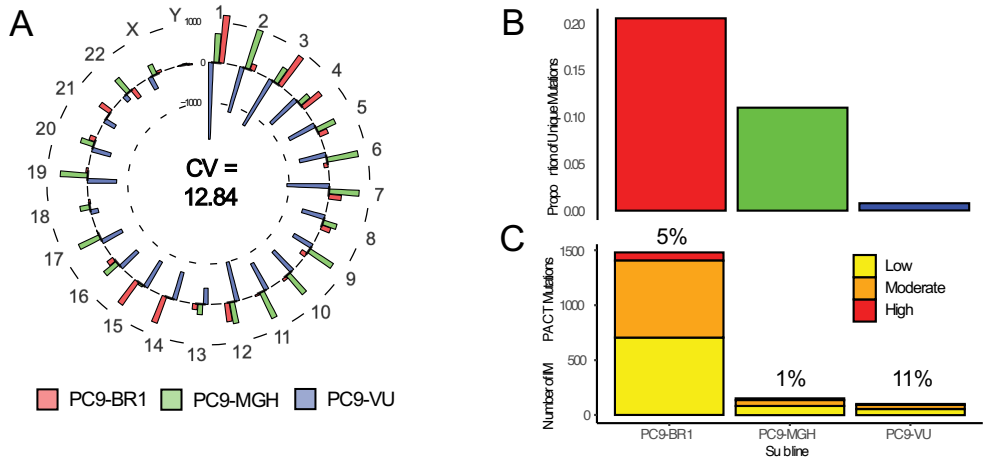

Figure 7 | Genomic characterization of PC9 cell line versions.
(A) Mean-centered mutation count by chromosome for all cell line versions. For each chromosome, versions with fewer mutations than the mean have a bar pointing inwards, while those with more mutations than the mean point outwards. Chromosome numbers are noted on the outside edge of the circle. Average coefficient of variation (CV) across all chromosomes is noted. (B) Proportions of unique mutations for all cell line versions. (C) Numbers of IMPACT mutations unique to each cell line version, stratified by IMPACT classification (*low*, *moderate*, *high*). Percentage of unique IMPACT mutations relative to the total number of unique mutations for each cell line version is denoted above each bar.

We also perform a mutational significance analysis on the unique mutations. Based on nonsynonymous-to-synonymous mutational load, genes are selected to create a mutational signature of

genetic differences within each cell line version (see Methods) and displayed as a heatmap. This signature does not reflect all mutated genes in the cell line versions but rather those of predicted importance, similar to the VEP IMPACT score analysis. We see that many mutations in the signature distinguish the cell line versions (FIG. 8A). Additionally, we generate a literature-curated, cancer-associated gene signature that includes mutations predicted to be implicated in cancer[12,140] (see Methods). Only PC9-BR1, which has a known resistance mutation (EGFR-T790M, noted as a missense mutation in the heatmap), harbors significant mutational load in the cancer-associated gene signature (FIG. 8B). Breakdowns of mutations by type provide additional evidence that PC9-VU mutation representation is distinct and proportionally impactful (FIG. 8C).

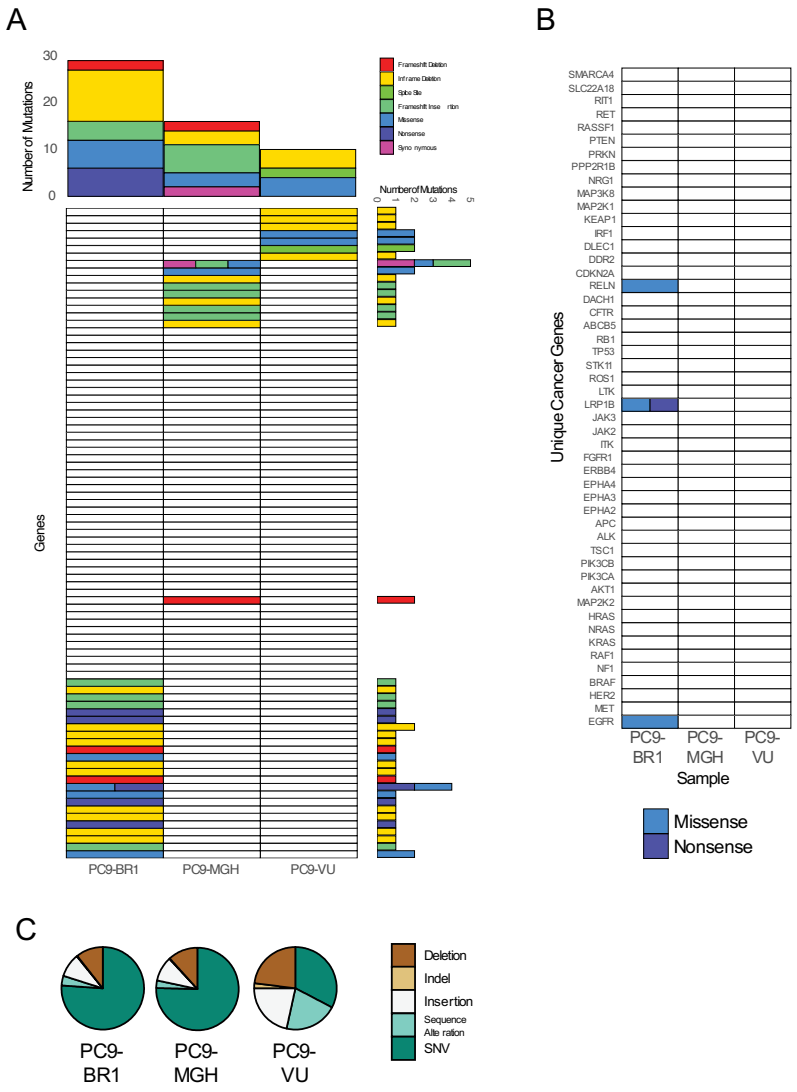

Figure 8 | Mutational significance of PC9 cell line versions.
(A) Quantification of mutational differences between cell line versions based on a signature of genes with a high nonsynonymous mutational load. Heatmap elements are colored based on type of mutation (see Methods). Total numbers of mutations (stratified by mutation type) across genes and cell line versions are shown as bar plots to the right and above of the heatmap, respectively. (B) Mutational differences between cell line versions for a literature-curated set of cancer-associated genes implicated in lung cancer (see Methods). Heatmap elements are colored based on type of mutation. (C) Mutation class pie charts. SNV: single nucleotide variant; indel: combination of insertion and deletion.

In addition to traditional variant analyses, we also perform copy number variant (CNV) detection at the single-cell level. By exploring the gene expression intensity across positions of the genome, CNVs are detected as gains (red) or deletions (blue) in large chromosomal regions, making it clear which regions

of the genome have different relative abundances (see Methods). For the cell line version analysis, the background abundance is an average of the three cell lines. We identity several key regions that differentiate the cell line versions (FIG. 9): PC9-MGH has multiple clear amplifications (chromosomes 6, 11, and 22) and deletions (chromosomes 16 and 19), while PC9-VU and PC9-BR1 are mostly concordant. This result fits with expectation, as PC9-BR1 was derived from PC9-VU and therefore likely had fewer opportunities to acquire large-scale chromosomal changes than PC9-MGH. Together, these genomic data (FIGS. 7-9) starkly illustrate the genetic differences among the cell line versions. Namely, PC9-BR1 is differentiated predominately by single-nucleotide and indel variants and PC9-MGH is differentiated by copy number variants.

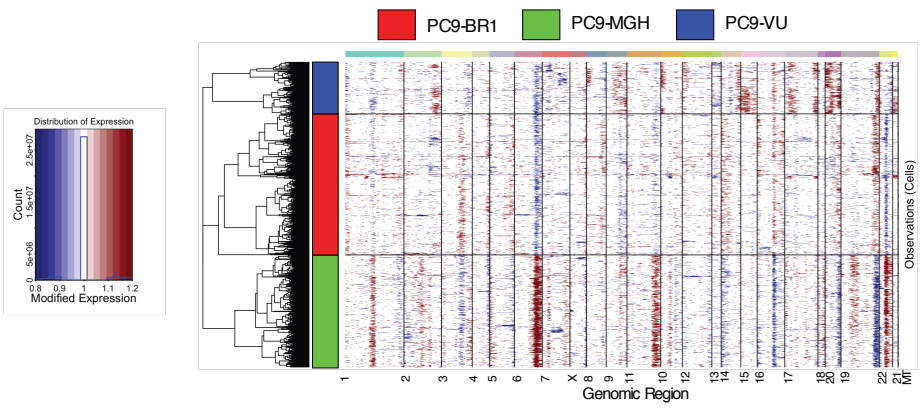

Figure 9 | ***Copy number Variant (CNV) detection for cell line versions.***
Red corresponds to amplifications, blue to deletions.

At the transcriptomic level, we use scRNA-seq to identify gene expression differences among the cell line versions (see Methods). After feature selection, we use Uniform Manifold Approximation and Projection[141,142] (UMAP) to project the transcriptional states for each cell into two-dimensional space (FIG. 10A). We see a clear separation of cell line versions in the UMAP space, with minimal overlap (FIG. 10B-C). Pairwise distances between the centroids of the single-cell clusters (FIG. 10D) show that PC9-VU and

PC9-BR1 are more similar to each other than either is to PC9-MGH, which is unsurprising given that PC9-BR1 was derived from PC9-VU.

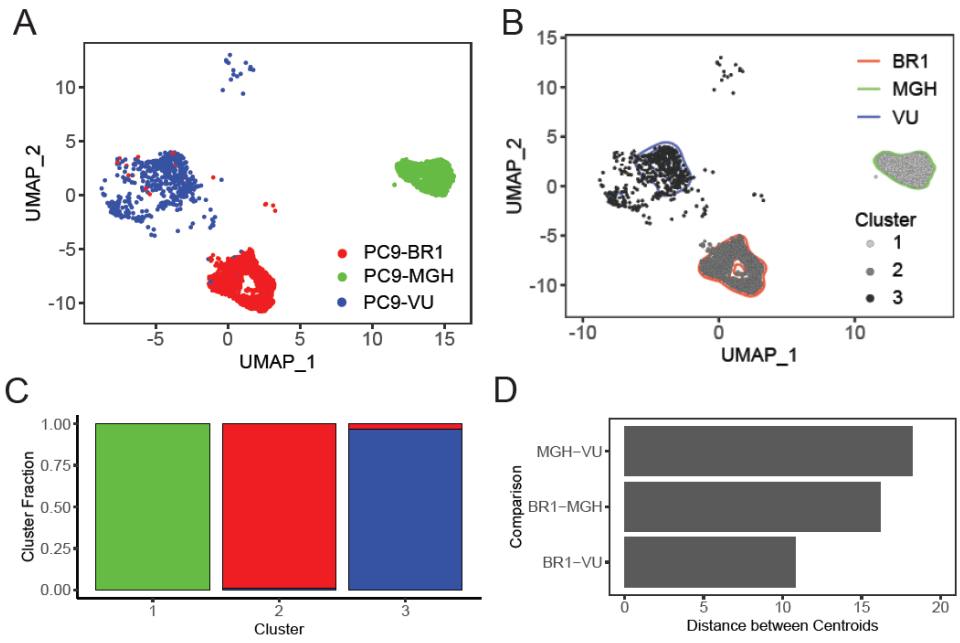

Figure 10 | ***Transcriptomic characterizations of PC9 cell line versions.***
(A) Uniform Manifold Approximation and Projection (UMAP) visualization of single-cell transcriptomes for cell line versions. For comparison purposes, the UMAP space is defined over all eight PC9 samples (including cell line versions and PC9-VU sublines; see Methods). (B) Clustering of cell line versions. Number of clusters (three) was defined based on majority rule from a consensus of 30 indices. Ward's minimum variance method was used. (C) Quantification of cluster fraction by cell line version. (D) Distance between cell line version centroids in UMAP space. Centroids were calculated as the mean across each UMAP axis using a Euclidian distance metric.

To quantify how predictive variations in the genomic states of cell line versions are of differential gene expression at the transcriptomic level, we compare Gene Ontology[143,144] (GO) terms associated with high consequence genetic variants (*low*, *moderate*, and *high* IMPACT scores) and GO terms associated with significantly differentially expressed genes (DEGs; adjusted $p<0.05$). We visualize these terms based on relative statistical significance (negative log of p-value for significant GO terms) and quantify the correlation (Spearman) between the genomics- and transcriptomics-derived terms (FIG. 11). Both PC9-

BR1 and PC9-MGH have a positive correlation between terms, indicating that terms shared between data modalities tend to agree with each other. Obvious exceptions exist that are more statistically significant for one data modality (top-left and bottom-right corners of the plot) but both the existence of terms for both modalities and the moderate correlation indicate the connection. Notably, PC9-VU has a slightly negative correlation.

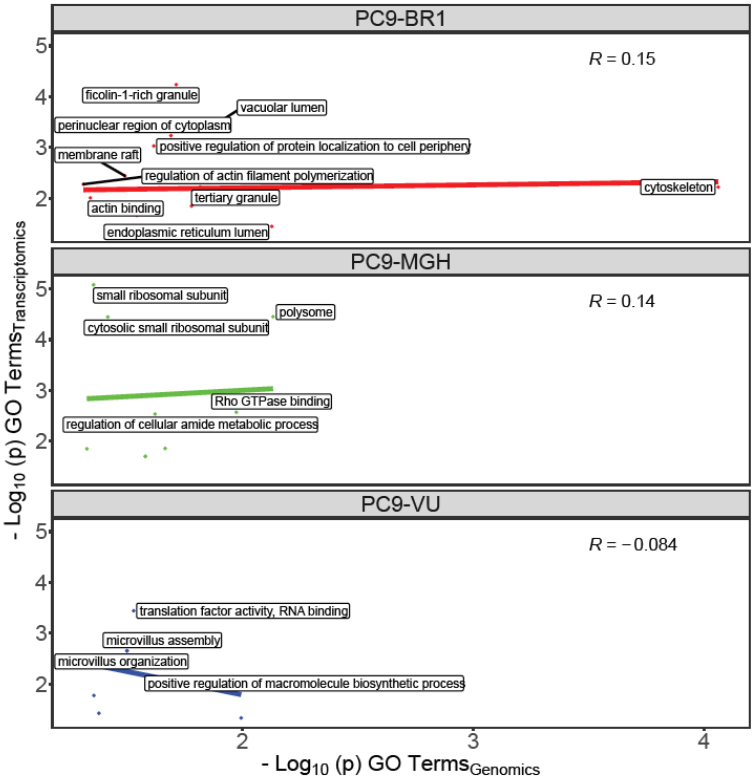

Figure 11 | *Gene Ontology comparison analysis of unique IMPACT mutations and differentially expressed genes (DEGs) for cell line versions.*
A correlation coefficient (Spearman) was calculated for each sample. Terms with a value greater than two on either axis are highlighted.

One PC9-VU subline is genetically distinct, while all others are transcriptomically distinct from each other

We perform the same genomic and transcriptomic analyses as above on five PC9-VU sublines (DS3, DS6, DS7, DS8, DS9) that exhibit differential responses to EGFRi as evidenced by their DIP rate

distributions (see FIG. 6B): DS3 has a peak in the negative DIP rate range, DS6 has a peak close to zero, DS7 and DS9 have peaks in the positive DIP rate range and nearly overlapping distributions, and DS8 stands out as an obvious outlier with a bimodal DIP rate distribution. Analysis of the total numbers of mutations by chromosome (FIG. 12A) shows significantly less variability in the variant count for the sublines relative to the cell line versions (average CV per chromosome = 6.27 vs. 12.84; cf. FIG. 6A). Additionally, unlike the cell line versions, most sublines exhibit similar proportions of unique mutations (FIG. 12B) and numbers of IMPACT mutations (FIG. 12C). The clear exception is DS8, which has more unique mutations and more than twice the number of predicted impactful mutations compared to the other sublines.

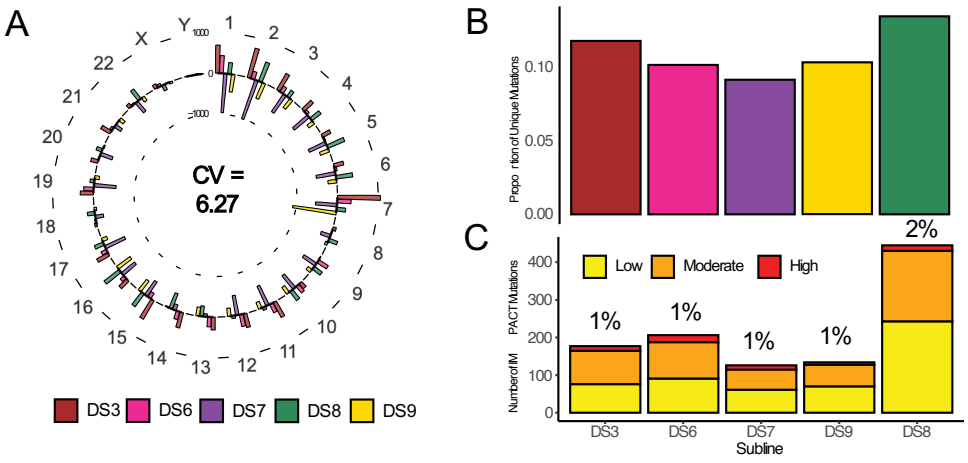

Figure 12 | Genomic characterization of PC9-VU discrete sublines.
(A) Mean-centered mutation count by chromosome for five (of the seven) sublines. Average coefficient of variation across all chromosomes is noted. (B) Proportions of unique mutations in each subline. (C) Numbers of IMPACT mutations unique to each subline, stratified by IMPACT classification (*low*, *moderate*, *high*). Percentage of unique IMPACT mutations relative to the total number of unique mutations in each subline is denoted above each bar.

Mutational significance analysis (using the same genomic signature as for the cell line versions; see FIG. 8A) shows similar numbers of total and impactful mutations in the sublines (FIG. 13A), while DS8 has the largest number of these mutations and most diverse types. This is true for the cancer-associated genes as well (FIG. 13B). The sublines also exhibit a nearly identical mutation class distribution (FIG. 13C).

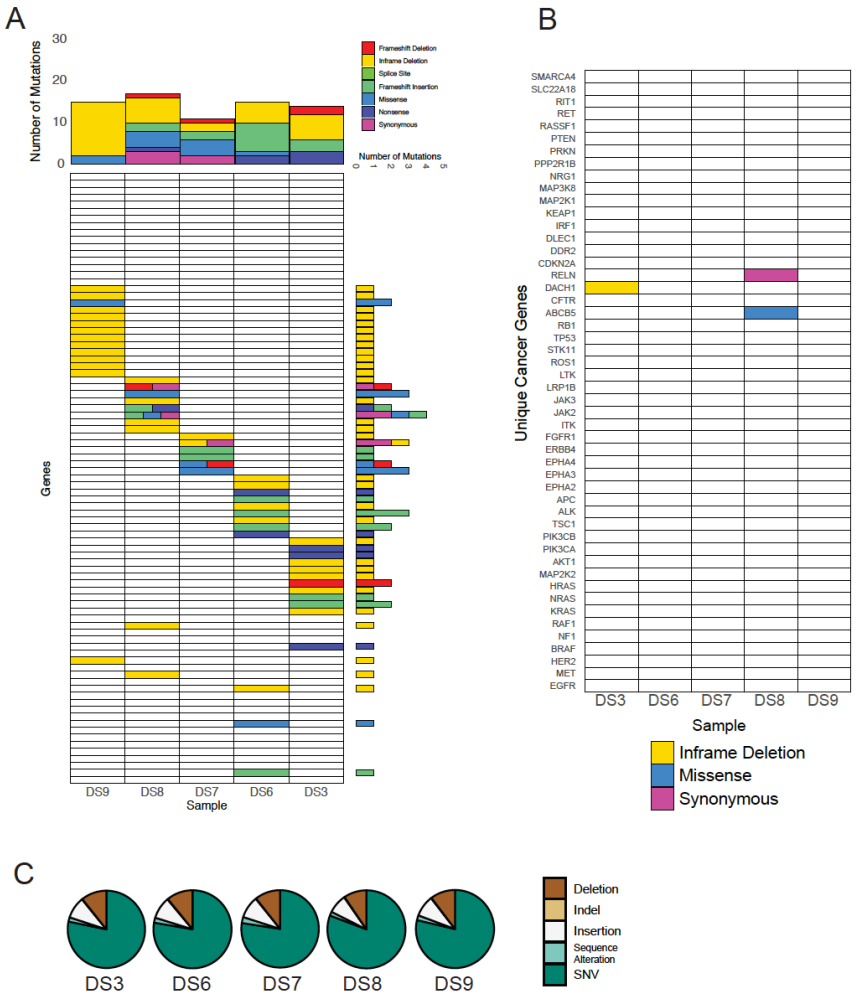

Figure 13 | *Mutational significance of PC9 discrete sublines.*
(A) Quantification of mutational differences between sublines based on gene signature in *FIG. 8A*. Total numbers of mutations (stratified by mutation type) across genes and cell line versions are shown as bar plots to the right and above of the heatmap, respectively. (B) Mutational differences between PC9-VU sublines for the cancer gene signature in *FIG. 8B*. (C) Mutation class pie charts. SNV: single nucleotide variant; indel: combination of insertion and deletion.

Interestingly, CNVs in the sublines show a slightly more nuanced result (FIG. 14): DS8 has major amplifications (chromosomes 6, 17, and 22) and deletions (chromosomes 7, 13, and 22) and some minor alterations not shared with other sublines; DS3 is missing a deletion present in chromosome 7 in DS6, DS7, and DS9 and also has a unique deletion in chromosome 9; and minor additional sharing is seen among

other sublines, such as DS8 and DS9 (deletions in chromosomes 13 and 16). On the whole, DS8 has the clearest cases of unique copy number variations in the sublines, while the other sublines remain largely similar except for a few instances. Taken together, with the exception of DS8, these genomic data (FIGS. 12-14) illustrate that there is significantly less genomic variability among the PC9-VU sublines than among the cell line versions. It is also important to note that DS8 does not harbor the same resistance conferring mutation (EGFR-T790M) that PC9-BR1 does (FIGS. 8B and 13B), indicating a different (unknown) resistance mechanism is at play (see *Chapter V: Discussion*).

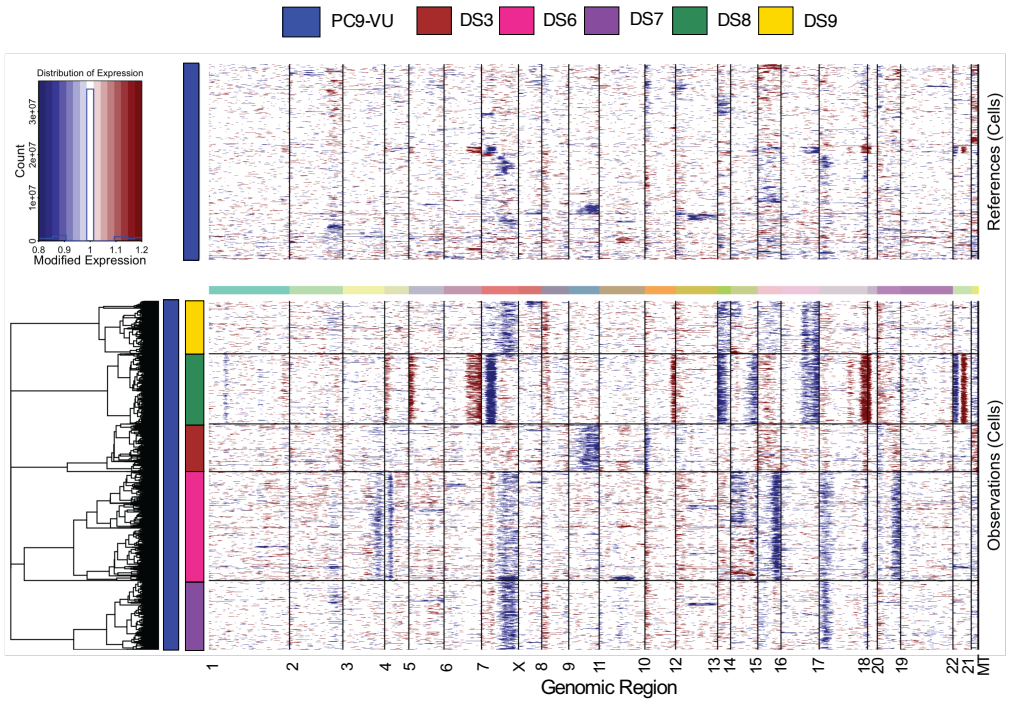

Figure 14 | *Copy number Variant (CNV) detection for discrete sublines.*
Red corresponds to amplifications, blue to deletions.

Comparing single-cell transcriptomes (FIG. 15A), we see distinctions among the sublines but to a much lesser extent than among the cell line versions, except for DS8, which, as in the genomics data, is a clear exception (FIG. 15B-C). Pairwise distances between centroids show virtually no separation between

DS7 and DS9, small but clear separations between DS3, DS6, and DS7/DS9, and a large separation between DS8 and the other sublines (FIG. 15D). We also see that DS3, DS6, DS7, and DS9 substantially overlap with the PC9-VU region of the UMAP space (see blue contour in FIG. 14A). These observations are largely consistent with the clonal drug responses observed in the cFP assays (i.e., the DS7 and DS9 distributions are almost identical, the DS3 and DS6 distributions are distinct from each other and from DS7/DS9, the DS8 distribution stands apart from the others in being bimodal with a large shoulder extending beyond the upper range of the PC9-VU parental distribution, and the DS3, DS6, DS7, and DS9 distributions overlap substantially with the PC9-VU distribution; see FIG. 6B). However, despite the clear separations of the centroids, we do see slight overlaps at the boundaries between the transcriptomic features for DS3 and DS7/DS9 and between DS6 and DS7/DS9, suggesting the possibility of phenotypic state transitions occurring between these states. We also see a very small number of DS9 cells (<2%) that overlap with DS8, which could have many potential explanations but is likely an outlier (see *Chapter V: Discussion*).

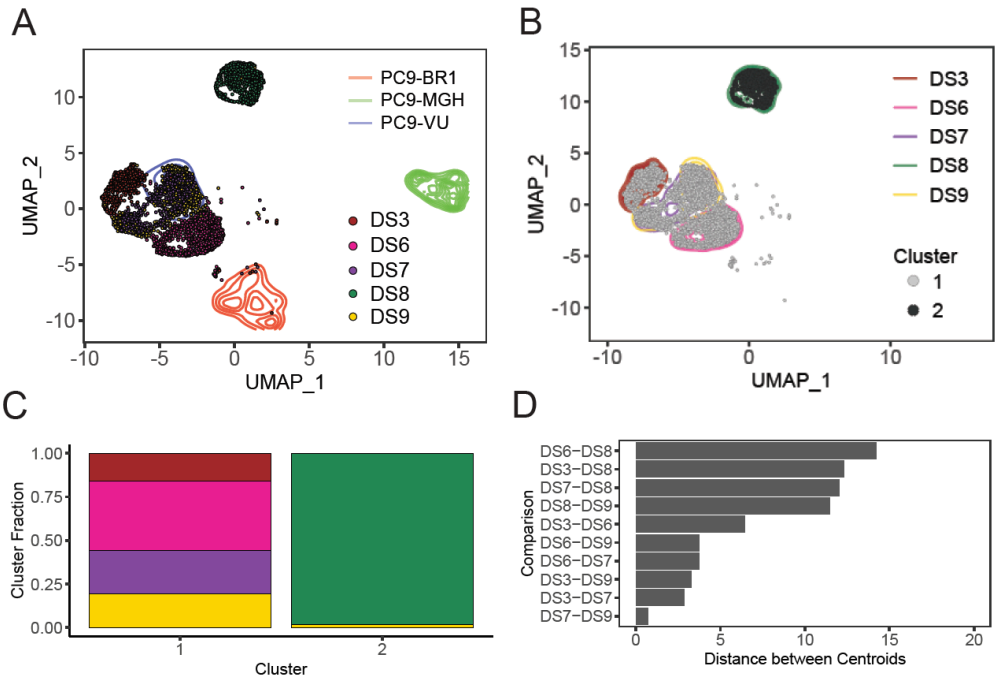

Figure 15 | ***Transcriptomic characterizations of PC9-VU discrete sublines.***
(A) Uniform Manifold Approximation and Projection (UMAP) visualization of single-cell transcriptomes for sublines. For comparison purposes, the UMAP space is defined over all eight PC9 samples (including cell line versions and PC9-VU sublines; see Methods). (B) Clustering of sublines. Number of clusters (two) was defined based on majority rule from a consensus of 30 indices. Ward's minimum variance method was used. (C) Quantification of cluster fraction by subline. (D) Distance between subline centroids in UMAP space. Centroids were calculated as the mean across each UMAP axis using a Euclidian distance metric.

Statistical comparisons between GO terms associated with high consequence genetic variants and DEGs support a connection between genomics and transcriptomics in DS8 but not in the other sublines (FIG. 16). Although DS3, DS6, and DS7 have a few terms significantly enriched and shared between data modalities, there were not enough data points to compute a correlation. DS9 is an interesting case, where many terms were similar between the modalities but showed a negative correlation (more so than PC9-VU; cf. FIG. 11). Taken together, these results suggest a strong connection between genomics and transcriptomics in DS8, a weaker but possibly not insignificant connection in DS9, and likely no connection in the other sublines.

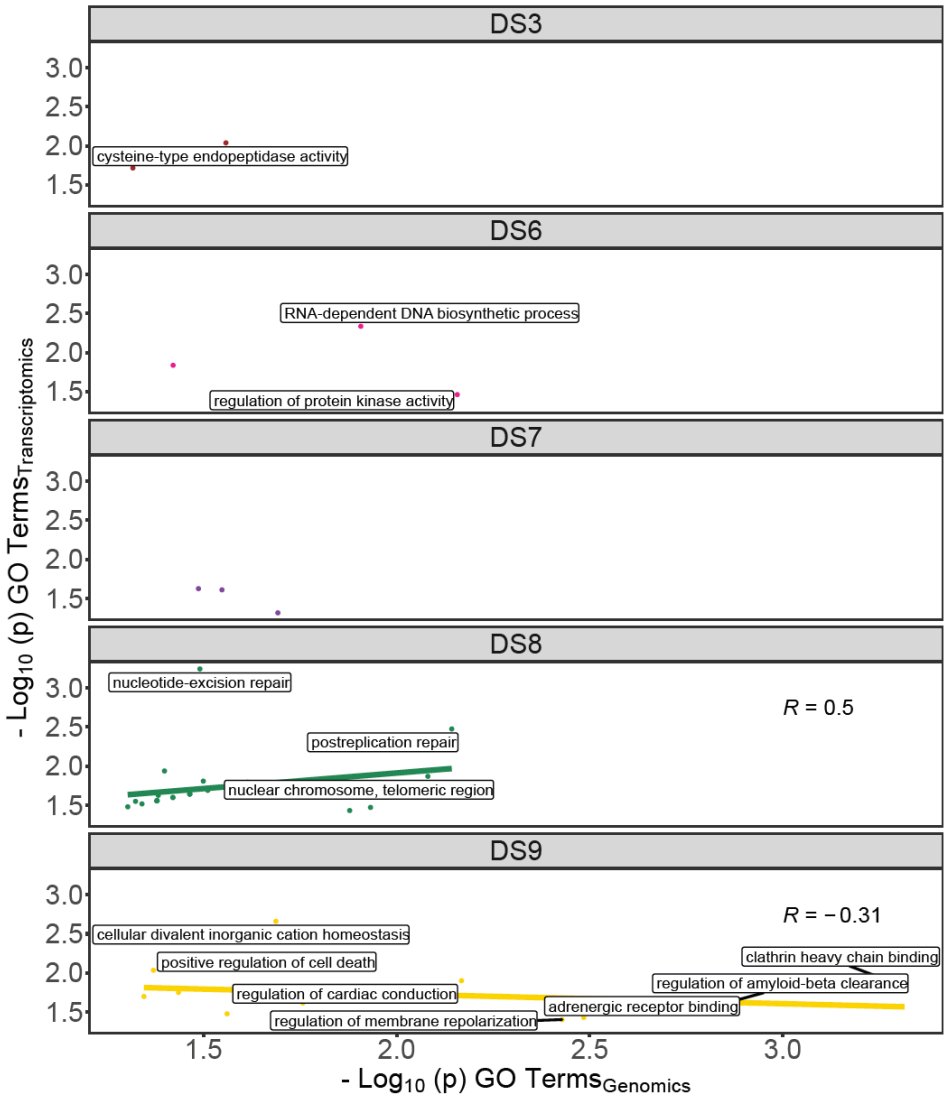

Figure 16 | *Gene Ontology comparison analysis of unique IMPACT mutations and differentially expressed genes (DEGs) for discrete sublines.*
A correlation coefficient (Pearson) was calculated for each sample. Terms with a value greater than two on either axis are highlighted.

Joint analysis of PC9 cell line family members

To validate these results, we perform a series of comparisons on all PC9 cell line family members jointly. Principal component analysis (PCA) on normalized RNA-seq counts (FIG. 17A) is strikingly consistent with scRNA-seq UMAP projections (FIGS. 10A and 15A). PC9-VU and its derivatives group with PC9-BR1 (upper left), while PC9-MGH (right) and DS8 (bottom left) group apart from other PC9 family

members. Hierarchical clustering of gene signatures shows a similar breakdown (FIG. 17B). PC9-VU falls in the same region as its derivatives, as expected. PCA on WES-derived SNPs (FIG. 17C) shows most PC9-VU sublines (DS3, DS6, DS7, DS9) and PC9-BR1 have similar genotypes (top right), while DS8 is distinct (bottom right), as seen in the transcriptomics datasets. Interestingly, PC9-VU and PC9-MGH reside in a similar region of the PCA space (left). An identity-by-state analysis on the same SNPs (FIG. 17D) shows a similar breakdown. Since PC9-MGH seems to differ from PC9-VU predominately by CNVs but shares a common progenitor, this result is consistent with earlier explanations. As shown in prior analyses, DS8 has a distinct genotype that is independent of other PC9 family members, signifying a distinct genetic mechanism of resistance. A surprising result is PC9-BR1 grouping with other sublines in both modalities, considering it differs so greatly from other cell line versions in earlier analyses (FIGS. 7-10). But since this analysis does not consider biological impact, it may indicate the importance of phenotypic context to mutations, as the *single* $EGFR^{T790M}$ mutation has been shown to confer resistance in PC9-BR1[130].

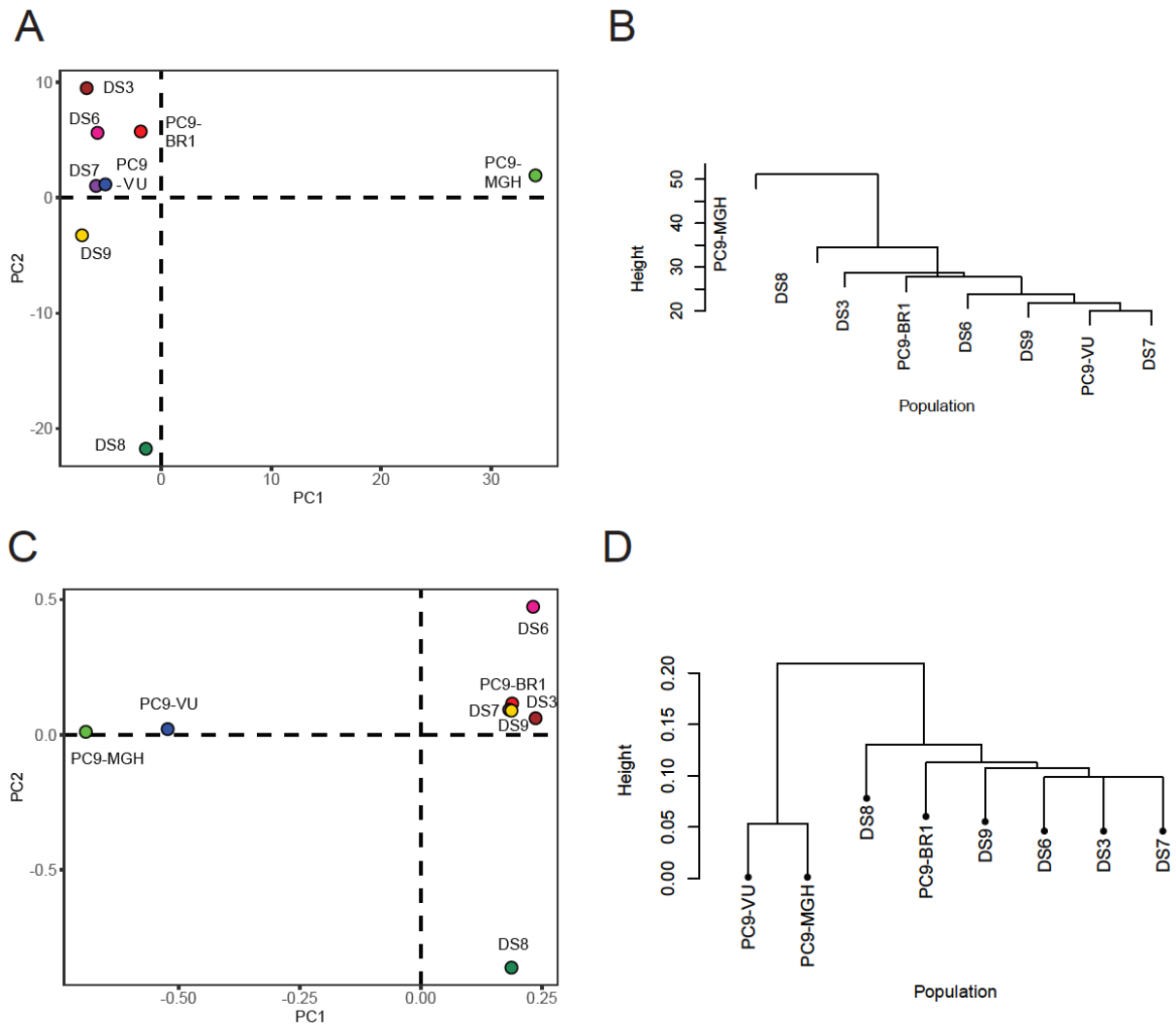

Figure 17 | *Dimensionality reductions and clustering of bulk sequencing datasets.*
(A) Principal Component Analysis (PCA) of single-replicate normalized RNA-seq count data. (B) Hierarchical clustering of RNAseq normalized count data. Clustering was performed on the pairwise Euclidian distance matrix created from the relative log transformed gene counts using the Ward's minimum variance method. (C) PCA of single nucleotide polymorphisms (SNP) genotypes. Using a subset of SNPs in approximate linkage equilibrium, a genetic covariance matrix was calculated. The covariance matrix was converted to a correlation matrix to achieve appropriate scaling and PCA was run to identify SNP eigenvectors (loadings of the principal components). PC9 cell line family members are plotted along the principal component axes. (B) Hierarchical clustering of PC9 genotypes. Using an identity-by-state analysis, a matrix of genome-wide pairwise identities was calculated. Hierarchical clustering was performed on these identities to determine sample relatedness.

To further supplement GO correlation analyses, we modified a semantic similarity metric[145] to compare the two sets of GO terms. For each version, we calculate pairwise similarity scores between GO

terms for genetic variants and DEGs and obtain an aggregate score between 0 and 1, with 1 indicating that differentially expressed genes at the transcriptomic level can be perfectly explained by variations at the genomic level, and vice versa (see Methods). Relative to a randomized baseline, we see elevated semantic similarity scores for PC9-BR1 in the "Biological Process" (BP) GO category and for PC9-VU in the "Molecular Function" (MF) category (FIG. 18). We also see significant semantic similarity scores for all three versions relative to baseline in the "Cellular Component" (CC) category. For the sublines, we see mixed results (FIG. 18): DS8 has high scores for both BP and MF GO categories but not for CC; DS3, DS6, and DS7 have low scores for BP and MF; DS6 also has a low score for CC but DS3 and DS7 have high scores; and DS9 has a high score for the BP category but low scores for the others. Note that based on the number of GO terms in each category (BP: 12,272, MF: 4,165, CC: 1,740), we consider BP to be the most predictive of the three, followed by MF and then CC (see Methods). Taken together, these results (FIGS. 11, 16, and 18) indicate a strong connection between mutations in the genomes of the cell line versions and expression at the transcriptomic level. Moreover, we use these results as benchmarks for ascertaining whether transcriptomic differences seen among single cell-derived sublines are rooted in genetic differences (like in the cell line versions) or are more likely non-genetic in origin. In most of the sublines, their transcriptomic states *cannot* be explained by variations seen at the genomic level, indicating that the transcriptomic differences are epigenetic in origin, not rooted in genetics. An important exception is DS8, which exhibits high scores (as well as a positive GO term correlation; cf. FIG. 16).

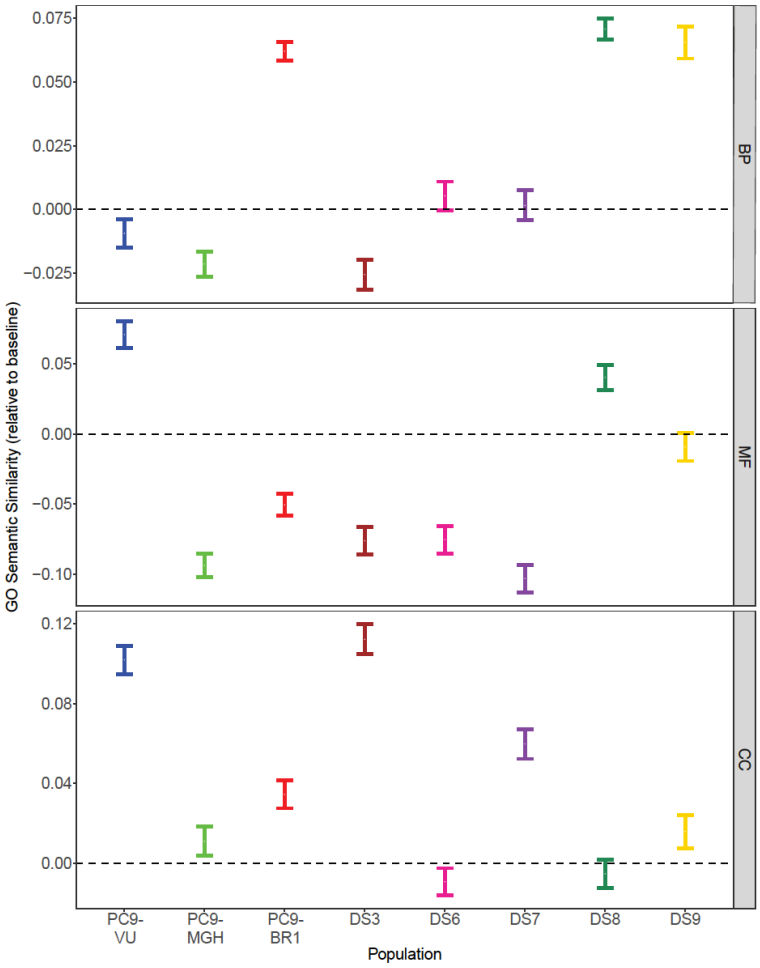

Figure 18 | GO semantic similarity scores for each GO ontology type.
Significantly enriched GO terms for each data modality pair (IMPACT mutations and differentially expressed genes) were compared for each cell line family member for each GO type ("Biological Process" – BP, "Molecular Function" – MF, "Cellular Component" – CC). The top 1000 similarity scores within each pair were compiled into a distribution to calculate a median (white circle) and 95% confidence interval (error bars). Scores are plotted relative to a baseline, defined as the median + one standard deviation of simulated distributions (dashed lines). Simulated score distributions were calculated based on random gene lists of identical lengths to the experimental gene lists (see Methods).

Furthermore, to address the potential biological interpretability of differences across PC9 cell line family members, we score each cell based on 50 hallmark gene signatures of well-defined biological states (see Methods). Many of these processes distinguish cell line versions, such as IL2/STAT5 and KRAS signaling being overexpressed in PC9-MGH, while PI3K/AKT/mTOR signaling has more expression in PC9-VU and PC9-BR1. Interestingly, PC9-BR1 has less expression in Hedgehog signaling than PC9-MGH and

PC9-VU, making it an interesting potential side effect of the increased mutational burden. For the sublines, DS8 has a larger hallmark gene signature score than the other sublines (and cell line versions) for DNA repair, unfolded protein response, and androgen response, while it has a slightly lower score in the p53 pathway. There are no clear cases where other sublines had a significantly larger hallmark gene signature score, with the exception of distribution extremes. Table 1 summarizes the findings of the biological interpretation analysis.

Table 1 | ***VISION transcriptome functional interpretation analysis.***
Single-cell gene expression matrix and MSigDB hallmark gene signatures were input to create a signature score for each cell (see Methods). Scores were totaled for each population across each hallmark and plotted as a density distribution. All 50 hallmark signatures were sampled. Note that "KRAS signaling" and "UV response" had hallmark signatures for both up- and down-regulated. We condensed these four signatures into two, leaving 48 hallmark signatures total.

PC9-MGH	PC9-BR1	DS8	PC9-VU (no DS8)
↑ Angiogenesis	↑ Cholesterol Homeostasis	↑ Allograft Rejection	↑ Hedgehog Signaling
↑ Apical Surface	↓ Hedgehog Signaling	↑ Androgen Response	↓ IL2/STAT5 Signaling
↑ Bile Acid Metabolism	↓ Xenobiotic Metabolism	↑ Complement	↓ WNT/β-catenin Signaling
↑ IL2/STAT5 Signaling		↑ DNA Repair	
↑ KRAS Signaling		↑ Interferon α/γ Response	
↑ WNT/β-catenin Signaling		↑ Unfolded Protein Response	
↓ Allograft Rejection		↓ P53 Pathway	
↓ Coagulation			
↓ Interferon α/γ Response			
↓ Pancreas β Cells			
↓ PI3K/AKT/MTOR Signaling			

Stochastic birth-death simulations suggest most PC9-VU sublines are epigenetically monoclonal, while one is polyclonal

The PC9-VU sublines exhibit variability in drug response, not just at the population level (FIG. 6A) but also at the sub-clonal level, as evidenced by variable colony growth in cFP assays (FIG. 19A) and quantified as distributions of DIP rates (FIG. 6B). To explore the origin of this subclonal variability, we perform stochastic simulations[137] on a simple birth-death model of cell proliferation to ascertain whether intrinsic noise in division/death decisions alone is sufficient to explain experimental observations (see Methods). We perform a battery of *in silico* cFP assays, where untreated single cells grow into colonies of variable size at a fixed proliferation rate (division rate constant – death rate constant) and are then treated with drug, modeled by reducing the proliferation rate. Colony sizes are tracked over time (FIG. 19B) and DIP rate distributions are calculated and statistically compared against experimental distributions (FIG. 19C). We repeat this procedure for a wide range of division and death rate constants to identify ranges of parameter values that can statistically reproduce experimental DIP rate distributions ($p>0.05$, bootstrapped Anderson-Darling test). For all sublines (except DS8, see next paragraph), we find ranges of parameter values that are physiologically plausible (FIG. 20). This is strong evidence that these sublines (DS1, DS3, DS4, DS6, DS7, DS9) are monoclonal, i.e., experimental DIP rate distributions can be reproduced with a birth-death model containing a single cell state (one division and one death rate constant) simulated stochastically.

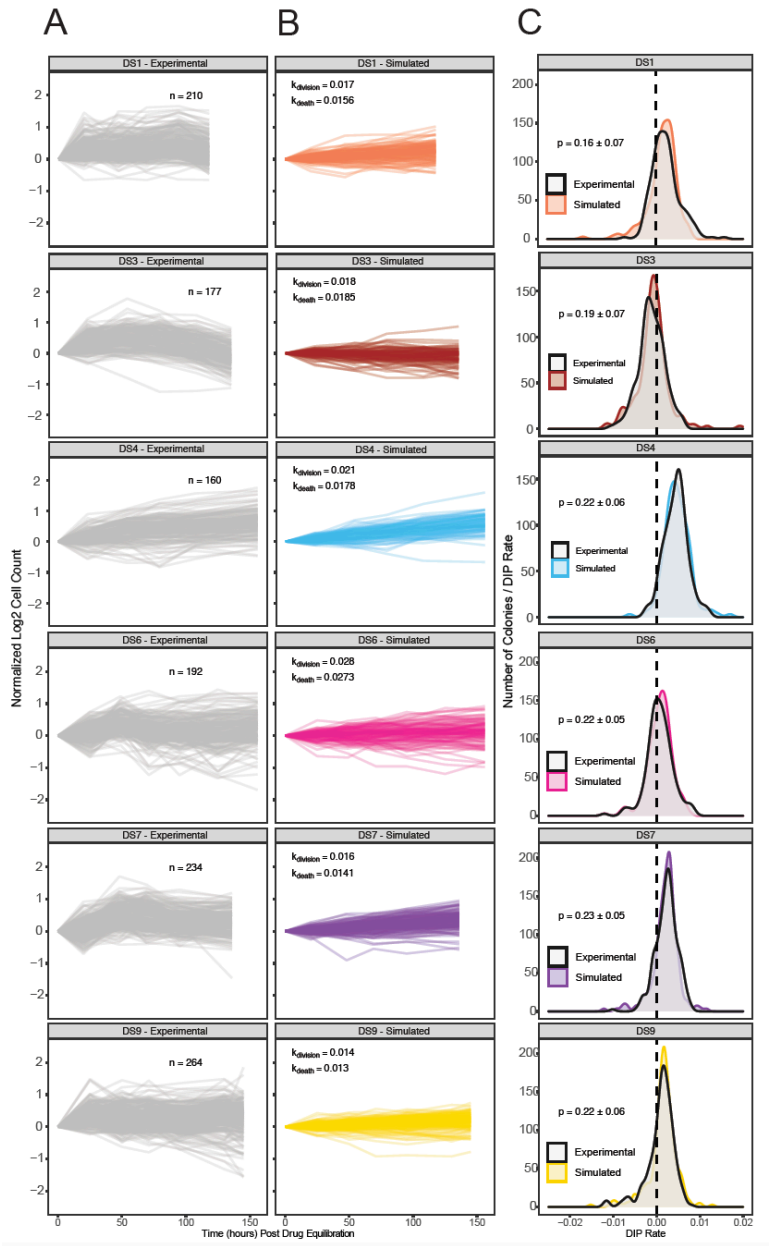

Figure 19 | Comparison of experimental and simulated cFP assays for PC9-VU sublines.
(A) Experimental clonal Fractional Proliferation (cFP) time courses for six PC9-VU sublines (DS1, DS3, DS4, DS6, DS7, DS9) in response to 3μM erlotinib (same data used to generate DIP rate distributions in FIG. 6B). Each trace corresponds to a single colony, normalized to 72h post-drug treatment. Only colonies with cell counts greater than 50 at the time of treatment were kept. *n* represents the number of colony traces for each subline. (B) Simulated cFP time courses generated using division and death rate constants that closely reproduce the experimental time courses in *A*. Trajectories are normalized to the time at which the simulated drug treatment was initiated and simulated cell counts are plotted only at experimental time points. Although the same number of simulations were initiated as the number of colonies (*n*) in the corresponding experiment (see panel *A*), only simulated colonies with cell counts > 50 at the time of simulated drug treatment are shown. (C) Comparison of experimental and simulated DIP rate distributions calculated from time courses in *A* and *B*. Distributions are compared statistically using the Anderson-Darling (AD) test (see Methods). Bootstrapped p-values are shown (mean and standard deviation). Dashed black line signifies zero DIP rate, for visual orientation.

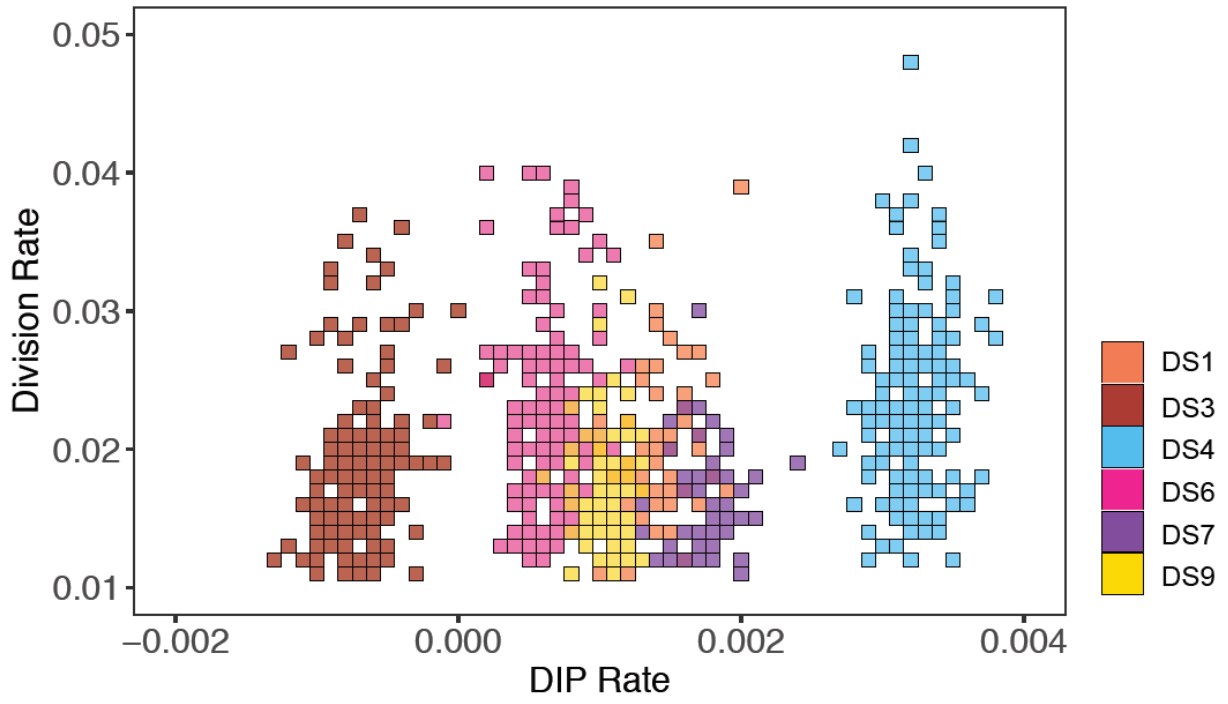

Figure 20 | ***Parameter scan of division and death rate constants for six PC9-VU sublines.***
For each pair of rate constants, we ran model simulations (same number as corresponding subline), calculated DIP rates and compiled them into a distribution, and then statistically compared against the corresponding experimental DIP rate distribution using the AD test (bootstrapped). All $p<0.05$ are colored white, indicating lack of statistical correspondence to experiment.

In contrast, DS8 is an exception once again, displaying greater variability than the other sublines in cFP colony growth rates (FIG. 21A) and a bimodal DIP rate distribution (FIG. 6B). We perform stochastic simulations on an expanded version of the birth-death model containing two cell states with distinct division and death rate constants (see Methods). As with the other sublines, we perform *in silico* cFP assays (FIG. 21B) and compare the simulated DIP rate distributions against the bimodal distribution seen experimentally (FIG. 21C). We again find physiologically feasible ranges of rate parameter values that can statistically reproduce the experimental result (FIG. 21D). Thus, these results provide strong evidence that DS8 harbors at least two distinct cell states. It is important to note that the mathematical model is agnostic

as to whether these two states are genetically or epigenetically distinct; it simply states that they have distinct division and death rate constants.

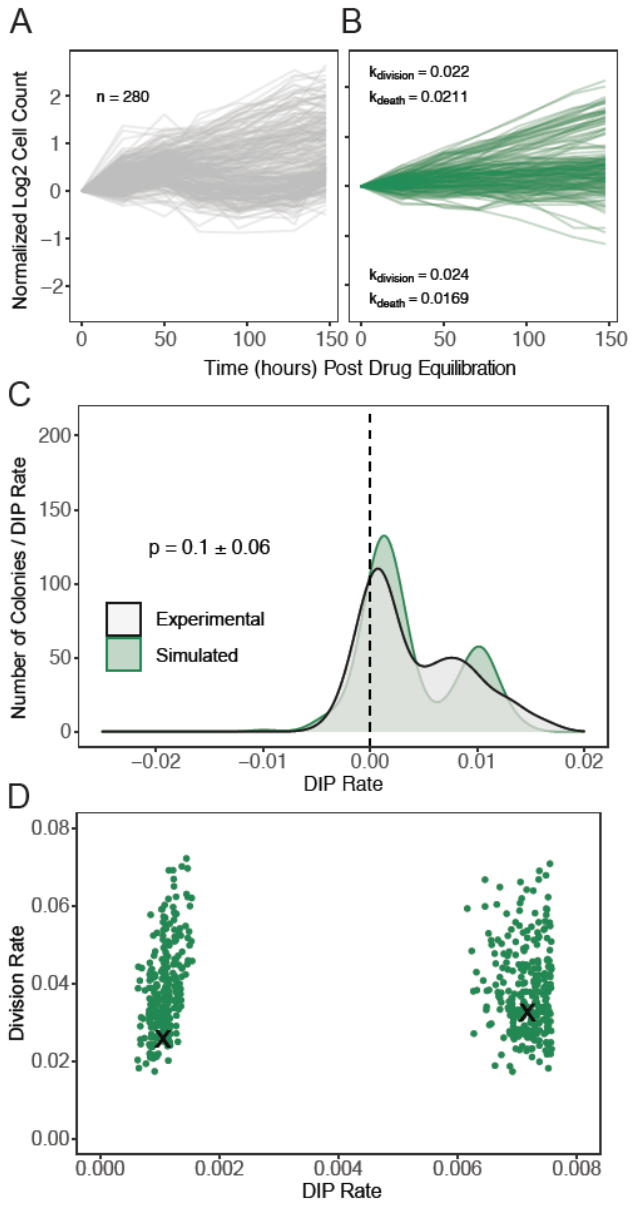

Figure 21 | ***Comparison of experimental and simulated cFP assays for DS8.***
(A) Same as FIG. 20*A* but for subline DS8. (B) Same as FIG. 20*B* but for DS8 using a two-state model. (C) Same as FIG. 20*C* but for DS8. (H) Same as FIG. 21 but for DS8 using a four-dimensional (two division-death rate constant pairs) parameter scan and projected into two dimensions. '×' denotes parameter values used to generate the simulated DIP rate distribution in *C*.

Methods

Cell culture and reagents

PC9-VU and PC9-BR1 were obtained as a gift from Dr. William Pao (Roche, formerly of Vanderbilt University Medical Center). PC9-MGH was obtained as a gift from Dr. Jeffrey Settleman (Massachusetts General Hospital). PC9-VU, PC9-MGH, and PC9-BR1 were individually fluorescently labeled with histone H2B conjugated to monomeric red fluorescent protein (H2BmRFP), as previously described[36,110,117,118,146]. The PC9 cell line versions and derivatives were cultured in Roswell Park Memorial Institute (RPMI) 1640 Medium (Corning) with 10% Fetal Bovine Serum (Gibco). Cells were incubated at 37°C, 5% CO_2, and passaged twice a week using TrpLE (Gibco). Cell lines and sublines were tested for mycoplasma contamination using the MycoAlert™ mycoplasma detection kit (Lonza), according to manufacturer's instructions, and confirmed to be mycoplasma-free. Cell line identity was confirmed using mutational signatures in whole exome sequencing. Erlotinib was obtained from Selleck Chemicals (Houston, TX) and solubilized in dimethyl sulfoxide (DMSO) at a stock concentration of 10mM and stored at -20°C. Cell lines were originally stored at -80°C, then moved into liquid nitrogen.

Derivation of single-cell sublines

The PC9-VU sublines were generated by limiting dilution of the parental cell line in 96-well plates. Wells with single cells were expanded for multiple weeks until large enough to be saved as frozen cell stocks. A single stock of each subline was brought back into culture, passaged for two weeks, and used for drug-response experiments. After 3-4 weeks of continued passaging, sublines were used for whole exome sequencing, bulk RNA sequencing, and single-cell RNA sequencing experiments. Since sublines were isolated from PC9-VU, they retained the same H2BmRFP nuclear label as cell line versions.

Population-level DIP rate assay

Cells were seeded in black, clear-bottom 96-well plates (Falcon) at a density of 2500 cells per well with six replicates for each sample. Plates were incubated at 37°C and 5% CO_2. After cell seeding, drug was added the following morning and changed every three days until the end of the experiment or confluency. Untreated samples were allowed to grow in DMSO-containing media until confluency, with media changes every three days. Plates were imaged using automated fluorescence microscopy (Cellavista Instrument, Synentec). Twenty-five non-overlapping fluorescent images (20X objective, 5x5 montage) were taken twice daily for a total of 500 hours or until confluency. Cellavista image segmentation software (Synentec) was utilized to calculate nuclear count (i.e., cell count) per well at each time point (Source = Cy3, Dichro = Cy3, Filter = Texas Red, Emission Time = 800μs, Gain = 20x, Quality = High, Binning = 2x2). Cell nucleus count across wells was used to calculate mean and 95% confidence intervals and normalized to time of drug treatment.

Clonal fractional proliferation assay

We modified the original cFP assay, which tracks multiple colonies in a single well of a plate[127]. Instead, here we flow-sorted single cells into a black, clear-bottom 384-well plate (Greiner) using fluorescence-activated cell sorting (FACS Aria III, RFP^+). Plates were incubated at 37°C, 5% CO_2 and cells were allowed to grow into small colonies over eight days in complete media (no media change). Drug was then added and changed every three days. Plates were imaged using the Cellavista Instrument (Synentec). Nine non-overlapping fluorescent images (3x3 montage of the whole well at 10X magnification) were taken once daily for a total of seven days. Cellavista image segmentation software (Synentec) was utilized to calculate nuclear count (i.e., cell count) per well at each time point (Source = Cy3, Dichro = Cy3, Filter = Texas Red, Emission Time = 800μs, Gain = 20x, Quality = High, Binning = 2x2). Depending on the number of wells that passed quality control thresholding (at least 50 cells per colony at the time of treatment), 160-280 replicates were included for each sample. DIP rates were calculated from 48h post-treatment to

the end of the experiment using the *lm* function in R. DIP rates for each sample were combined and plotted as a kernel density estimate. Mood's median test was performed to determine statistical difference between subline DIP rate distributions using the *RVAideMemoire* R package.

DNA bulk whole exome sequencing

Data acquisition: Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen), according to the manufacturer's protocol. Libraries were prepared using 150 ng of genomic DNA by first shearing the samples to a target insert size of 200 bp. Illumina's TruSeq Exome kit (Illumina, Cat: 20020615) was utilized per manufacturer's instructions. The samples were then captured using the IDT xGen® Exome Research Panel v1.0 (IDT, Cat: 1056115). The resulting libraries were quantified using a Qubit fluorometer (ThermoFisher), Bioanalyzer 2100 (Agilent) for library profile assessment, and qPCR (Kapa Biosciences, Cat: KK4622) to validate ligated material, according to the manufacturer's instructions. The libraries were sequenced using the NovaSeq 6000 with 150 bp paired end reads. RTA (version 2.4.11; Illumina) was used for base calling and sequence-specific quality control analysis was completed using MultiQC v1.7. Reads were aligned to the University of California, Santa Cruz (UCSC) hg38 reference genome using BWA-MEM[147] (version 0.7.17) with default parameters.

Genomic mutational analysis: Mutation analysis for single nucleotide polymorphisms (SNPs) and insertion/deletions (InDels) was performed using an in-house variant calling pipeline based on the Genome Analysis Toolkit (GATK, Broad Institute) recommendations. Duplicate reads were marked and replaced using PICARD (Broad Institute). Base recalibration and variant calling were performed using GATK version 3.8 (Broad Institute). Variants were selected and filtered based on gold standard SNPs and InDels, as well as a hard filtration according to GATK recommendations (SNPs: QD<2, QUAL<30, SOR>3, FS>60, MQ<40, MQRankSum<-12.5, ReadPosRankSum<-8; InDels: QD<2, QUAL<30, FS>200, ReadPosRankSum<-

20). Total variants were counted using VCFtools[148] (version 0.1.15). Sequencing metrics were calculated using vcfR[149] (version 1.8.0). These metrics included read depth, mapping quality, and a Phred quality score[150]. Variants were annotated by Variant Effect Predictor[82] (VEP; Ensembl Genome Browser version 95, available at ensembl.org) with multiple indicators, including chromosome name, gene symbol, mutation class, mutation type, and IMPACT rating. Mutation class corresponds to whether a variant is a substitution (SNV, sequence alteration) or Indel (insertion, deletion, or both - also referred to as indel). Type corresponds to the result of a variant on the amino acid sequence: synonymous (no effect), missense (codon change), nonsense (codon stop or start), splice site (boundary of intron and exon), or a shift in frame (inframe or frameshift). IMPACT rating is a subjective classification of the severity of variant consequence, as defined by Ensembl and based on genetic variant annotation and predicted effects (e.g., amino acid change, protein structure modification). The IMPACT rating categories are: *modifier* (no evidence of impact), *low* (unlikely to have disruptive impact), *moderate* (non-disruptive but might have effect), and *high* (assumed to have disruptive impact). *Modifier* variants were not plotted but constituted a majority of variants in all samples. Variants categorized into *low*, *moderate*, or *high* are referred to as IMPACT mutations in the text. The variant count distribution was organized as a mean-centered mutation count per chromosome for samples within each comparison set (cell line versions, sublines). Variant overlap analysis was conducted using VCFtools and visualized using the *UpSet* (version 1.4.0) R package. Variants unique to each sample were plotted as proportions of the total mutations in that sample. The *SNPRelate* R package (version 1.18.1) was used to project sample genotypes into PCA space and cluster based on similarity.

Automatic generation of a genetic mutational signature: Unique cell line version mutations were input into *dNdScv*[151] to generate a mutational signature that could define the genetic heterogeneity within the PC9 cell line family (all cell line versions and sublines). In this analysis, genes are first annotated by type.

dNdScv then uses a maximum-likelihood model to detect genes under positive selection (i.e., potential driver mutations). For each gene, a variety of models are utilized to identify genes that have substantially more nonsynonymous mutations than expected in each of the nonsynonymous mutation types, as compared to synonymous mutational load. These metrics are combined together to calculate a global p-value (see original publication for details). We use this approach, with a maximum number of mutations per gene per sample = 5 (tool recommendation; limits a hyper-mutator phenotype), to determine genes with a global p-value <0.05 in all eight members of the PC9 cell line family. The resulting gene signature sets the baseline for the genetic heterogeneity for all cell populations (Table 2). We visualized the mutation data as a heatmap of these significant genes and cell line versions or sublines, colored by annotated mutation type (number of variants in each gene-population pair are not annotated in heatmap but are shown in adjacent bar charts for number of variants per gene and sample).

Table 2 | ***List of genes associated with mutation heatmaps.***

Index	Gene	Index	Gene	Index	Gene
1	BHLHA15	35	IL32	69	AC092647.5
2	NCR2	36	SOX21	70	ARID1B
3	TMEM158	37	GDF11	71	CNPY3
4	JAGN1	38	MECP2	72	FOXD1
5	DLX4	39	VKORC1L1	73	TPPP
6	RAD51B	40	SMAP1	74	PRDM8
7	FAU	41	XBP1	75	OLIG1
8	AGAP3	42	RAPH1	76	DBNDD2
9	KRTAP10-6	43	MMP17	77	MYL1
10	CGREF1	44	TERF1	78	LRP1B
11	RPS15	45	POU3F2	79	CCDC74A
12	TUBB8P12	46	BCHE	80	ZNF257
13	HCRT	47	POU3F3	81	YY1
14	SKA3	48	ZFP36L2	82	FOXN3
15	OR8G1	49	ZNF787	83	PCDH9
16	GPR37L1	50	BCL6B	84	MMP1
17	BHLHE22	51	TDG	85	RRP8
18	RRAGD	52	IGSF21	86	TCHH
19	VEGFC	53	VCX3B	87	PRAMEF6
20	NAP1L5	54	OR2A1		
21	SOX10	55	SLC35F1		
22	IGFBP2	56	CADM4		
23	ADRA2B	57	NPIPB3		
24	SOX11	58	FAM155A		
25	FAM98C	59	PHLDA1		
26	CEBPA	60	RAB21		
27	PHOSPHO1	61	OR2T35		
28	PHGR1	62	FOXO6		
29	BRI3BP	63	POU3F1		
30	CARNMT1	64	FMR1NB		
31	TAF11L2	65	ASTN2		
32	EPB41L1	66	SNX31		
33	CAPNS1	67	MNX1		
34	ACSM2A	68	ZNF713		

Literature-curated, cancer-associated mutational signature: A cancer-associated gene list was established to supplement the predicted genetic heterogeneity signature from *dNdScv*. The gene list was created from the NIH Genetics Home Reference (GHR) key lung cancer genes (ghr.nlm.nih.gov/condition/lung-cancer#genes) and two additional publications of key mutations in lung cancer[12,140]. Associated heatmaps were generated of this gene list for cell line versions and sublines, colored by annotated mutation type.

RNA single-cell transcriptome sequencing

Data acquisition: scRNA-seq libraries were prepared using the 10X Genomics gene expression kit (version 2, 3' counting[152]) and cell hashing[153]. Cells were prepared according to recommendations from the cell hashing protocol on the CITE-seq website (cite-seq.com/protocol). After cell preparation, 1 ng of eight different cell hashing antibodies (TotalSeq-A025(1-8) anti-human Hashtag, Biolegend) were used to label each of the eight samples. Hashed single-cell samples were combined in approximately similar proportions and 'super-loaded' (aiming for ~20,000 cells, ~15,400 cells were obtained) onto the Chromium instrument. Cells were encapsulated according to manufacturer guidelines. Single-cell mRNA expression libraries were prepared according to manufacturer instructions. The leftover eluent of the mRNA expression library, containing the hashtag oligonucleotides (HTOs), was utilized to further size select the HTO library. The size-selected HTO library was PCR amplified to obtain higher quality reads. Libraries were cleaned using SPRI beads (Beckman Coulter) and quantified using a Bioanalyzer 2100 (Agilent). The libraries were sequenced using the NovaSeq 6000 with 150 bp paired-end reads targeting 50M reads per hashed sample for the mRNA library and a spike-in fraction for the HTO library. RTA (version 2.4.11; Illumina) was used for base calling and MultiQC (version 1.7) for quality control. Gene counting, including alignment, filtering, barcode counting, and unique molecular identifier (UMI) counting was performed using the *count* function in the 10X Genomics software *Cell Ranger* (version 3.0.2) with the GRCh38 (hg38) reference transcriptome. We utilized *CITE-seq-Count* (github.com/Hoohm/CITE-seq-

Count) to count HTOs from the HTO library. We then used the *Demux* function in the R package Seurat[154] (satijalab.org/Seurat, version 3.2.2) to demultiplex the HTO and mRNA libraries and pair cells to their associated hashtag. Data was integrated into a count matrix with genes and cells, with HTO identity as a metadata tag.

Data analysis: After creating the demultiplexed, single-cell gene expression matrix, we removed cell multiplets using both cell hashtags and automated doublet detection (*DoubletFinder* version 3 with default parameters). Additionally, poor quality cells were removed based on a minimum cutoff of features (number of genes detected in each cell = 3000) and count (number of RNA molecules detected within each cell = 15000). These numbers were chosen subjectively but with respect to the overall sequencing depth in order to remove droplets with ambient RNA. Cells below these thresholds fell in isolated regions of the UMAP space and had a large degree of overlap with all other samples. Feature selection was performed according to Seurat guidelines (0.1 < average gene expression < 8, log variance-to-mean ratio > 1; 574 genes met criteria). A cell cycle score was calculated on the expression matrix (default implementation; satijalab.org/seurat/v3.1/cell_cycle_vignette.html), which was used to classify the cell cycle phase. Data were visualized using the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction algorithm, as implemented in the Seurat[154] R package. To facilitate comparisons across cell line versions and sublines, we performed the UMAP projections in the space of all eight cell line versions and PC9-VU sublines (PC9-VU, PC9-MGH, PC9-BR1, DS3, DS6, DS7, DS8, DS9). To quantify overlap of cells between transcriptomic features, we performed k-means clustering of the cell line versions (k=3) and the sublines (k=2) using the *NBClust* package, which also identified the optimal number of clusters based on 30 different methods. Distances between cluster centroids were calculated in the common UMAP space using the *dist* function (Euclidian) in R. Differential expression analysis was performed between a single sample (cell line version or subline) and the rest of the PC9 cell line family

members using the Wilcoxon rank sum test (as implemented in the *FindMarkers* function in Seurat, default settings). Differentially expressed genes (DEGs, adjusted-p < 0.05) were used for downstream analyses (see "Gene Ontology analysis" below).

Copy number variation analysis

Copy number variants were inferred from scRNA-seq data using InferCNV (Trinity CTAT Project, github.com/broadinstitute/inferCNV). The single-cell transcriptome count matrix (see "RNA single-cell transcriptome sequencing: *Data acquisition*" above), an annotation file (pairing each cell to its corresponding PC9 cell line family member), and a gene order file (derived from the GRCh38/hg38 gtf file) were used to create an inferCNV object. Separate objects were created for cell line versions (no reference group; an average of the three versions was used, default setting) and sublines (PC9-VU reference group). The inferCNV analysis was run with a cutoff of 0.1 (default for droplet-based experimental methods). Clustering was performed based on annotation file groups (i.e., cell line versions, sublines). Analysis settings were to denoise the dataset and use a hidden Markov model for CNV predictions. Heatmaps of relative expression values (to the reference group) were output by chromosome for all cells in the analysis. Red corresponds to increased expression (amplification) and blue to decreased expression (deletion).

scRNA-seq functional interpretation analysis

The single-cell transcriptome count matrix (see "RNA single-cell transcriptome sequencing: *Data acquisition*" above) was scaled by multiplying counts by the median RNA molecules across all cells and dividing that number by the number of RNA molecules in each cell (as recommended by VISION[155] developers). Gene signature files were obtained from the molecular signatures database (MSigDB). *Hallmark* gene sets (50 in total) were downloaded from MSigDB (gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H). Both the scaled counts matrix and each of the

hallmark gene sets were input into VISION[155] to identify gene signature scores for each cell-signature pair. Four hallmark gene sets (KRAS_SIGNALING_UP, KRAS_SIGNALING_DOWN, UV_RESPONSE_UP, UV_RESPONSE_DOWN) were condensed into two (KRAS_SIGNALING, UV_RESPONSE) by VISION to leave 48 total gene signatures. Scores were compiled into a distribution and plotted by PC9 cell line family member for each gene set.

RNA bulk transcriptome sequencing

Data acquisition: Total RNA was extracted using a Trizole extraction (ThermoFisher), according to the manufacturer's protocol. RNA-seq libraries were prepared using 200 ng of total RNA and the NEBNext rRNA Depletion Kit (NEB, Cat: E6310X), per manufacturer's instructions. The kit employs an RNaseH-based method to deplete both cytoplasmic (5S rRNA, 5.8S rRNA, 18S rRNA and 28S rRNA) and mitochondrial ribosomal RNA (12S rRNA and 16S rRNA). The mRNA was enriched via poly-A-selection using oligoDT beads and then the RNA was thermally fragmented and converted to cDNA. The cDNA was adenylated for adaptor ligation and PCR amplified. The resulting libraries were quantified using a Qubit fluorometer (ThermoFisher), Bioanalyzer 2100 (Agilent) for library profile assessment, and qPCR (Kapa Biosciences, Cat: KK4622) to validate ligated material, according to the manufacturer's instructions. The libraries were sequenced using the NovaSeq 6000 with 150 bp paired-end reads. RTA (version 2.4.11, Illumina) was used for base calling and MultiQC (version 1.7) for quality control. Reads were aligned using STAR[156] (version 2.5.2b) with default parameters to the STAR hg38 reference genome. Gene counts were obtained using the featureCounts[157] package (version 1.6.4) within the Subread package. The gene transfer format (GTF) file for the genes analyzed in the scRNA-seq data (provided by 10X Genomics and used in the *Cell Ranger* pipeline, generated from the hg38 reference transcriptome) was used to better facilitate internal comparison between scRNA-seq and bulk RNA-seq datasets.

Data analysis: RNA-seq data was analyzed using the DESeq2[158] R package. Counts were transformed using the regularized logarithm (rlog) normalization algorithm. PCA was performed using the *prcomp* function in R and hierarchical clustering using the *hclust* R function with a Ward's minimum variance method. Data was visualized using the *ggplot2* R package.

Gene ontology analysis

Motivation: It is well known that a mutation in a gene does not always alter the expression level of that gene[159]. Rather, if gene *X* encodes for a protein that acts as a transcription factor for gene *Y*, then a mutation in *X* could alter the protein structure such that it affects the expression level of *Y*. Thus, rather than making gene-by-gene comparisons of mutations and expression levels, we compare data modalities at the level of gene ontology (GO) terms.

Setup: Genes associated with unique IMPACT mutations (classified as *low*, *moderate*, or *high* IMPACT scores; see "DNA bulk whole exome sequencing: *Genomic mutational analysis"*) were identified for each comparison set (i.e., cell line versions or sublines). Additionally, DEGs from the scRNA-seq statistical comparisons for each comparison set were determined (see "RNA single-cell transcriptome sequencing: *Data analysis*" above). The two gene lists were independently subjected to a GO enrichment analysis using *EnrichR*[160] (version 2.1). Genes were compared to the ontology databases *GO Biological Process 2018* (BP)*, GO Molecular Function 2018* (MF)*,* and *GO Cellular Component 2018* (CC), which we refer to as GO "type" in the text.

Correlation analysis: GO terms significantly enriched in the IMPACT mutations ($p < 0.05$) and in DEGs ($p < 0.05$) were identified and stored independently as separate GO term lists for each PC9 cell line family member. For terms shared between the lists, we calculated $-\log_{10}$(p-value) to rank terms based on

statistical significance. Spearman correlation was calculated between the significant GO terms using the *ggpubr* R package (version 0.4.0), as long as a minimum of five significant terms were shared between the IMPACT and DEG GO term lists. Terms with a $-\log_{10}$(p-value) > 2 for the IMPACT or DEG GO term lists were highlighted. Outlier GO terms ($-\log_{10}$(p-value) > 10) were excluded from the analysis (two terms in PC9-BR1, one in PC9-MGH), in order to not unfairly skew correlation calculations.

Semantic similarity analysis: GO term lists for each PC9 cell line family member were further separated into GO types, which created GO term lists unique for each combination of sample (cell line version or subline), dataset (IMPACT mutations or DEGs), and GO type ("BP", "MF", or "CC"). For each sample, the mutation and DEG GO term lists associated with each GO type were compared using the semantic similarity metric from Wang et. al.[145], as calculated in the *GOSemSim* package[161] (version 2.12.1) using the *goSim* function. This approach compares two individual GO terms using the underlying GO term network structure. Pairwise similarities were calculated on the lists of terms to generate similarity matrices for each sample. In order to avoid the dismissal of terms near any proposed statistical cutoff and ensure lists were of a minimum length, mutation and DEG GO term lists associated with each GO type for each sample were chosen randomly based on a modified p-value metric from the GO enrichment analysis. Specifically, terms were chosen from each list if a random number (between 0 and 1) was greater than the GO enrichment p-value. Semantic similarity distributions had a large skew, biased heavily towards lower values, primarily due to the size of the GO type graph network structure and, therefore, the "distance" between terms in the graph. To combat this issue, a maximum range of semantic similarity scores were chosen for each comparison (similar to but more robust than the "best max average" option provided in *GOSemSim*). The median of the top 1000 scores and a 95% confidence interval were calculated for each sample-GO type comparison. Semantic similarity scores were also correlated with the number of genes input into the GO enrichment analysis. To address this problem, random gene lists of the same lengths as

IMPACT and DEG gene lists were chosen and input into the same process as experiment-derived gene lists in order to generate a simulated semantic similarity distribution. Depending on the length of the lists, these simulated distributions varied in the both the median and variance. Therefore, instead of comparing experimental and simulated distributions, experimental semantic similarity scores were normalized to the median + one standard deviation of the simulated score, for each sample. These relative GO semantic similarity score distributions are represented in plots. Importantly, the number of GO terms varied across GO types (according to *GOSemSim*; BP: 12,272, MF: 4,165, CC: 1,740). We assume that GO types with more terms are more biologically significant, i.e., BP is more predictive than MF, followed by CC.

In silico modeling of clonal fractional proliferation

Birth-death population growth models. Mathematical models of population growth dynamics were constructed using PySB[162], a Python-based kinetic modeling and simulation framework. We modeled cell proliferation as a simple birth-death process,

$$Cell_i \xrightarrow{k_{div,i}} Cell_i + Cell_i \quad (1)$$

$$Cell_i \xrightarrow{k_{dth,i}} \emptyset \quad (2)$$

where *i* is an integer index, $k_{div,i}$ and $k_{dth,i}$ are division and death rate constants, respectively, for cell type *i*, and Ø denotes cell death. Note that there is no state switching included in the model. Models with one cell type were used to compare against experimental cFP data for the majority of the PC9-VU sublines (DS1, DS3, DS4, DS6, DS7, DS9), while a two-cell-type model was used in one case (DS8).

Stochastic simulations and in silico DIP rate distributions. All model simulations were run using the stochastic simulation algorithm[137] (SSA), as implemented in BioNetGen[163] (invoked from within PySB), to capture the effects of random fluctuations in division and death on cell population proliferation. We

performed *in silico* cFP assays, where numerous single cells (run as independent simulations) were grown into colonies of variable size over eight days of simulated time using the SSA and fixed rate constants for division and death (k_{div} = 0.04*ln(2) h^{-1}, k_{dth} = 0.005*ln(2) h^{-1}; Table 3), based on vehicle-control proliferation data (FIG. 6A, dashed lines; in the case of two cell types, both types were assumed to grow at the same rate outside drug). We ran as many simulations as there were experimental cFP trajectories for the PC9-VU subline being compared against. Drug treatment was then modeled by changing the rate constants for division and death (for two cell types, each was assumed to proliferate at different rates in drug; Table 3) and running for the additional days of simulated time corresponding to each subline experiment. Simulated time courses were plotted at the same time points as in the corresponding experimental cFP assays for direct comparison. *In silico* DIP rates were obtained by taking $\log_2$ of the total cell counts and calculating the slope of a linear fit to the time course from the time of drug addition to the end of the simulation using the SciPy *linregress* function. DIP rates for all *in silico* colonies were compiled into a distribution and compared to the corresponding experimental cFP distribution using the Anderson-Darling (AD) test. The p-value for each simulation result was bootstrapped based on 100 resamples of the experimental distribution. For DS8, a two-state model was used. All aspects of the model were the same, except drug treatment was modeled by two sets of division and death rate constants. These simulations were constrained by the approximate DIP rate ranges for both DIP rate distribution modes (FIG. 6B).

Table 3 | Rate parameters used for monoclonal and polyclonal growth models.
Ranges are shown in square brackets, units are in hr^{-1}.

Subline / Parameter	DS1	DS3	DS4	DS6	DS7	DS8	DS9
$k_{div,pre\text{-}drug}$	0.04*ln(2)	0.04*ln(2)	0.04*ln(2)	0.04*ln(2)	0.04*ln(2)	0.04*ln(2)	0.04*ln(2)
$k_{DIP,pre\text{-}drug}$	0.035*ln(2)	0.035*ln(2)	0.035*ln(2)	0.035*ln(2)	0.035*ln(2)	0.035*ln(2)	0.035*ln(2)
$k_{div1,post\text{-}drug}$	[0.01, 0.08]	[0.01, 0.08]	[0.01, 0.08]	[0.01, 0.08]	[0.01, 0.08]	[0.01, 0.08]	[0.01, 0.08]
$k_{DIP1,post\text{-}drug}$	[0.0005, 0.0025]	[-0.0015, 0.0005]	[0.0025, 0.0045]	[-0.0005, 0.0015]	[0.001, 0.003]	[0.0005, 0.0015]	[0, 0.002]
$k_{div2,post\text{-}drug}$	N/A	N/A	N/A	N/A	N/A	[0.01, 0.08]	N/A
$k_{DIP2,post\text{-}drug}$	N/A	N/A	N/A	N/A	N/A	[0.0055, 0.0075]	N/A

Parameter scans. We repeated the simulation procedure above over ranges of physiologically plausible division and death rate constant values (Table 3). For each parameter set (either one or two pairs of post-drug division/death rate constants, depending on subline), Anderson-Darling tests for simulated vs. experimental DIP rate distributions were performed. To account for variabilities in individual comparisons, only simulations with a mean p-value minus one standard deviation (from the bootstrapped result) were kept. Any simulation with p-sdev > 0.05 (i.e., we cannot reject the null hypothesis that the DIP rates are drawn from the same distribution) were plotted in a heatmap using the *ggplot2* R package. Thus, all colored elements in the heatmap represent sets of division/death rate constants that produce DIP rate distributions statistically indistinguishable from the experimental distributions obtained from cFP assays. Note that the scan for the two-cell-type model (DS8) was performed in four dimensions (two division and two death rate constants) but results were plotted in two dimensions for visual simplicity.

Conclusion

Cancer heterogeneity remains a persistent problem for cancer treatment, and often results in tumor drug tolerance or resistance. This chapter uses an *in vitro* cancer cell line model to attribute various sources of tumor heterogeneity to multiple 'versions' and single-cell derived sublines of commonly used EGFR-mutant cell line PC9. We find that genetic heterogeneity dominates drug response variability across cell line versions (PC9-VU, PC9-MGH, PC9-BR1). PC9-BR1 exhibits the canonical $EGFR^{T790M}$ resistance mutation, as well as various sequence and copy number variants that make it resistant to EGFRi. PC9-MGH varies predominately in CNVs, which translates to a distinct epigenetic state and increased drug sensitivity. PC9-VU exhibits mostly epigenetic variability, as seen by various single-cell derived sublines with similar genetic states but differential transcriptomic signatures. One subline, DS8, harbors a distinct genetic state, resulting in a separate transcriptomic signature and drug insensitivity. Mathematical modeling of subline population dynamics show intrinsic stochastic fluctuations in a single state can explain drug response variability, again with the exception of DS8, which requires at least two states.

Together, these results advocate for a framework of cancer heterogeneity comprising genetic, epigenetic, and stochastic components. Through the lens of this framework, we can explain both the data in this chapter, as well as additional studies on cancer heterogeneity. It also offers an alternative to the traditional gene-centric view of heterogeneity in the cancer research community. By considering cancer heterogeneity as a complex system of interacting levels, one can appreciate the difficulty in treating the disease, as well as identify alternative treatment strategies.

CHAPTER III

A Non-Quiescent "Idling" Population State in Drug-Treated, BRAF-Mutant Melanoma[2]

Introduction

Targeted therapy is an effective standard of care in BRAF-mutated malignant melanoma. However, the duration of tumor remission varies unpredictably among patients, and resistance is almost inevitable. Common among these patients is residual disease, the tumor population that remains after treatment that often falls under the limit of detection. Residual disease has been hypothesized to be comprised of drug-tolerant cells that 'persist' after treatment[12,21,69], and is considered to be the reservoir from which resistant tumors arise. Understanding the nature of these persister cells is likely crucial to the eradication of tumors prior to resistance.

In this chapter, we examine the responses of several BRAF-mutated melanoma cell lines (including isogenic clonal sublines) to BRAF inhibitors. We observe complex response dynamics across cell lines, with short-term responses (<100 h) varying from cell line to cell line. In the long term, however, we observe equilibration of all drug-treated populations into a non-quiescent state characterized by a balanced rate of death and division, which we term the "idling" state. Using mathematical modeling, we propose that the observed population-level dynamics are the result of cells transitioning between basins of attraction within a drug-modified phenotypic landscape, each associated with a DIP rate. The idling population state represents a new dynamic equilibrium in which cells are distributed across the landscape such that the population achieves zero net growth. By fitting our model to experimental drug-response data, we infer the phenotypic landscapes of all considered melanoma cell lines and provide a unifying view of how BRAF-

[2] Adapted with permission from Paudel B.B., Harris L.A., Hayford C.E. *et al.* A Nonquiescent "Idling" Population State in Drug-Treated, BRAF-Mutated Melanoma. *Biophysical Journal* 6, 114 (2018). https://doi.org/10.1016/j/bpj.2018.01.016

mutant melanomas respond to BRAF inhibition. We also identify differential drug responses of a clonal subline as it diversifies in the absence of drug, eventually adopting a similar response to the parental. We hypothesize that the residual disease observed in patients after targeted therapy is composed of a significant number of idling cells, and differential short-term dynamics can be attributed to the occupancy of phenotypic states. This view provides a theoretical underpinning of cancer drug tolerance and paves the way toward therapies based on rational modification of the landscape to favor basins with greater drug susceptibility.

Results

BRAF-mutant melanoma cells enter a drug-tolerant 'idling' state upon treatment with BRAFi

To investigate the effects of BRAF inhibition (BRAFi) on BRAF-mutant melanoma, we subjected commonly-used BRAF-mutant melanoma cell line SKMEL5 to small-molecular BRAFi PLX4720 (analogous to Vemurafenib) for approximately seven days at a variety of different doses. Cell counts were collected multiple times a day over the course of drug treatment and compiled into drug-response trajectories (FIG. 22A). In all BRAFi dose concentrations, nonlinear dynamics were captured in the response to treatment. Short-term BRAFi treatment moderately reduces the DIP rate in a dose-dependent way, followed by a further long-term reduction at higher concentrations. At the highest BRAFi concentrations (8-32 uM), cells achieve a somewhat stable net-zero DIP rate over long-term BRAFi (FIG. 22B). Importantly, this net-zero DIP rate is not due to dish confluence (FIG. 22C) or quiescence (FIG. 22D), as cells continue to divide and die throughout prolonged BRAFi without further expanding. For these reasons, we termed these cells 'idling,' as they continue to die and divide without expanding, much like a car engine continuing to turn over but not moving while parked.

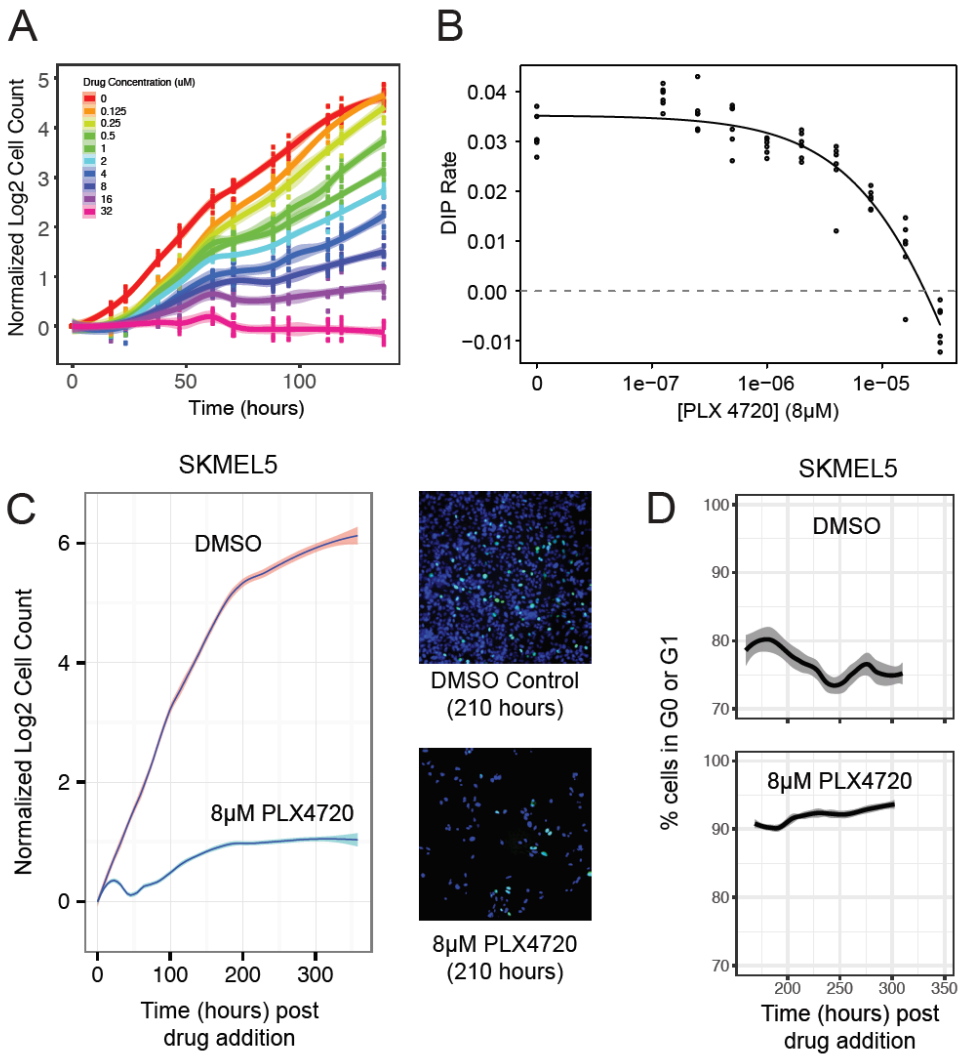

Figure 22 | *BRAFi-Induced responses of BRAF-mutant melanoma cell populations.*
(A) Population growth curves (log2 normalized) for SKMEL5 parental cells treated with varying concentrations of BRAFi. (B) DIP rate-based dose-response curve (C) (Left) Response dynamics of drug- (8μM PLX4720) and control-treated (DMSO) cells. (Right) Hoechst-stained (blue) and FUCCI (cell cycle indicator; green) images of drug- and control-treated cells. The idling state is not due to confluence in a cell culture dish. (D) Percentage of FUCCI-negative cells for SKMEL5 cells post initial treatment (168-310h) with DMSO or drug. Drug-treated idling cells are still dividing, but at approximately half the rate of control-treated cells.

To expand our understanding of this 'idling' behavior across BRAF-mutant melanoma, we treated six additional BRAF-mutant melanoma cell lines with the same prolonged BRAFi treatment (FIG. 23A). In each case, the cell lines exhibited differential short-term dynamics, but eventually adopted the net-zero DIP rate characteristic of idling. Interestingly, when given a drug holiday (~48 hours) and treated again

with BRAFi, drug response dynamics are similar to pre-treatment (FIG. 23B). This drug holiday experiment suggests that the idling state is reversible. Taken together, these data suggest BRAF-mutant melanoma cell lines enter a drug-tolerant 'idling' state with prolonged BRAFi, which is reversible.

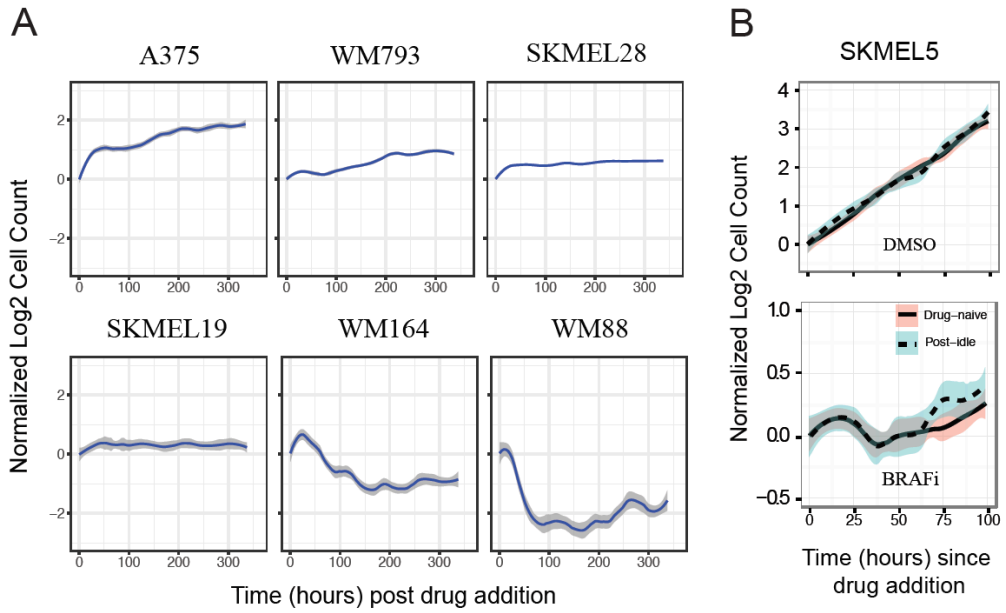

Figure 23 | *BRAF-mutant melanoma cell lines idle under continued BRAFi.*
(A) Population growth curves (log_2 normalized; 3+ technical replicates) for six BRAF-mutant melanoma cell lines treated with saturating concentrations of BRAFi . (B) Idling cell populations (7 days BRAFi treatment) return to normal pre-drug function after a 24h drug holiday: (Top) Drug-naïve and idling cells expand at equal rates in control treatment (DMSO); (Bottom) Drug-naïve and idling cells response similarly to BRAFi treatment in BRAFi (8μM PLX4720). Mean responses are shown as solid or dashed lines. 95% point-wise confidence intervals are shown in shades regions with a loess fit.

Idling occurs for populations of single-cell derived clonal sublines with varying short-term BRAFi sensitivities

To better understand the short-term drug response dynamics in BRAFi, we subjected SKMEL5 to a clonal drug response assay. This assay, called clonal fractional proliferation[127] (cFP), dissociates a cell population into many clones, which are expanded into colonies and treated with drugs. We tracked 84 single-cell derived colonies using the cFP assay, and quantified their responses in terms of the DIP rate

(FIG. 24A). Drug responses varied from clone to clone, encompassing a wide rate of behaviors, from expanding to regressing. DIP rates were compiled into a distribution (FIG. 24B) to quantify the variability. The resulting distribution for SKMEL5 showed a majority of the cells had negative DIP rates, with a long tail. However, a small minority of cells have slightly positive DIP rates. The combination of these drug responses seem to lead to the short-term drug response dynamics (FIG. 23), where cells continue to expand as cells equilibrate to drug treatment (up to ~60 hours), followed by a regression (~60-90 hours, tied to the negatively proliferating cells) and further expansion (~90-125 hours, tied to positive proliferating cells).

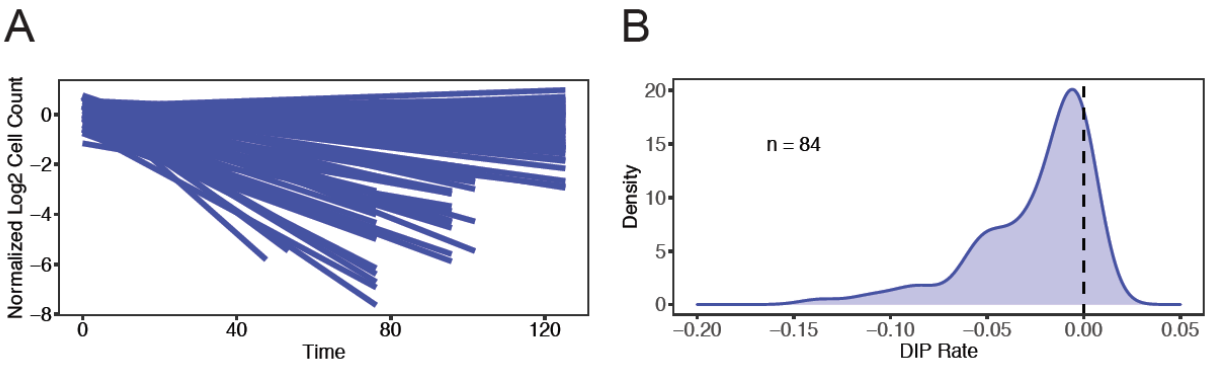

Figure 24 | *Short-term drug response dynamics reveal pre-existing clonal heterogeneity.*
(A) Linear fits of population growth curves (log_2 normalized) for BRAFi-treated SKMEL5 single-cell derived colonies (n=84) obtained using the cFP assay. (B) DIP rate distribution obtained from fits in *A*.

To better understand the long-term idling behavior, we isolated sixteen sublines from the SKMEL5 cell line prior to treatment and treated them with the same BRAFi therapy. As in the cFP assay, drug responses varied significantly across the clonal sublines upon BRAFi in the short term (<100 hours), with some expanding, some regressing, and others maintaining a stationary population size (FIG. 25A). We selected three sublines representative of the range of observed short-term responses for further experimentation: SC01 (regressing), SC07 (stationary), and SC10 (expanding). Upon prolonged BRAFi

exposure (>120 hours), despite their initial divergent responses, both SC01 and SC10 converged to near-zero DIP rates, where SC07 maintained its initial zero-net-growth response (FIG. 25B). Furthermore, by manually tracking the fates of multiple individual cells over time, we determined that all three subline populations exhibit death and division while in the idling state (FIG. 25C). Thus, although we cannot exclude the possibility that some cells might be quiescent, these results confirm that balanced rates of death and division are largely responsible for idling cell populations. This suggests, therefore, that the idling population state is a characteristic feature of BRAF-mutated melanoma, which is achieved by all clonal sublines in the continued presence of BRAF inhibition regardless of their initial responses.

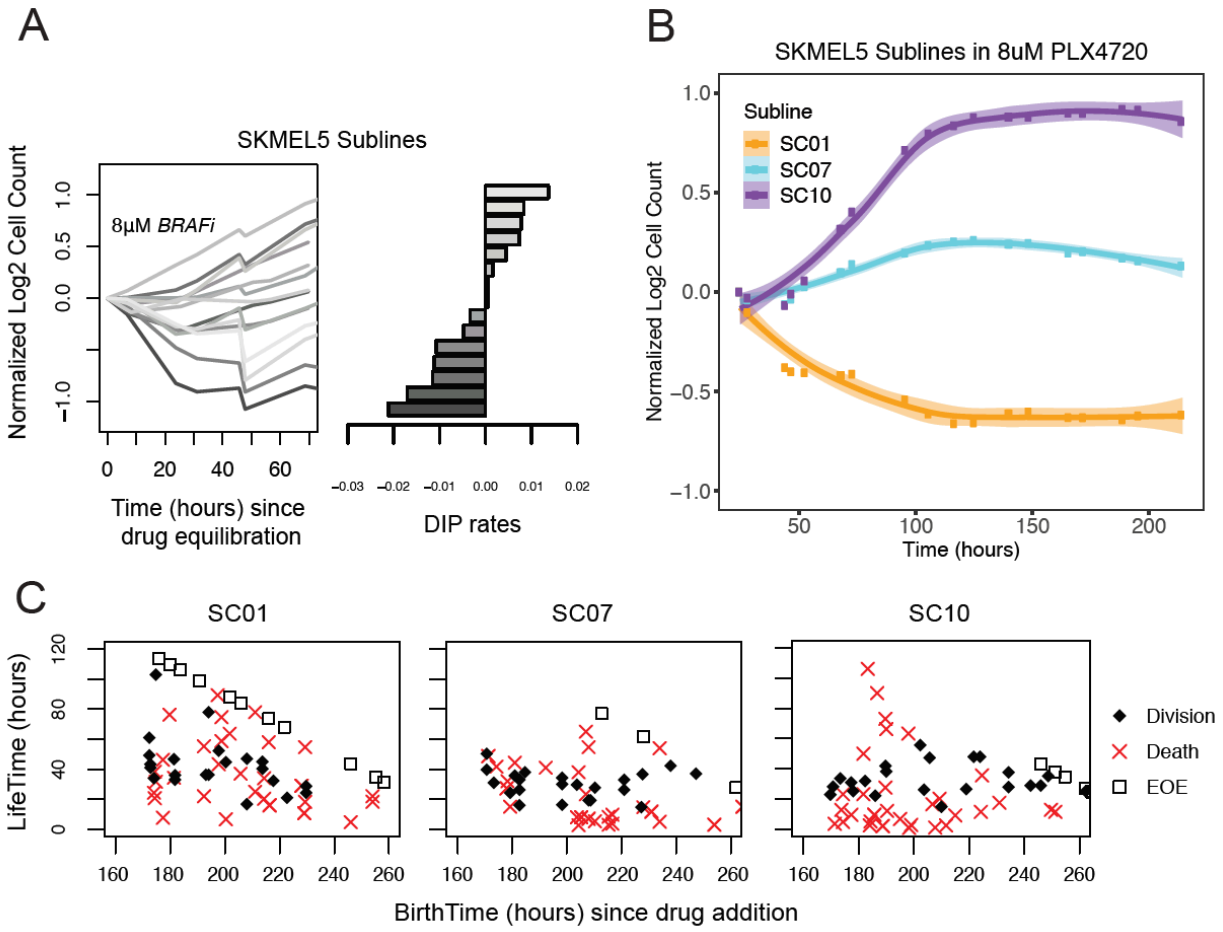

Figure 25 | *Single-cell derived clonal sublines idle independent of short-term drug sensitivity.*
(A) Sixteen single cell-derived SKMEL5 clonal sublines treated with 8 μM BRAFi: (Left) Population growth curves (log_2 normalized) and (right) bar plot of BRAFi-treated DIP rates (calculated as linear fits to the growth curves). (B) Population growth curves (log_2 normalized) for three clonal sublines treated with BRAFi (6 technical replicates; means are points; Lines are loess fits; Shaded regions are 95% point-wise confidence interval). (C) Single-cell lifespans versus birth times (time of first mitotic event) for the three sublines in *B*. Experiment started after 7 days in BRAFi and continued an additional 100 h. Cells born during the experiment that reached the end of experiment (EOE) without a second mitotic event are plotted on the diagonal.

Mathematical modeling qualitatively reproduces complex, population-level drug-response dynamics

Due the reversibility of the idling phenotype (FIG. 23B), we assumed that differences in the short-term clonal variability (FIG. 25A-B) originate from non-genetic processes, i.e., are not based on genetic alterations (differs from the traditional "clone" definition[164]). This result suggested that nongenetic (or more specific, epigenetic) heterogeneity was present prior to treatment with BRAFi. The long-term

convergence in drug-response phenotypes (FIGS. 22A,C, 23A, 25A-B) to a net-zero DIP rate further suggests that cells transition between epigenetic states during treatment.

To mathematically formalize these observations, we constructed a simple kinetic model of cell proliferation that qualitatively captures the treatment responses of all BRAF-mutant melanoma cell lines considered here (FIG. 26A). Briefly, we defined three drug-response phenotypes: regressing (R), stationary (S), and expanding (E), with negative, zero, and positive DIP rates, respectively (corresponding to example clonal sublines in FIG. 25B). A negative DIP rate results from the rate of cell death exceeding the rate of cell division, a zero DIP rate is due to balanced rates of cell death and division, and a positive DIP rate is due to the rate of cell division exceeding death. We assume that cells in each subpopulation can (reversibly) transition into ‘‘adjacent’’ phenotypes, thus changing the proportion of cells in each phenotype over time. With cells distributed across the three phenotypes, we expect a period of short-term, nonlinear dynamics driven by differences in rates of death and division of each subpopulation (i.e., clonal competition), followed by transitions of cells between phenotypes, resulting in the population achieving idling. Intuitively, the nature of the short-term dynamics and the timescale for transition into idling will depend upon the initial proportions of cells in each phenotype and the values of the transition rate constants.

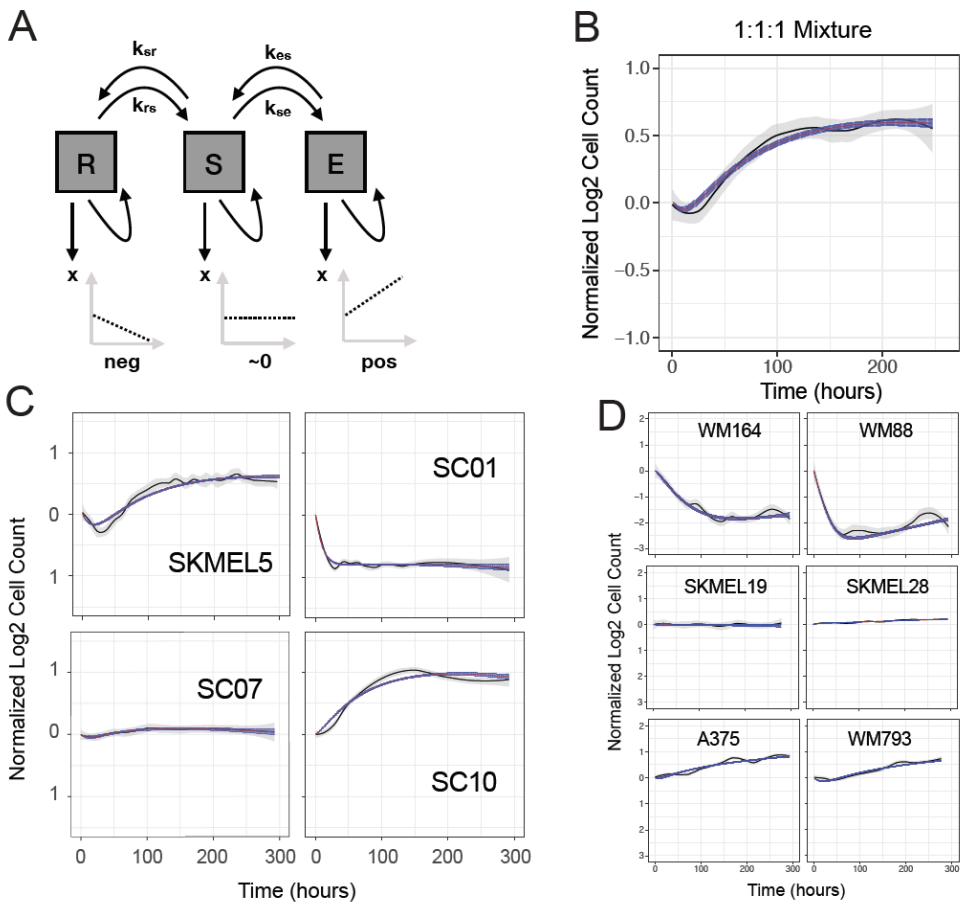

Figure 26 | ***Three-state kinetic model qualitatively captures complex drug-response dynamics.***
(A) Graphical representation of three-state model. Arrows represent cell fates (i.e. death, division, and state transition), and each state corresponds to different growth dynamics (bottom). (B) Model fitting to a 1:1:1 clonal mixture of SKMEL5 clonal sublines (SC01, SC07, SC10). (C) Model fits of BRAFi-induced responses of the SKMEL5 cell line and its clonal sublines using the rate constants inferred from *B*. (D) Model fits of BRAFi-induced responses of six additional BRAF-mutant melanoma cell lines (same as in *FIG. 23A*). All plots have experimental (black line, grey shading) and simulated (blue shading - one standard deviation envelope for 1000 parameter sets) trajectories.

For the SKMEL5 cell line, we estimated model parameters by fitting the mathematical model to experimental data for a 1:1:1 clonal mixture of the SC01, SC07, and SC10 clonal sublines (FIG. 26B) using MCMC sampling (see Methods). We then used the estimated transition rate constants to predict drug-response dynamics for individual SKMEL5 clonal sublines as well as the parental line (allowing the initial cell proportions to vary; FIG. 26C). In each case, the model predictions matched closely with the experimental time courses, capturing both the short-term dynamics and the transition into the idling

population state. Importantly, the model cannot explain the observed dynamics if the transition rate constants are set to zero, demonstrating that the nonlinear growth curves cannot simply be explained by clonal selection. We also fitted the model to experimental time courses for six additional melanoma cell lines (FIG. 26D). In all cases, the model predicted that the idling state is composed of very few cells in the R subpopulation but significant proportions of cells in both the S and E subpopulations, ranging between 20 and 80% (FIG. 27A). This is a significant result because it demonstrates that the idling state is not characteristic of an individual cell but is rather an emergent state of the population as a whole. In summary, our model, incorporating both clonal competition and phenotypic state transitions, captured the key features of the drug-response dynamics of several BRAF-mutated melanoma cell populations, which differ significantly in the shape and duration of their short-term response but eventually converge into an idling population state.

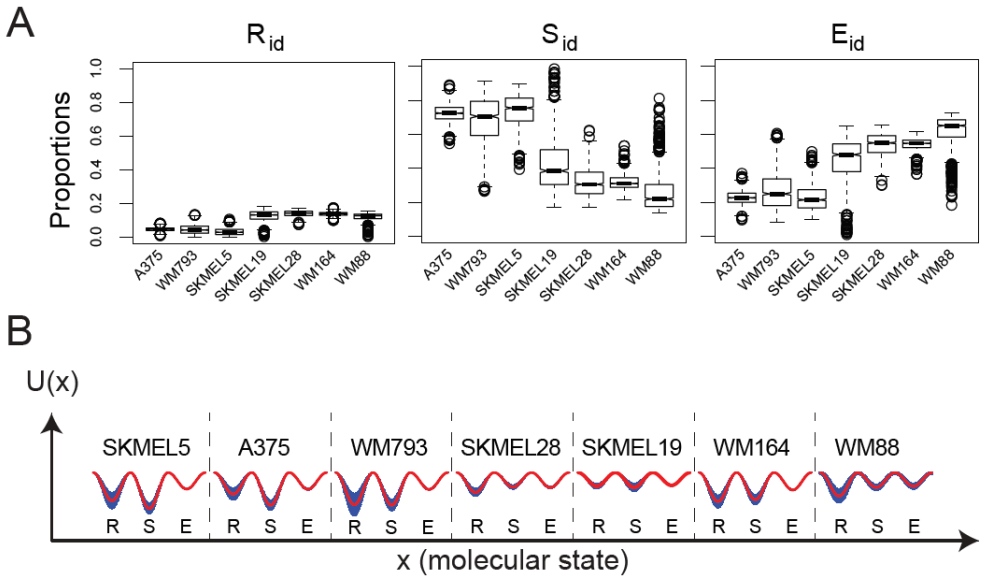

Figure 27 | ***Model-inferred drug-modified landscapes provide insight into drug-response dynamics.***
(A) Model-predicted proportions of cells in the regressing (R_{id}), stationary (S_{id}), and expanding (E_{id}) subpopulations of the idling state. (B) Calculated landscapes for all seven BRAF-mutant melanoma cell lines considered in this chapter. U(x) is the quasi-potential energy, and x is the molecular state (sometimes called "reaction coordinate"). Landscapes are based on 2000 random samples of MCMC-generated parameter ensembles. Mean basin depths and barrier heights are shown as red lines; Blue shaded regions correspond to one standard deviation around the mean.

Using the values of the transition rate constants from the MCMC-generated parameter ensembles, we inferred BRAFi-induced quasi-potential-energy landscapes for all BRAF-mutated melanoma cell lines considered (Fig. 27B; see Methods). Within a quasi-potential-energy landscape, each subpopulation is associated with a basin of attraction, and transitions between subpopulations amount to traversals of energy barriers separating basins[89]. Our results showed that the basin associated with the expanding subpopulation (E) was consistently the shallowest across cell lines (i.e., had the smallest exit barrier). This makes intuitive sense, since for the cell population to reach the idling state (zero net growth), cells must rapidly evacuate the basin for the E state, otherwise the cell population would continue expanding. However, at equilibrium, a proportion of cells remained in this basin, providing a source to counterbalance the cell depletion occurring in the regressing subpopulation. The depths of the basins associated with the regressing and stationary subpopulations were more variable than the basin for the expanding state (across the MCMC-generated parameter ensemble), but they generally showed the basin for the S state to be deeper than the basin for the R state. Notable exceptions were the WM164 and WM88 cell lines, which exhibited significant short-term cell loss. Again, it makes intuitive sense because for this to occur, because the exit barrier from the basins for the R to the S state has to be large enough to reduce the flux into the basin for the S state after drug exposure. Overall, the inferred epigenetic landscapes are powerful theoretical tools for understanding the basis of the complex population-level dynamics observed in BRAF-mutated melanoma cell populations and for reconciling differences in drug response seen across cell lines in terms of variations in an epigenetic landscape's topography.

Subclonal diversification in the absence of drug leads to differential in-drug responses

In addition to cellular adaptation to drug-induced epigenetic landscape changes, cancer cells have also been known to change in the absence of treatment[165] . Stochastic, or transcriptional state switching

has been known to happen in the absence of drug treatment, and can sometimes have an effect on drug sensitivity phenotypes[61,98,99]. To address this issue, we subjected previously isolated clonal subline SC01 to prolonged cell passaging in the absence of selective pressure to mimic drug-naïve diversification (FIG. 28A). At different points in the diversification process, samples of cells were drawn from the population and tested in BRAFi (passages 9, 13, 22, and 26; FIG. 28B). Interestingly, each successively higher passage sampled adopted a slightly larger DIP rate when tested in BRAFi. In passages 22 and lower, sampled cells seemed to still acquire the long-term net-zero DIP rate characteristic of idling. However, at least in the timeframe tested, passage 26 cells continued a stable positive DIP rate in prolonged BRAFi, similar to the short-term dynamics identified in the parental cell line (FIGS. 22A and 26C). A qualitative look back at the BRAF-mutant melanoma three-state model suggests a potential explanation (FIG. 28C). SC01 cells, at low passage, are presumably monoclonal in nature. Given sufficient time, epigenetic diversification allows cells to overcome barriers in the epigenetic landscape. Based on the drug-response dynamics seen in across cell passages, we predicted that SC01 cells may diversify into adjacent basins in the absence of drug treatment, which are only revealed in BRAFi.

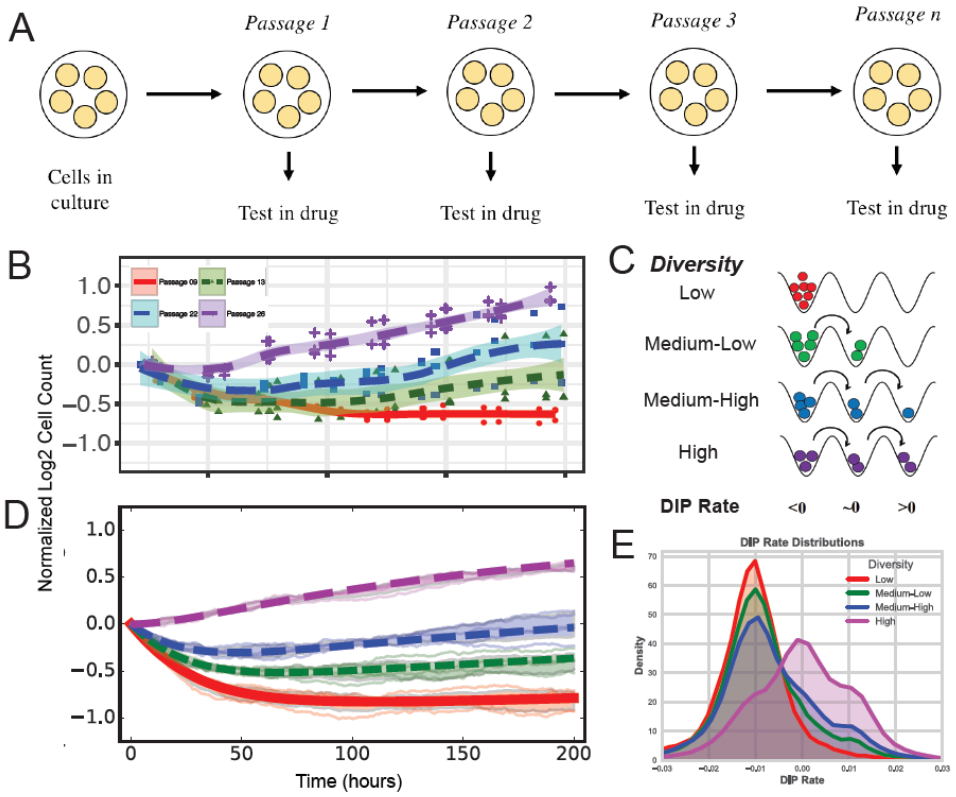

Figure 28 | *Clonal phenotypic diversification in the absence of selective pressure*.
(A) Design for population DIP rate experiment to identify clonal diversification. (B) Population growth curves (log_2 normalized) for SC01 at different cell passages (9, 13, 22, 26) treated with BRAFi (6 technical replicates; means are points; Lines are loess fits; Shaded regions are 95% point-wise confidence interval). (C) Hypothesized diversification of SC01 over increasing passage number. Basins in landscapes correspond to DIP rate phenotypes, and colors are shared with *B*. (D) Simulations of drug treated based on different initial model conditions (diversity across states). Six stochastic simulations were run for each condition. Simulation means are noted by the thick lines, and shaded regions represent a 95% confidence interval. (E) Simulated DIP rate distribution from *in silico* cFP assays. 1000 clonal drug response trajectories were simulated for each initial condition, and linear fits were compiled into distributions.

Therefore, we revisited the three-state model to quantitatively address this clonal diversification hypothesis by varying the initial clonal composition (i.e., model initial conditions). Using the MCMC model fits as a guide (FIG. 26C), we chose model initial conditions corresponding to the subclonal diversity for each passage number and simulated drug response dynamics (FIG. 28D). Specifically, we used six stochastic simulations, using the SSA (Gillespie) algorithm[137], to model six experimental replicates for each

initial condition. These simulations are strikingly concordant with the experimental drug response trajectories, even down to the lack of idling being present in passage 26 over the experimental timeframe.

We next simulated single-cell DIP rate distributions using *in silico* cFP assays (see Methods) to ensure that the population DIP rates roughly reflect the input distributions (FIG. 28E). However, although distributions at the diversity extremes were distinguishable, it became clear that it would be difficult to statistically quantify distribution differences between conditions using the standard experimental cFP assay. To address this concern, we created a modified cFP assay that changes the number of initial cells (from 1 to >5) to enrich the number of potential diversified clones and therefore shift the distribution mean rather than various modes. This assay, termed the low seeding density (LSD) assay (see Methods), was used to test a similar progression of drug-naïve diversification as in *FIG. 28B*. We first simulated the LSD assay, which showed a mild distribution shift over increasing passage number (FIG. 29A). We then performed experimental LSD assays over a similar diversification course, which shows a similar moderate shift over increasing passage number (FIG. 29B). We then use a quantitative distribution overlap metric, known as the *histogram distance*, to quantify this diversification. Compared to the self-distance (a statistical resampling; see Methods), all distributions after passage 11 show a statistically distinguishable difference from the earliest passage (FIG. 29C).

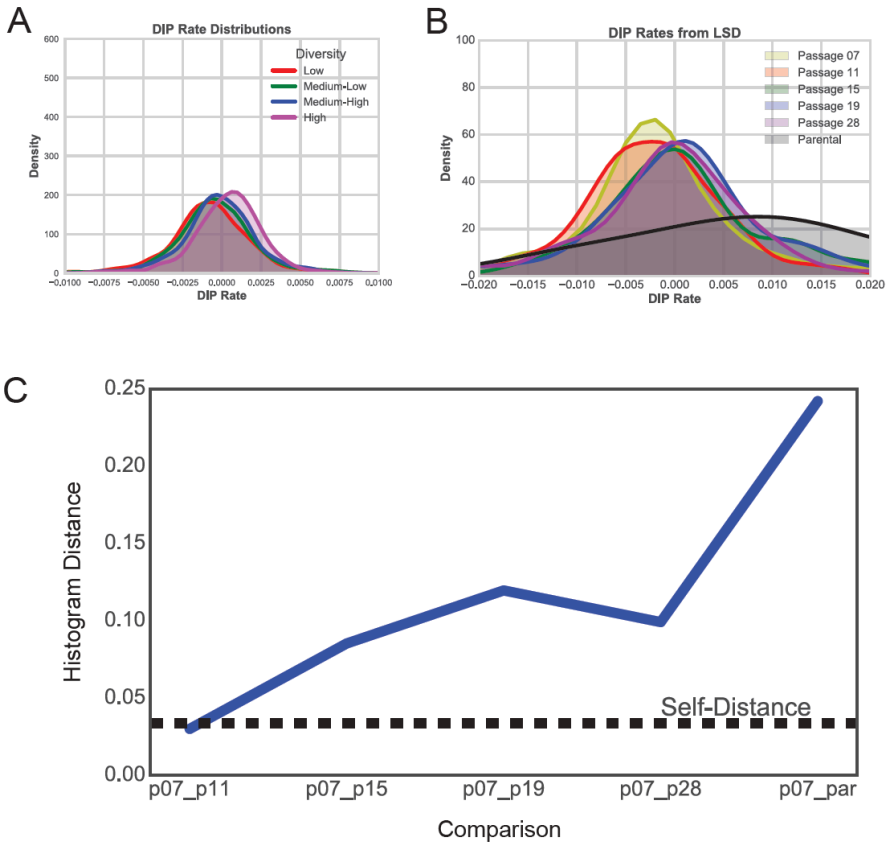

Figure 29 | *Clonal diversification using the LSD assay.*
(A) Simulated DIP rate distributions from *in silico* LSD assays. 1000 clonal drug response trajectories were simulated for each initial condition, and linear fits were compiled into distributions. Trajectories correspond to the clonal mixture of approximately five cells. (B) Experimental DIP rate distributions from the LSD assay. Linear fits of approximately 300 clonal mixtures (approximately 5 clonal colonies per mixture) were calculated and compiled into distributions for each passage number tested. (C) Histogram distance calculations for each passage number compared to the earliest tested (passage 7). Self-distance was calculated by resampling the experimental distribution and calculating a distance between the original and resampled distributions.

Methods

Cell culture and reagents

SKMEL5 cell line was purchased from ATCC© and labeled with either a fluorescent histone H2B conjugated to the green fluorescent protein (H2B-GFP), or monomeric red fluorescent protein (H2BmRFP; Addgene plasmid No. 18982) and geminin1–110 monomeric Azami-Green. Cells were cultured a mixed media of DMEM and Ham F-12 media (DMEM:F12 1:1, catalog no. 11330-032), supplemented with 10%

fetal bovine serum (FBS). Cells were incubated at 37°C, 5% CO_2, and passaged twice a week using TrpLE (Gibco). Cell lines and sublines were tested for mycoplasma contamination using the MycoAlert™ mycoplasma detection kit (Lonza), according to manufacturer's instructions, and confirmed to be mycoplasma-free. Cell lines were originally stored at -80°C, then moved into liquid nitrogen. BRAF inhibitor PLX 4720 (analog to Vemurafenib) was obtained from MedChem Express (Monmouth Junction, NJ) and solubilized in dimethyl sulfoxide (DMSO) at a stock concentration of 10mM and stored at -20°C.

Derivation of single-cell derived sublines

Sixteen SKMEL5 clonal sublines were derived from single cells by serial dilution. Briefly, cells were serially diluted to less than one cell per well in 96-well imaging plates and imaged to identify wells containing a single cell. Cells were expanded in complete growth medium (in the absence of BRAFi) and sequentially transferred to 48-, 24-, and 6-well plates until sufficient numbers of cells were available for cryopreservation. All sublines were tested for their sensitivity to BRAFi before cryopreservation.

Population-level DIP rate assay

Cells were seeded in black, clear-bottom 96-well plates (Falcon) at a density of 2500 cells per well with six replicates for each sample. Plates were incubated at 37°C and 5% CO_2. After cell seeding, drug was added the following morning and changed every three days until the end of the experiment or confluency. Untreated samples were allowed to grow in DMSO-containing media until confluency, with media changes every three days. Plates were imaged using automated fluorescence microscopy (Cellavista Instrument, Synentec). Twenty-five non-overlapping fluorescent images (20X objective, 5x5 montage) were taken twice daily for approximately two weeks or until confluency. Cellavista image segmentation software (Synentec) was utilized to calculate nuclear count (i.e., cell count) per well at each time point (Source = Cy3, Dichro = Cy3, Filter = Texas Red, Emission Time = 800μs, Gain = 20x, Quality =

High, Binning = 2x2). SKMEL5 and clonal derivatives were treated with 8 uM BRAF inhibitor (BRAFi; PLX4720 unless otherwise stated). Cell nucleus count across wells was used to calculate mean and 95% confidence intervals and normalized to time of drug treatment.

Clonal fractional proliferation assay

We modified the original cFP assay, which tracks multiple colonies in a single well of a plate[127]. Instead, here we flow-sorted single cells into a black, clear-bottom 384-well plate (Greiner) using fluorescence-activated cell sorting (FACS Aria III, GFP^+). Plates were incubated at 37°C, 5% CO_2 and cells were allowed to grow into small colonies over eight days in complete media (no media change). Drug was then added and changed every three days. Plates were imaged using the Cellavista Instrument (Synentec). Nine non-overlapping fluorescent images (3x3 montage of the whole well at 10X magnification) were taken once daily for a total of seven days. Cellavista image segmentation software (Synentec) was utilized to calculate nuclear count (i.e., cell count) per well at each time point (Source = FITC, Dichro = FITC, Filter = FITC, Emission Time = 800μs, Gain = 20x, Quality = High, Binning = 2x2). Wells that passed quality control thresholding (at least 50 cells per colony at the time of treatment) were included in DIP rate calculations for each sample. DIP rates were calculated from 48h post-treatment to the end of the experiment using the *lm* function in R. DIP rates for each sample were combined and plotted as a kernel density estimate.

Low Seeding Density Assay

Analogous to the cFP assay (above), we plated cells into a black, clear-bottom 384-well plate (Greiner) targeting ~5 cells per well. Incubation and treatment were completed identically to the cFP assay. Plates were imaged using the Cellavista Instrument (Synentec). Five non-overlapping fluorescent images (3x3 montage of the whole well at 10X magnification) were taken once daily for a total of seven days. Cellavista image segmentation software (Synentec) was utilized to calculate nuclear count (i.e., cell

count) per well at each time point (Source = Cy3, Dichro = Cy3, Filter = Texas Red, Emission Time = 800μs, Gain = 20x, Quality = High, Binning = 2x2). Wells that passed quality control thresholding (at least 50 cells per well at the time of treatment) were included in DIP rate calculations for each sample. DIP rates were calculated from 48h post-treatment to the end of the experiment using the *lm* function in R. DIP rates for each sample were combined and plotted as a kernel density estimate.

Time-lapse single-cell tracking

Fluorescence images of cellular nuclei were obtained as described above. Images were acquired using a BD Pathway 855 in (spinning disk) confocal mode with a 20X (0.75 NA) objective in a CO2- and temperature-controlled environment every 20 minutes for 260 hours from the time of the first drug treatment. Media was replaced with a freshly prepared drug every 3 days. Images from each well were organized into stacks of a time series. Fluorescent nuclei were manually tracked across sequential images to obtain cell life spans and resultant cell fates (death or division). “Birth time” denotes the time at which a mitotic event occurred, resulting in two sister cells. “Lifetime” denotes the duration of single-cell viability until the cell either died or underwent another mitosis. “End of the experiment (EOE)” represents the cells that were born in a drug but did not exhibit any cell fate during the remaining observation time. Data are displayed as two-dimensional plots of birth time versus lifetime, with death, division, and EOE signified with different markers.

Mathematical modeling and parameter calibration

We consider three cell subpopulations, defined in terms of their net proliferation rates: *R* (regressing), *S* (stationary), and *E* (expanding). Cells within each subpopulation can divide, die, or transition into “adjacent” subpopulations. The ordinary differential equations (ODEs) describing the temporal dynamics of the system are

$$\frac{dN_R}{dt} = \left(k_{pR} - k_{rs}\right)N_R + k_{sr}N_S \quad (1)$$

$$\frac{dN_S}{dt} = \left(k_{pS} - k_{sr} - k_{se}\right)N_S + k_{rs}N_R + k_{es}N_E \quad (2)$$

$$\frac{dN_E}{dt} = \left(k_{pE} - k_{es}\right)N_E + k_{se}N_S \quad (3)$$

where N_R, N_S, and N_E are the numbers of cells in subpopulations *R*, *S*, and *E*, respectively; k_{pR}, k_{pS}, and k_{pE} are the DIP (net-proliferation) rates of subpopulations *R*, *S*, and *E*, respectively; k_{rs} and k_{sr} are the forward and reverse transition rate constants between subpopulations *R* and *S*, respectively; and k_{se} and k_{es} are the forward and reverse transition rate constants between subpopulations *S* and *E*, respectively (see Table 4). DIP rates were set to $k_{pR} = -0.055\ \mathrm{h}^{-1}$, $k_{pS} = 0\ \mathrm{h}^{-1}$, and $k_{pE} = 0.015\ \mathrm{h}^{-1}$ and a total initial cell population of 10,000 was assumed. The remaining six parameters [k_{rs}, k_{sr}, k_{se}, k_{es}, and the initial cell proportions R_0 and S_0 (≥0 and ≤1)] were determined by calibrating to experimental data (see below). The model was encoded in R (https://www.r-project.org/) and ODE simulations were performed using the *ode* function of the R package *deSolve*.

Table 4 | *Three-state model variables and parameters.*

Variable	Definition	
t	Time (h)	
N_R	Number of cells in state R	
N_S	Number of cells in state S	
N_E	Number of cells in state E	
T	Total number of cells	
Parameter	**Definition**	**Units**
k_{pR}	Net proliferation rate of cells in state R	h^{-1}
k_{pS}	Net proliferation rate of cells in state S	h^{-1}
k_{pE}	Net proliferation rate of cells in state E	h^{-1}
k_{rs}	Rate of transition of cells from state R to state S	h^{-1}
k_{sr}	Rate of transition of cells from state S to state R	h^{-1}
k_{se}	Rate of transition of cells from state S to state E	h^{-1}
k_{es}	Rate of transition of cells from state E to state S	h^{-1}
R_0	Initial proportion of cells in state R	unitless
S_0	Initial proportion of cells in state S	unitless
E_0	Initial proportion of cells in state E	unitless

Parameter calibration was performed using Markov chain Monte Carlo (MCMC) sampling[166,167] (1.5×10^5 iterations) using the *modMCMC* function of the R package *FME*. Goodness of fit was quantified using the cost function

$$Cost = \sum_{i=1}^{n} \frac{(M_i - O_i)^2}{\sigma_i} \quad (4)$$

where n is the number of measured time points and M_i, O_i, and σ_i are the model prediction, experimentally observed value, and standard experimental error (automatically determined by *modMCMC*) at time point i, respectively. For the SKMEL5 cell line, the model was calibrated against an experimental time course for a 1:1:1 clonal mixture of three single cell-derived subclones (SC01, SC07, and SC10). Predictions for the

dynamics of the SKMEL5 parental line and subclones were then made by selecting 1000 random parameter sets from the last 50% of iterations (accounting for burn-in) in the MCMC-generated parameter ensemble. Specifically, for each of the 1000 parameter sets, we recalibrated the model using MCMC keeping the transition rate constants (k_{rs}, k_{sr}, k_{se}, k_{es}) fixed at the values for that particular iteration and allowing the initial cell proportions (R_0 and S_0) to vary as free variables. For other *BRAF*-mutated melanoma cell lines (WM88, WM164, SKMEL28, SKMEL19, A375, WM793), model calibration was performed against experimental time courses for the parental lines. In all cases, we plot simulated time courses as one-standard-deviation envelopes around the mean from 1000 random samples of the MCMC-generated parameter ensemble.

Inferring quasi-potential energy landscapes

We assume the probability that a cell transitions from subpopulation X to Y follows Arrhenius' equation. Within this view, each subpopulation constitutes a basin of attraction within a quasi-potential energy landscape and transitions between subpopulations require traversal of an energy barrier separating adjacent basins. The height of this barrier, ΔU_{xy}, is proportional to the negative logarithm of the transition rate constant, i.e.,

$$\Delta U_{xy} \sim -ln\ k_{xy} \quad (5)$$

Intuitively, the higher the barrier the less probable is the transition. For each cell line considered, we randomly select 2000 parameter sets from the MCMC-generated parameter ensemble (see above) and estimate barrier heights between basins for each set using Eq. (5). A pictorial representation of the inferred quasi-potential energy landscape is then generated as a one-standard-deviation envelope around the mean barrier heights from the 2000 sampled parameter sets.

In silico modeling of clonal fractional proliferation

Birth-death population growth models. Mathematical models of single-cell population growth dynamics were constructed using PySB[162], a Python-based kinetic modeling and simulation framework. We modeled cell proliferation identical to above (i.e., three state model with adjacent state switching).

Stochastic simulations and in silico DIP rate distributions. Model simulations were run using the stochastic simulation algorithm[137] (SSA), as implemented in BioNetGen[163] (invoked from within PySB), to capture the effects of random fluctuations in division and death on cell population proliferation. We performed *in silico* cFP and LSD assays, where numerous single (cFP) or few (LSD) cells (run as independent simulations) were grown into colonies of variable size over eight days of simulated time using the SSA and fixed rate constants for division and death (k_{div} = 0.04*ln(2) h^{-1}, k_{dth} = 0.005*ln(2) h^{-1}), based on vehicle-control proliferation data. We ran 1000 simulations for each condition (diversification of SC01, represented by passage number). Drug treatment was then modeled by changing the rate constants for division and death and running for the additional days of simulated time corresponding to each subline experiment. Cells were allowed to transition between states throughout the simulated experiment, according to transition rate probabilities (see above). *In silico* DIP rates were obtained by taking $\log_2$ of the total cell counts and calculating the slope of a linear fit to the time course from the time of drug addition to the end of the simulation using the SciPy *linregress* function. DIP rates for all *in silico* colonies (cFP) or wells (LSD) were compiled into distributions.

Conclusion

Plasticity in response to cancer targeted therapies continues to be a primary contributor to drug evasion and eventual recurrence. This chapter provides an example of such plasticity in the form of a drug-

tolerant idling population state in BRAF-mutant melanoma. Akin to other forms of drug-tolerant persisters[12,21,69], idling cells survive at low levels in continued treatment. However, idling cells differ in that they continue dividing and dying during treatment, rather than the quiescence or senescence identified in other systems. Additionally, we find that idling is characteristic of all clonal derivatives of a representative cell line. Using a population dynamics model of division, death, and cell state transitions, we determined that the idling state was actually a population state, i.e. multiple states in dynamic equilibrium, albeit defined by various properties in each cell line. Finally, we show that drug treatment is not necessary for plasticity, as seen through diversification of a clonal subline over successive cell passage numbers. A novel technique was created to quantitatively survey clonal diversification, which showed experimental and simulated success.

Ultimately, these results point towards a view of cellular plasticity where genetically similar cells diversify over a variety of epigenetic states. Upon treatment, the landscape of epigenetic states changes, resulting in a newly equilibrated population state. Using this view, we can conceptualize hypotheses for epigenetic diversification in response to treatment and better understand treatment evasion and recurrence. This chapter provides an initial hypothesis for BRAF-mutant melanoma with the discovery of the idling state. Chapter IV will test this hypothesis and provide molecular features for both the drug-naïve and idling epigenetic landscapes.

CHAPTER IV

Ion Channel Dysregulation Induced by BRAF Inhibition Transitions Clonal Lineages into a Drug-Tolerant Idling State Susceptible to Ferroptosis

Introduction

Tumor plasticity in response to targeted therapy is a primary method by which cancer cells evade treatment[12,13,21]. Genetic mutations often cannot explain this plasticity, suggesting that non-genetic factors may play a role. In tumors without clear genetic indicators, epigenetic variability can explain some drug tolerance[21,36,168]. However, it is largely unclear which cells (or cell types) will evade treatment and the mechanism by which they achieve drug tolerance. Chapter III identified of a novel 'idling' population state in BRAF-mutant melanoma, and proposed a hypothesis by which cells evade treatment via epigenetic plasticity to allow the population to persist at low levels in BRAFi. Here, we aim to test this hypothesis and understand the molecular underpinnings of idling state drug tolerance in order to develop novel therapeutic strategies.

In this chapter, we use BRAF-mutant melanoma cell line SKMEL5 as an *in vitro* model of tumor plasticity to quantify the nature by which cells adopt the idling state in continued BRAFi. This state is achieved by an overwhelming majority of the drug-naïve population, identified by lineage tracing of barcoded SKMEL5 clones, as opposed to 'special' cells via clonal selection. Although more homogeneous than the drug-naïve population, idling cells are still heterogeneous, making up a 'population' state of dividing and non-dividing transcriptomic states. The relative fraction of clonal lineages that represent the idling state corresponds to lineage-specific occupancy of idling transcriptomic states, with some stochastic variability. Evidence of this heterogeneity is confirmed by transcriptomic characterization of single-cell derived clones in response to BRAFi. Idling state transcriptomic features point to ion channel activity as

the central process underlying drug tolerance, which is further supported by epigenomic characterization. Differential ion channel fluxes were verified between untreated and idling cells, which led to treatment with ferroptosis inducers to maximize idling cell death. Ferroptosis inducers increase drug potency, and provide a potential secondary therapy to work in sequence with BRAFi. This chapter provides support for the view that BRAFi induces re-equilibration of cancer cells to a new idling epigenetic landscape, which is collaterally sensitive to other treatments.

Results

Response to BRAFi is driven by a shift in the majority of cells

We chose BRAF-mutant melanoma cell line SKMEL5 as a model system to understand cellular plasticity in melanoma. This cell line is characterized by a $BRAF^{V600E}$ point mutation, making it sensitive to inhibition of the mutant BRAF protein. We have used this cell line in the past[36,133,134], as it has an average response to BRAFi compared to other BRAF-mutant melanoma cell lines. Interestingly, this cell line shows non-linear drug response dynamics to BRAFi. In response to BRAFi (>=8 μM), cells have an initial phase of growth, regression, and rebound, which we hypothesized was to clonal selection. However, then these cells move from a net-positive to a net-zero growth rate that remains stable for days. This change in the underlying phenotype from relatively high fitness in drug to an inability to proliferate further is a stark contrast to the traditional theory of cancer clonal selection, whereby cells acquire key genetic mutations that allow for survival in changing environments[14], such as drug treatment[12,13]. Therefore, we aimed to understand how these cells adopt the idling state, i.e., whether the idling state could be explained by clonal selection or if other mechanisms were at play.

To this end, we barcoded the SKMEL5 cell line with a gRNA barcoding library that allowed for both lineage tracing at the level of isolated, amplified barcode sequences (barcode sampling[32]) or mRNA in single-cell transcriptomics (scRNA-seq) (unpublished). This type of approach allows for the simultaneous

high-depth coverage of lineage dynamics and their dynamics in the transcriptomic space, in order to determine the distribution of barcodes that survive treatment (FIG. 30A). Cells were barcoded with a low multiplicity of infection (MOI, ~2.5%) with a limited barcode library (16 bp strong-weak balanced, 2^{16} potential barcodes) to achieve a minimum of 425 uniquely barcoded cells (FIG. 30B). Upon treatment with BRAFi PLX 4720 (8μM for 8 days), the barcode library complexity was reduced by less than 10% (FIG. 30B). These barcodes were shared within a large majority of replicates within each treatment condition (FIG. 30C), as well as between conditions (FIG. 30D). These results indicate that the idling state consists of cells from an overwhelming majority of the original population, providing evidence against the clonal selection hypothesis.

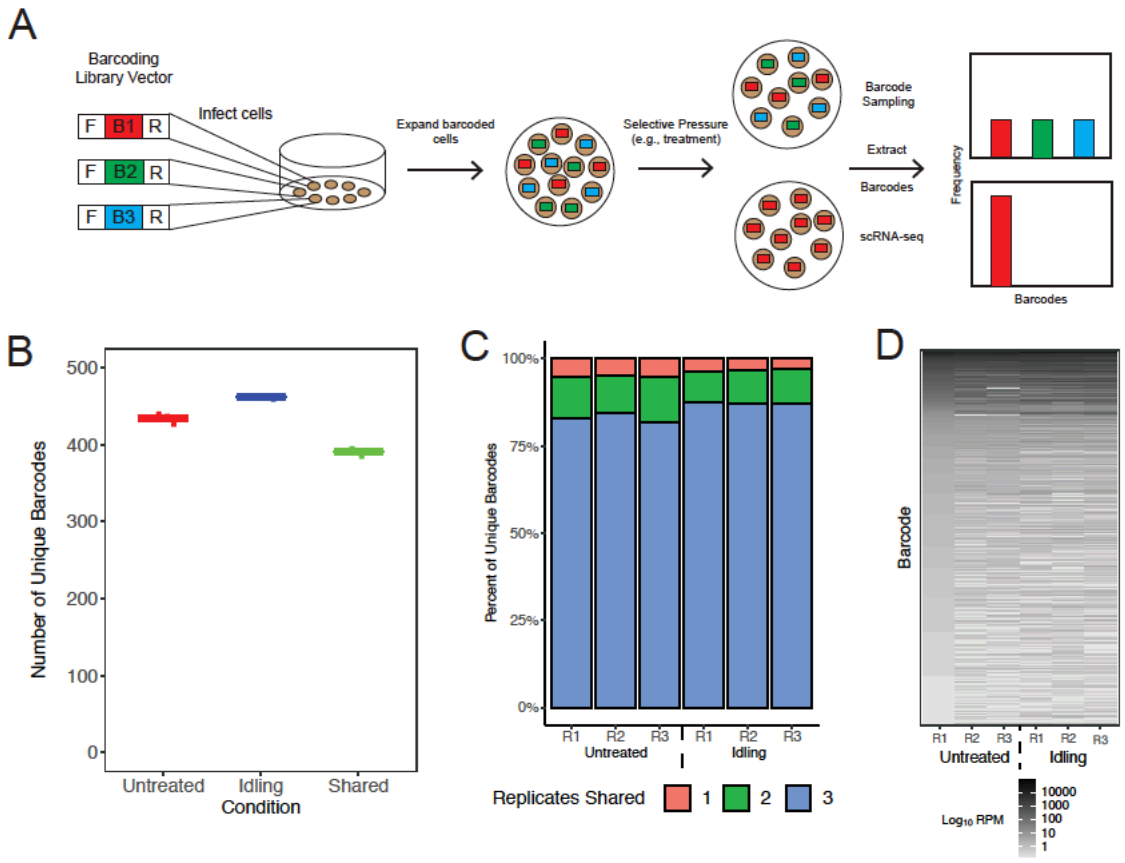

Figure 30 | ***Most clonal lineages survive treatment with BRAFi into idling.***
(A) Schematic of example lineage tracing experiments using cellular barcoding. (B) Number of unique barcodes in each treatment condition. Lines correspond to the means of three experimental replicates (points). A minimum cutoff of 100 counts per million (CPM) was used. (C) Proportional sharing of barcodes among experimental replicates (i.e. R1 = replicate 1) for each treatment condition. (D) Heatmap of relative barcode abundances (log_{10} CPM) for each experimental replicates across all captured barcodes. Heatmap is organized by decreasing barcode abundance in untreated condition.

Some fluctuations in relative barcode abundance did exist after treatment with BRAFi (FIG. 31A). Nearly all of the lineages that do not survive treatment come from clones that have an exceedingly small representation in the overall distribution, suggesting that the loss of those lineages is due to stochastic loss. Interestingly, the underlying fold change distribution for the entire barcoded cell population from untreated to idling is a normal distribution centered at zero (FIG. 31B). No clear exceptions were noted, and the top barcodes reflected this distribution (colored blocks in the histogram; same lineages as *FIG. 31A*; these barcodes will be used for downstream analyses). Therefore, although lineages exhibit some

expansion and contraction upon BRAFi (FIG. 31A), this behavior is not indicative of clonal selection. Instead, this is consistent with a hypothesis that cell state transitions dominate in the response to BRAFi, leading a diverse set of clonal lineages to adopt phenotypic states associated with drug tolerance. Rationale for the barcode fluctuations is explored in later sections.

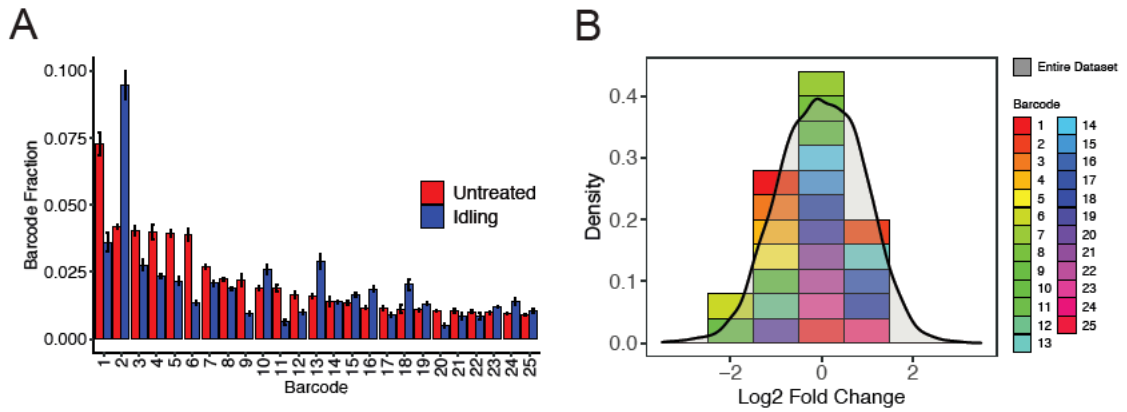

Figure 31 | ***Lineage dynamics in response to treatment reflected as relative barcode abundance.***
(A) Relative fraction of the top 25 ranked (in untreated) barcoded cell lineages in untreated and idling conditions. Bar height corresponds to the average of three experimental replicates (line is standard deviation). (B) Distribution of (log_2) fold change for barcoded clonal lineages from untreated to idling. Means of fold changes were compiled into a distribution for all captured lineages (grey), as well as the top 25 most abundant lineages noted in *A*.

BRAFi induces melanoma cells into a convergent, yet heterogeneous transcriptomic state

In order to determine the phenotypic state associated with idling drug tolerance, single-cell transcriptomics was performed on the barcoded cell line in both untreated and idling (see Methods). We use Uniform Manifold Approximation and Projection[141,142] (UMAP) to project the transcriptional states for each cell into a two-dimensional space (FIG. 32A). Untreated and Idling cells clears fall in different regions of the UMAP space, with minimal overlap. Additionally, the idling population falls in a much more constrained region of the UMAP space. To quantify variability of localization in the UMAP space, pairwise distances were calculated between cells in each condition. Pairwise distances were compiled into a cumulate density distribution for each condition (FIG. 32B) and an Earth mover's distance (EMD, see

Methods) was calculated between the distributions. More than 30% of the untreated cell-cell distances are greater than 10 distance units (Euclidian, UMAP space) apart, and nearly all idling cell-cell distances are lower than 10. The EMD between distributions indicates a rather large distance (2.53) separate the two population distance distributions.

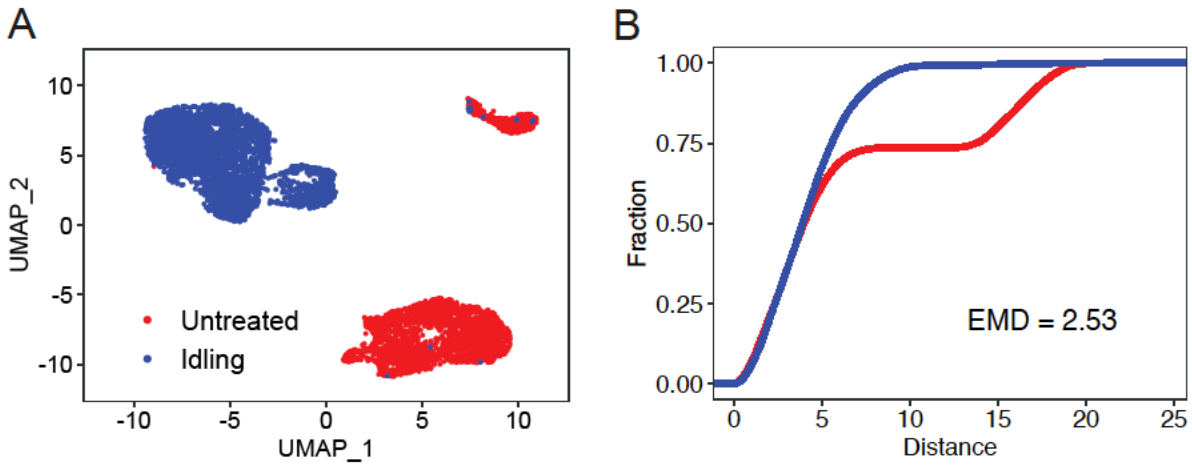

Figure 32 | ***Idling cells represent a convergent, yet still heterogeneous transcriptomic state.***
(A) UMAP projection of untreated and idling single-cell transcriptomes. 6410 cells are shown, with an approximately equal split between conditions. (B) CDF of pairwise cell distances (random sampling of 15,000) on the UMAP space in *A*. An EMD was calculated between the distributions.

To determine the biological factors that differentiate the axes that separate populations, we perform differential expression analysis between treatment conditions, clusters, and combinations to determine biological interpretation of UMAP axes. Although no clear biological processes separate UMAP_2, UMAP_1 could be separated based on cell cycle stage, even after cell cycle regression (within the context of treatment condition; see Methods). To simplify the interpretation, cell cycle stages (G1, G2M, S) were simplified into nondividing (G1) and dividing (G2M/S) states and overlay it on the UMAP plot (FIG. 33A). A first look indicates nondividing and dividing cells are represented in both untreated clusters, but clearly separate idling clusters (see Methods). Therefore, we classified cell cycle state proportional representation across untreated and idling clusters (FIG. 33B). Within the untreated

population, across both clusters, we see a gradient of cell cycle states, with nondividing cells representing the smaller UMAP_1 values and dividing cells the larger values. The large cluster (~84% of all untreated cells) does have a larger proportion of nondividing cells (~35%) than the small cluster (~16%). However, in the idling state, clusters separate solely based on cell cycle state, with the large cluster (~83% of all idling cells) composed of nearly all nondividing cells, while the small cluster has a large majority (~97%) of dividing cells.

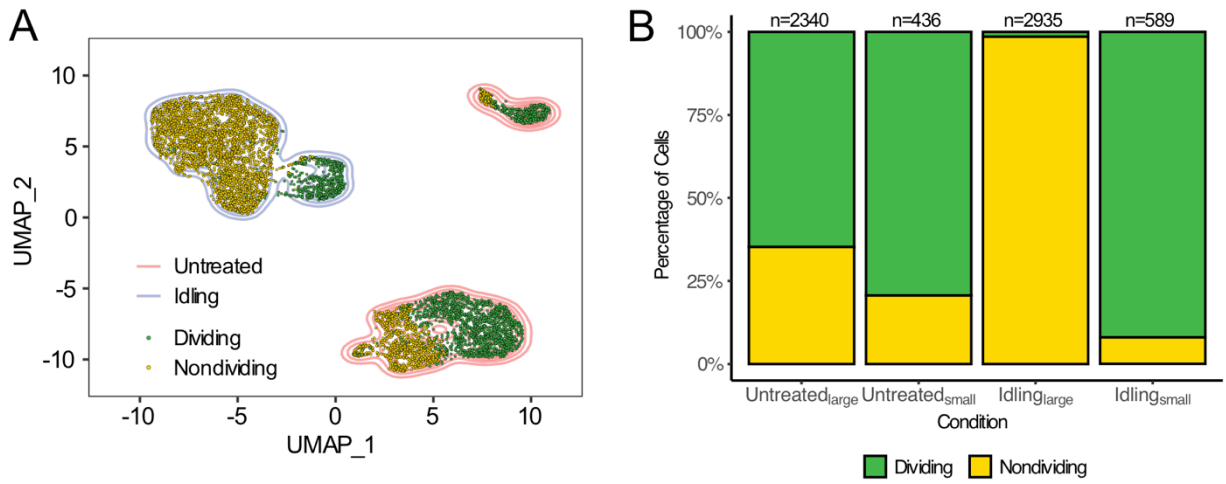

Figure 33 | *Cell cycle state separates major clusters in idling cells.*
(A) Overlay of cell cycle state (see Methods) on UMAP projection of single cell transcriptomes (colored contours represent information in *FIG. 32A*)**.** (B) Relative proportion of cells in cell cycle state for major untreated and idling clusters. The total number of cells in each cluster (n) is noted above each bar.

This result is interesting because the clusters seem to be maintained after BRAFi treatment, even down to the relative proportion of the population, but shifted in the transcriptomic space (see next paragraph for further rationale from transcriptomic lineage analysis). Considering that dividing cells represent a majority of both untreated clusters, but separate clearly between clusters in idling cells, it seems likely that the idling phenotypic distribution (i.e. epigenetic landscape) may have shifted upon BRAFi. Further, the decrease in the proportion of dividing cells from untreated (~67% total) to idling (~15%

total) maintains that the idling state drug-tolerant persister cells have a reduction in cell division events, but not a total loss, as predicted in Chapter III. It also seemingly indicates that the population idling state is due to a combination of quiescence/senescence and balanced division/death, unifying previous claims[12,21,69]. In addition, it supports our previous claims that the BRAF-mutant melanoma cell lines, specifically SKMEL5, are a *population state* after treatment, meaning that multiple cell states coexist even after treatment. It may be that these cell cycle states broadly reflect proliferation states, with a large quiescent/senescent state, and a small actively dividing state. Importantly, in this view, cells can interconvert between these states, which we will further investigate in the next section.

Lineage re-equilibration to the drug-modified idling transcriptomic landscape is predictive of the corresponding drug response

To investigate the lineage distribution across the states in each population, we overlay barcoded clonal lineages on the underlying transcriptomic distributions (four representative examples in FIG. 34A). We then quantify the relative proportions of each barcode (top 25 from *FIG. 31*) in the dividing cell cycle state for untreated and idling cells (FIG. 34B). Interestingly, in untreated cells, lineages exhibit a tight distribution across the dividing cell cycle state (FIG. 34B, top). This result is contrary to what one might expect without taking into account epigenetic variability and transitions between epigenetic states, i.e., isolated lineages *should* fall in distinct transcriptomic regions. However, the likely explanation is barcoded lineages quickly equilibrated in the untreated distribution, spreading across the epigenetic landscape. Therefore, we can conclude that the transcriptomic regions that they originate from are not very stable (i.e., shallow basins in the epigenetic landscape). The lineage distributions in the idling state have a slightly more nuanced behavior, with more variable occupancies across cell cycle states (FIG. 34B, bottom). However, a closer looks shows that the proportion of idling cells in the dividing cell cycle state is predictive of lineage response to BRAFi (FIG. 31). Lineages like barcodes 1 and 5 have a smaller relative proportion

of cells in the idling dividing state, corresponding to a smaller relative fraction after BRAFi. Conversely, barcodes 2 and 13 have a larger proportion of cells in the idling dividing state, which results in a larger relative fraction after treatment. This result is consistent with a short-term, relative lineage distribution shift as the barcodes re-equilibrate to the new idling epigenetic landscape. In the long term, lineages presumably reach a dynamic equilibrium where they reflect the proportions of cells in each cell state, similar to the untreated condition.

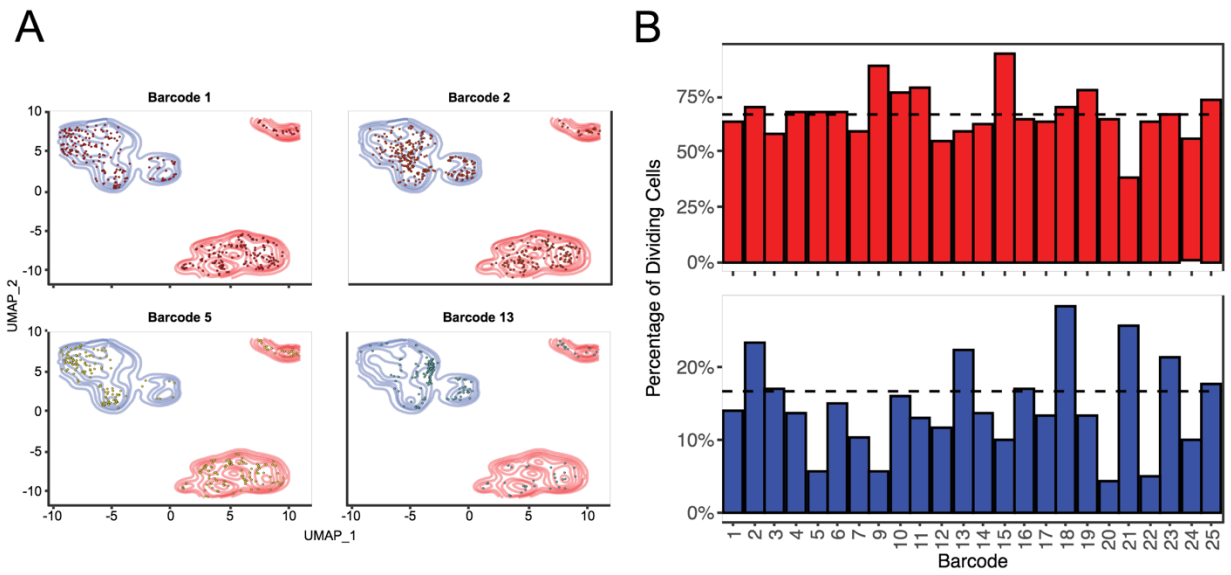

Figure 34 | *Lineage distribution across cell cycle states are reflective of clonal dynamics.*
(A) Projections of lineage transcriptomic distributions on UMAP projection in *FIG. 32A*. Lineages correspond to colored dots, while contours reflect treatment condition. (B) Proportion of cells in the dividing transcriptomic state for the top 25 most abundant barcodes. Dashed line represents average of all barcodes.

To confirm single-cell transcriptomic phenotypes in response to BRAFi, we subjected previously isolated clonal sublines (SKMEL5 SC01, SC07, and SC10; see Chapter III) to bulk RNA sequencing (RNA-seq) at three time points over the course of BRAFi (0, 3, and 8 days post treatment). These populations were chosen because they represent the extremes of the drug-response distribution (SC01 – most negative short-term proliferation rate; SC10 – most positive short-term proliferation rate) and an average

representative (SC07 – near zero short-term proliferation rate). All three clonal sublines eventually adopt the long-term net-zero proliferation rate characteristic of idling (3-8 days; see FIG. 25B). Bulk transcriptomics data of the clonal sublines were projected into reduced dimensionality space by PCA (see Methods, FIG. 35A). Interestingly, clonal sublines predominately vary on PC1 prior to BRAFi treatment. In the short term (day 0 → day 3), sublines predominately change on PC2, and maintain the overall variance from baseline. From the short-term to long-term response (day 3 → day 8), cells converge on both PC1 and PC2. A differential expression analysis was performed between untreated (day 0) and idling (day 8) bulk RNA-seq data to identify transcriptomic signatures characteristic of the transition to idling. Differentially expressed genes (see Methods) were input into a gene ontology (GO) over-enrichment analysis to determine processes that are upregulated in the idling state (see Methods). Interestingly, GO terms associated with ion transport and homeostasis are upregulated in idling cells (FIG. 35B). This result will be further explored in later sections.

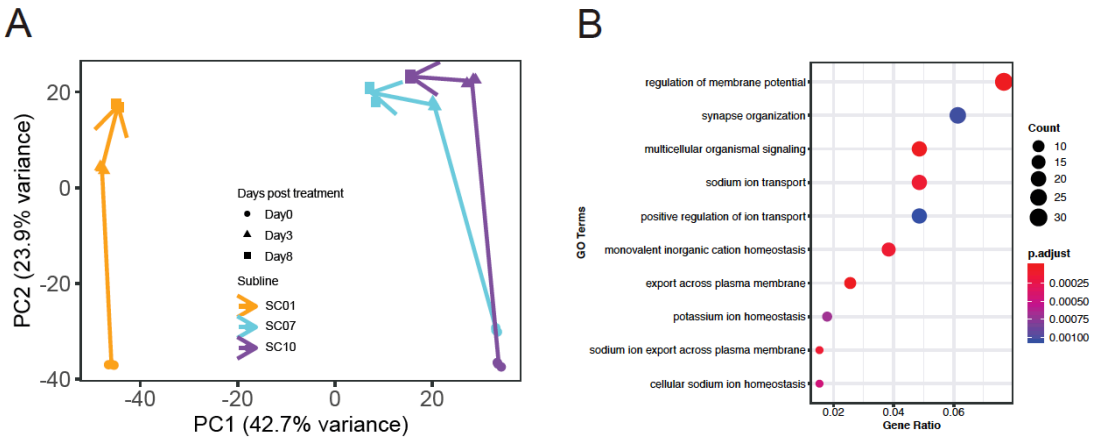

Figure 35 | *Bulk transcriptomics on single-cell derived subclone response to BRAFi.*
(A) PCA projection of subclones (SC01, SC07, and SC10) at multiple times (0, 3, and 8 days) in BRAFi. Each condition (e.g. SC01 at day 0) was completed in triplicate. Lines are drawn between the centroids of triplicates across the time series. (B) GO over-representation analysis on differentially expressed genes with increased expression between untreated and idling. GO terms with the top 10 largest gene ratios (see Methods) are shown.

Together, these results qualitatively agree with the lineage-resolved single-cell transcriptomics data, in that BRAFi seems to induce a change in the transcriptomic states (i.e., epigenetic landscape) into a more convergent transcriptomic space. Since bulk measurements muddle subpopulation separation, these results could indicate either that clonal sublines' transcriptomic signatures are becoming closer upon prolonged BRAFi (i.e. independently becoming more homogeneous) or that the underlying landscape has shifted (i.e. cells originating from the same subline are equilibrating to the new landscape). Considering the lineage-resolved transcriptomics data, the latter is more consistent. The drug-induced shift seems to be tied to a re-equilibration of ion channel transport and homeostasis, which we attempted to validate with different data modalities below.

Epigenomic data shows a broad shift upon BRAFi treatment and has strong connection with transcriptomics

To understand the connection between transcriptomic and epigenomic phenotypes, as well as validate the functional findings, we performed by bulk Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) on untreated and idling cells. We first calculated the fragment size distribution in both populations (FIG. 36A). Both populations show enrichment for nucleosome-free and mono-nucleosome fragments, indicative of a successful experiment. Fragments were then aligned to a reference genome and peaks were identified (see Methods). Peaks were compared to identify intersections and unique peaks (FIG. 36B). Approximately twice as many peaks were identified in idling than untreated, but could be attributed to duplicated reads (see Methods).

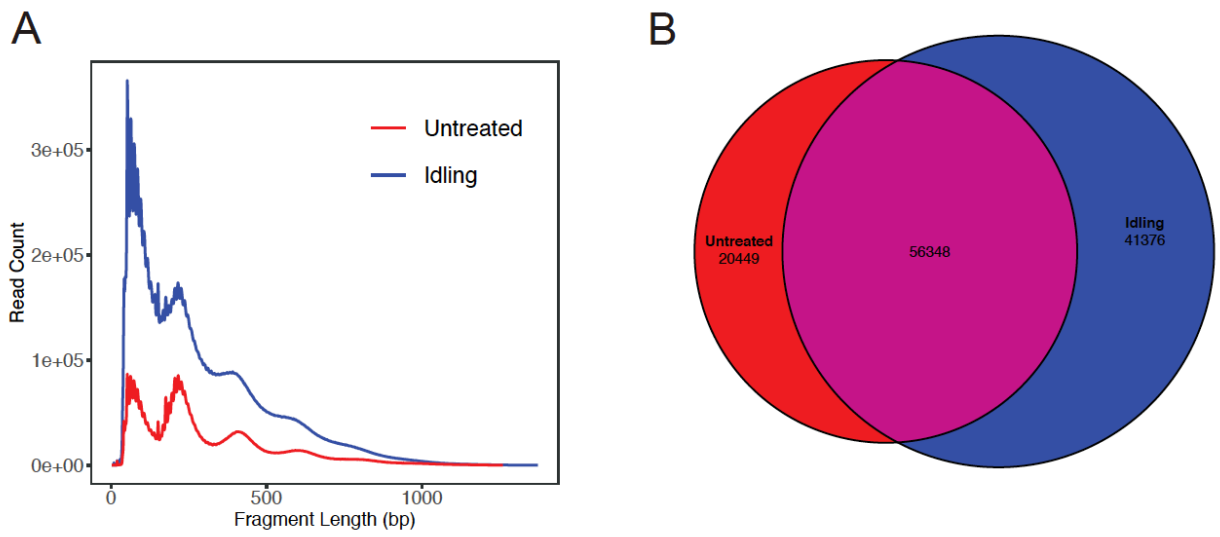

Figure 36 | *Quality control of bulk epigenomics ATAC-seq data.*
(A) Insert size distribution of aligned reads from ATAC-seq data on untreated and idling cells. Both conditions follow traditional nucleosome patterning. (B) Venn diagram of ATAC-seq identified peaks of open chromatin.

Unique and shared peaks were normalized to the transcription start site (TSS) to determine the distribution of binding loci (FIG. 37A), and used to quantify the peak feature distribution (FIG. 37B). Idling peaks have much fewer proximal features (e.g., promoter regions) and more distal elements (e.g., distal intergenic and intronic regions) compared to the untreated condition, which seems to suggest that idling cells are the result of dynamic epigenetic changes in response to treatment. Unique peaks were also assigned to a corresponding gene using the nearest gene paradigm, and genes associated with unique genes were input into a GO over-representation analysis (like the single-cell transcriptomics data). GO over-representation identifies ion transport and activity as clear differentiators of idling (FIG. 37C), which compliments findings in transcriptomics data. This connection between transcriptomics and epigenomics indicates a clear, conserved shift in the epigenomic signature upon treatment with BRAFi. Additionally, the connection between idling GO terms across data modalities suggests a new “epigenetic landscape” is formed upon treatment with BRAFi, as seen in transcriptomic and epigenomic landscape “reflections.”

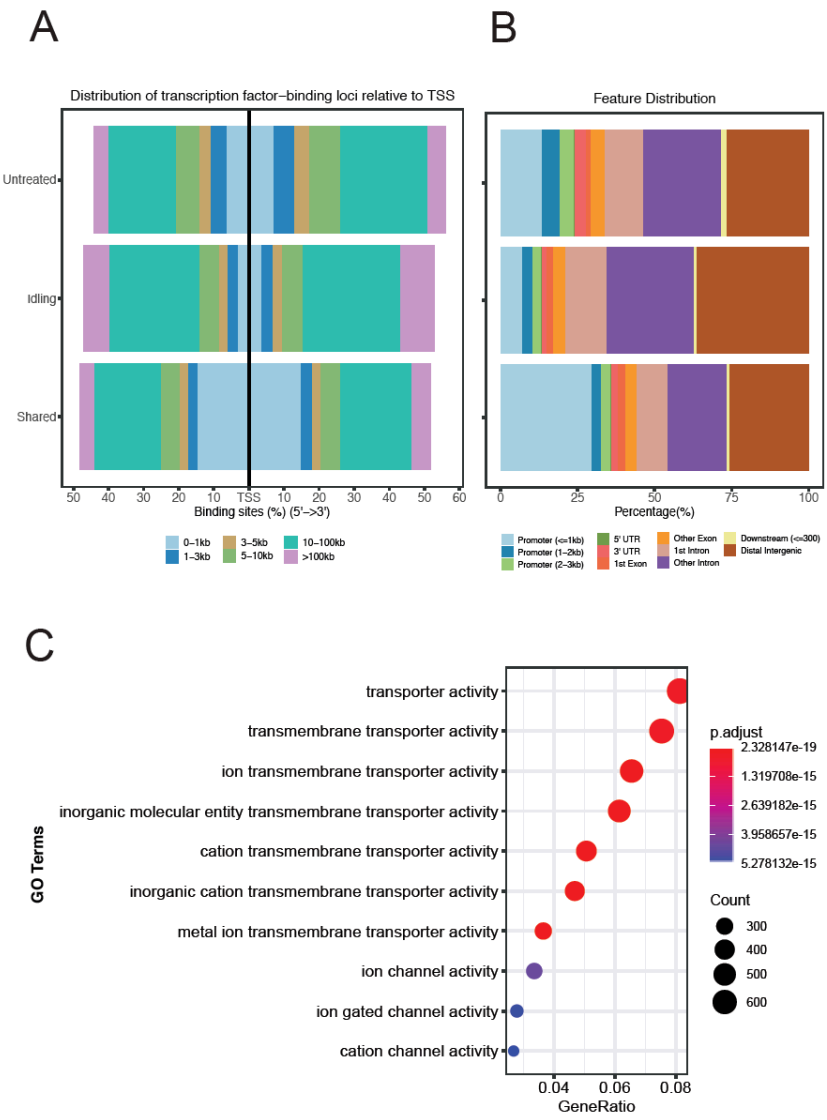

Figure 37 | *Molecular epigenomic profiling of the idling state.*
Alignment of peaks to the TSS allows for prediction of epigenomic features. (A) Peak binding site distribution for untreated, idling, and shared peaks. X-axis represents kb distances from the TSS. (B) Predicted feature distribution in untreated, idling, or shared peaks. Peaks were assigned to features based on proximity to TSS (see Methods). (C) GO over-enrichment analysis of genes associated to unique idling peaks. The top 10 terms with the largest gene ratio are shown.

Verification of ion channel molecular mechanism

Considering both the transcriptomics and epigenomics data point toward ion channel activity as a major connection point in idling, we intended to test the baseline ion channel flux exhibited in idling versus untreated cells. To do this, we used a previously described calcium flux assay[169] that measures the

amount of chelated calcium that accumulates in the intracellular space upon treatment with various agonists. In the first assay, we use Ionomycin as an agonist, which in this assay specifically facilitates calcium transport from intracellular stores (FIG. 38A). Here, the maximum amplitude of the untreated cells reached almost twice the height of the idling cells, and both were followed by a re-equilibration period. This suggests that idling cells have a lower flux capacity than untreated. A possible mechanism is that high expression of these ion channels in idling (FIG. 35B) means that more channels exist on idling cell *compartment* (i.e. mitochondria) membranes. Therefore, the idling calcium ion concentration in cytosol is higher than untreated cells prior to agonist treatment, meaning cells have lower potential flux from the compartments to the cytosol. To garner a more specific mechanism, we perform the same assay with Thapsigargin, an endoplasmic reticulum (ER) stress active transport inhibitor (FIG. 38B). Here, we see a slightly more nuanced result, with similar short-term responses to the agonist but divergent long-term responses. The lack of a second peak in the long-term idling response to Thapsigargin indicates possible impairment of store operated calcium entry (SOCE), which has been implicated previously in melanoma disease progression[170]. Regardless of the mechanism, these results confirm differences in ion channel activity in idling cells, as predicted by transcriptomics and epigenomics data.

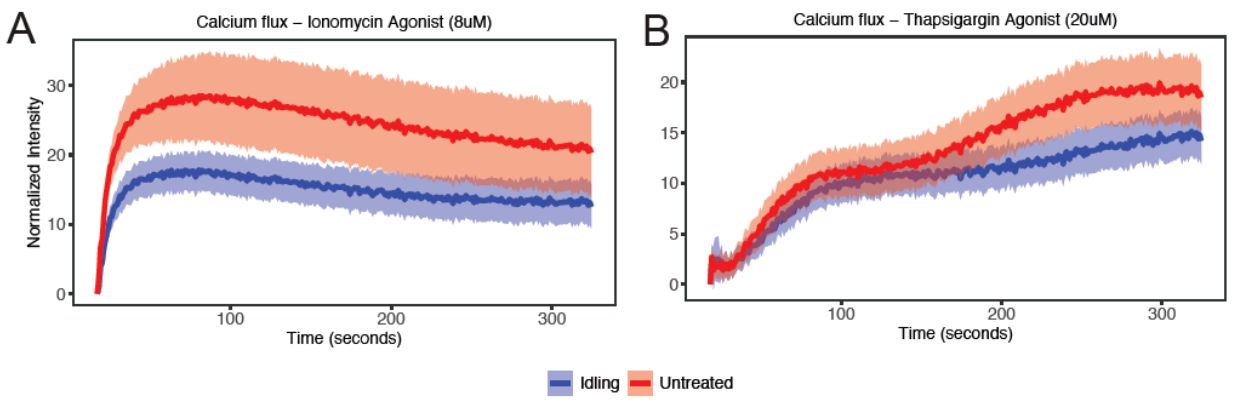

Figure 38 | *Untreated and idling cells have differential ion flux when challenged by agonists.*
Calcium flux assays for (A) Ionomycin and (B) Thapsigargin agonists. Trajectories were normalized to the time of agonist addition. Data points were captured every second for 320 seconds. Means are shown as a solid line, while the shaded envelope represents one standard deviation.

Idling BRAF-mutant melanoma cells are susceptible to ferroptotic death

BRAF-mutant melanoma cells have a previously established convergent gene signature upon prolonged BRAFi treatment (FIGS. 32A and 35A), but the molecular implications of that gene signature remain unclear. Bulk RNA-seq and ATAC-seq data suggest that ion channel activity is an important characteristic of idling cells (FIGS. 35B, 37C, and 38), but there is not a clear way to target ion channel activity in a therapeutic manner. Previous publications[171] have indicated a connection between ion channels, cancer signaling, and ferroptosis, a type of necrotic cell death. Thus, induction of ferroptosis may provide a potential way to eradicate idling cancer cells. Recent publications on the idling state[36,133,134] have indicated that BRAFi induces mitochondrial oxidative stress through reactive oxygen species (ROS), which is also known to result in ferroptosis. Interestingly, ferroptotic cell death has been marked in the literature more by sensitivity to ferroptosis inducers than specific molecular indicators[69,170]. However, some biological processes do seem to have an effect on ferroptosis, including glutathione (GSH) regulation, polyunsaturated fatty acid (PUFA) synthesis, and iron regulation.

To create a ferroptosis molecular signature, genes associated with these three processes were curated from the from the KEGG database (see Methods) and used to categorize clonal subline transcriptomic signatures. Using genes already identified in the earlier differential expression analysis (see Methods, FIG. 35), we further reduced that significant gene set to those also in the KEGG ferroptosis signature, and further classified genes according to GSH, PUFA, and iron regulation (see previous paragraph). Using this reduced signature, gene expression z-scores were compiled into a heatmap to indicate changes over time in BRAFi (FIG. 39). An initial observation is that the clonal sublines have vastly different gene expression scores, with SC01 exhibiting a markedly lower score across all genes over the course of treatment with BRAFi. Second, we see large shifts in gene expression from pre- to post-treatment, providing some credence to ferroptosis as a potential cell death target. Interestingly, these clones have relatively similar scores by the 8 days in BRAFi, suggesting this signature may be characteristic of the idling state and pointing toward a potential way to target idling cells.

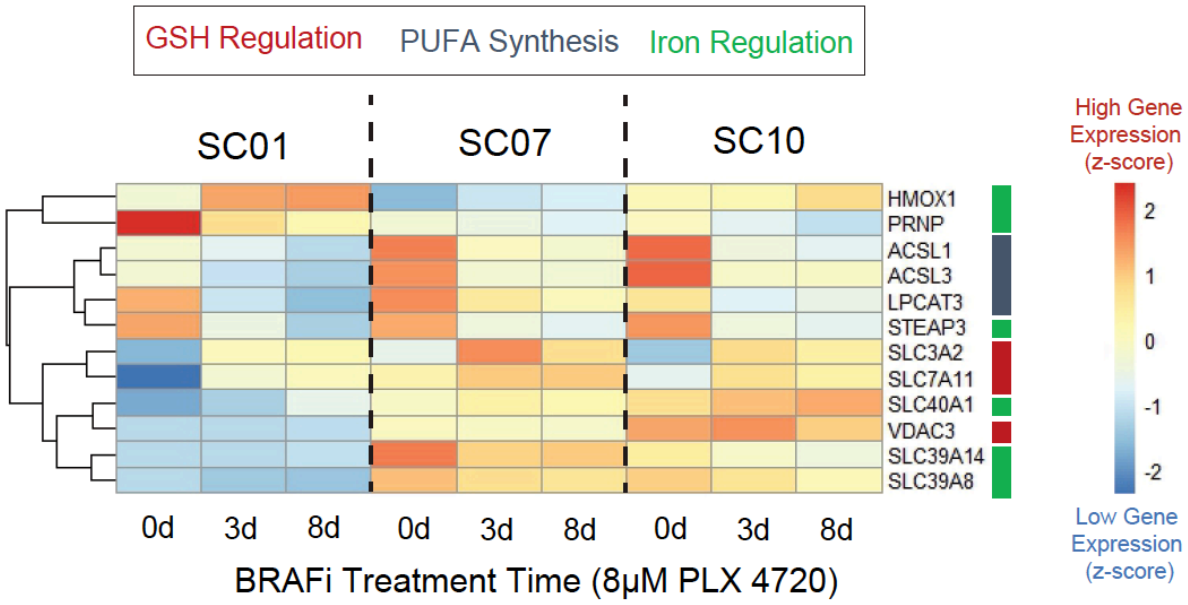

Figure 39 | ***Ferroptosis gene expression of clonal sublines across BRAFi time course.***
Gene expression values (z-scores; see Methods) for multiple clonal sublines (SC01, SC07, SC10) across the BRAFi treatment time course (0, 3, and 8 days post treatment). Genes are further broken down into three contributing groups (GSH – glutathione regulation; PUFA – polyunsaturated fatty acid synthesis; Iron regulation).

Erastin and RSL3 (see Methods) are ferroptosis inducers that have been known to modify glutathione metabolism and prevent cells from removing ROS and addressing lipid peroxidation, which leads to increased cell death[69,133]. Erastin works by targeting system X_c, a cysteine-glutamate antiporter on the cell membrane, while RSL3 targets the downstream GPX4, a key regulator of glutathione oxidation[172] (FIG. 40A). We subjected both drug-naïve and idling cells to Erastin and RSL3 treatment (DIP rate dose response curve, see Methods) in order to identify a potential mechanism . Interestingly, Erastin did not induce cell death in untreated or idling cells (FIG. 40B). However, RSL3 had a major impact on idling cells compared to untreated, massively increasing drug potency (FIG. 40C; blue to red solid line). Since Erastin did not actively target idling cells as well as RSL3, it became clear that idling sensitivity was more localized to a region of glutathione signaling than originally thought (FIG. 40A). To further clarify the important signaling processes, we performed a drug rescue experiment (see Methods) with ferrostatin-1

(Fer-1), a ferroptosis inhibitor, on SKMEL5 cells treated with the sequential BRAFi + RSL3 (FIG. 40C; dashed lines). Fer-1 inhibits lipid free radicals (FIG. 40A), reducing opportunities for ferroptotic death[172]. The Fer-1 treatment rescued the drug-response behavior except at very large doses (uM range) in both treatment conditions (FIG. 40C; solid to dashed lines), but is specifically important in idling (FIG. 40C; red solid to red dashed line) because it shows that idling drug sensitivity is associated with a narrow region of glutathione signaling and lipid peroxidation that makes them particularly susceptible to ferroptotic cell death. Given these results, it is possible that RSL3 may be a candidate for sequential therapy (post-BRAFi) to eradicate cells that persist BRAFi and eventually acquire resistance (i.e. the idling state).

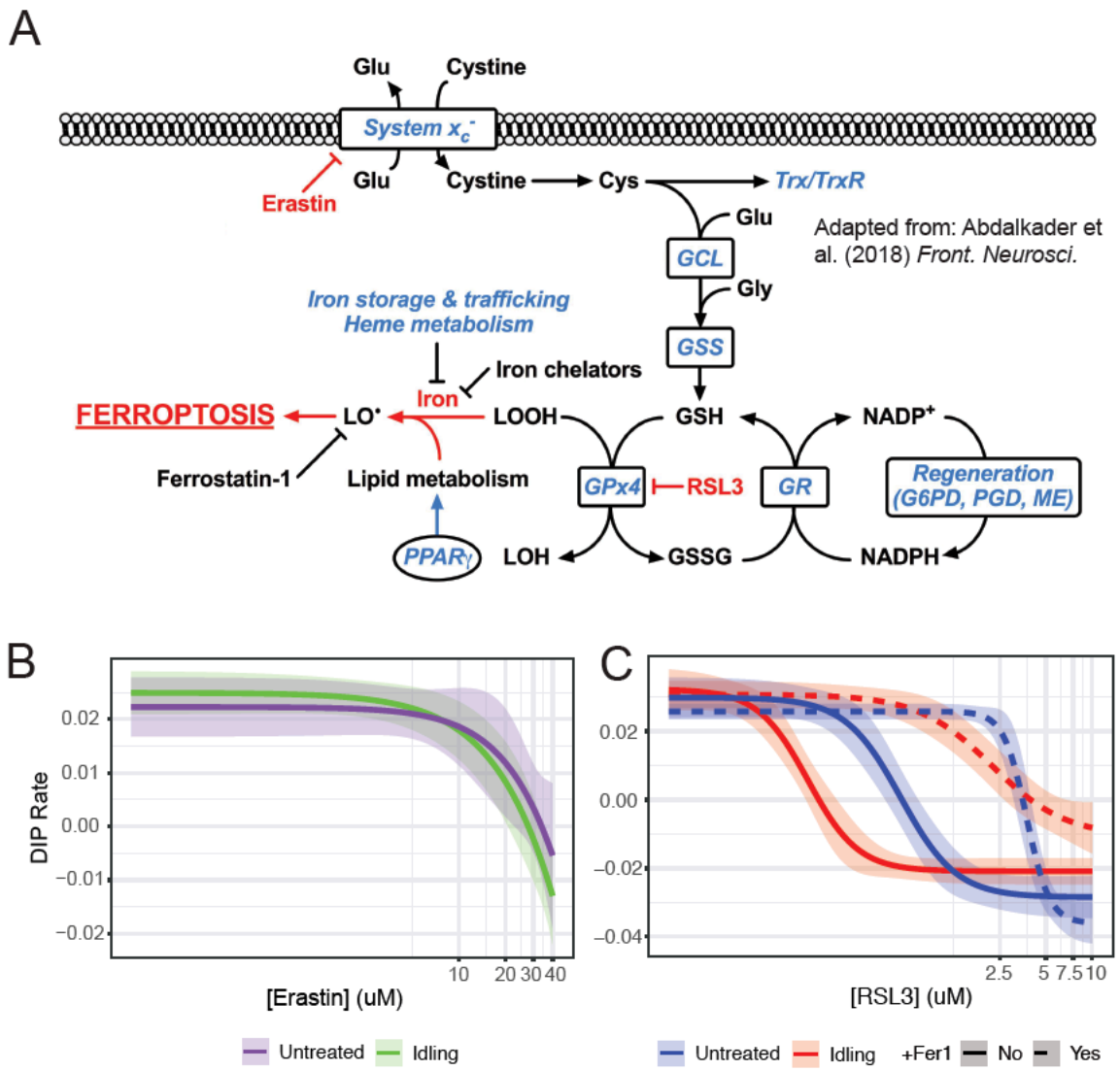

Figure 40 | *Idling cell populations are susceptible to ferroptotic cell death.*
(A) Schematic of signaling upstream of ferroptosis cell fate determination. Adapted from *Abdalkader et al.* (2018)[172]. (B) DIP rate dose response curves for untreated and idling cells treated with ferroptosis inducer Erastin. (C) DIP rate dose response curves for untreated and idling cells treated with ferroptosis inducer RSL3 (solid lines). Rescue experiments on the post-treated cells using ferroptosis inhibitor Fer-1 are noted with dashed lines.

Methods

Cell culture and reagents

SKMEL5 cell line was purchased from ATCC© and labeled with either a fluorescent histone H2B conjugated to the monomeric red fluorescent protein (H2BmRFP) and a cellular barcoding library (see

next section) or H2B conjugated to the green fluorescent protein (H2B-GFP). SKMEL5 single-cell derived clones were selected and derived by limiting dilution as described previously. Cells were cultured a mixed media of DMEM and Ham F-12 media (DMEM:F12 1:1, catalog no. 11330-032), supplemented with 10% fetal bovine serum (FBS). Cells were incubated at 37°C, 5% CO_2, and passaged twice a week using TrpLE (Gibco). Cell lines and sublines were tested for mycoplasma contamination using the MycoAlert™ mycoplasma detection kit (Lonza), according to manufacturer's instructions, and confirmed to be mycoplasma-free. BRAF inhibitor PLX 4720 (analog to Vemurafenib), ferroptosis inducers RSL3 and Erastin, and ferroptosis inhibitor Fer-1 were obtained from MedChem Express (Monmouth Junction, NJ) and solubilized in dimethyl sulfoxide (DMSO) at a stock concentration of 10mM and stored at -20°C. Cell lines were originally stored at -80°C, then moved into liquid nitrogen.

Cellular barcoding

Barcode Library Creation: Cellular barcoding library was constructed by cloning a guide RNA (gRNA) library of barcodes into a CROP-seq-BFP-TSO vector as previously described[32]. The vector was engineered in a that barcodes could be captured by isolation and amplification (barcode sampling) or mRNA capture in a scRNA-seq experiment. gRNAs were built as a 20 nucleotide sequence of 4 nucleotides identical among all barcodes followed by a 16 strong-weak (SW) paired nucleotides (i.e., XXXXSWSWSWSWSWSWSW). The SW pairing of the barcode sequence was designed to prevent PCR amplification bias, and has a maximum complexity of 2^{16} (~65k unique barcodes). The barcode library vector was used to produce lentiviral libraries using a lipofectamine transfection of HEK293T cells. Media containing lentiviral particles were collected at 48 and 72 hours post-transfection, pooled, filtered through a 0.45 um Nalgene syringe filter (Thermofisher) and concentrated using a 50 mL size-exclusion column (Millipore) by centrifugation at 2200 RCF at 4°C for 2 hours. Concentrated virus was stored in -80°C.

Experimental Setup: SKMEL5 cells were seeded in a 6-well plate at ~1x10^6 cells per well in 2.5 mL culture media. Cells were transduced with the barcoded CROP-seq-BFP-TSO-Barcode_sgRNA lentivirus using 0.8 µg/mL in each well and a multiplicity of infection (MOI) of 0.05. 24 hours after incubation, transduction media (containing polybrene) was exchanged for fresh culture media. 48 hours after incubation, barcoded cells were isolated by FACS and subsequently cultured until confluence in a T-150 dish, and then cryopreserved. Cryopreserved cells were thawed in a T-25 dish and scaled up for ~2 weeks in two separate sets. The first set of thawed cells were treated with 8µM PLX4720 (and an untreated control) for 8 days and subjected to barcode sampling (see *Barcode Sampling* section). The second set was plated in 3 T-75 flasks (parallel replicates) and independently treated with 8µM PLX4720 (or untreated control) for 8 days and subjected to scRNA-seq by the 10X genomics Chromium platform (version 2 chemistry, see *RNA single-cell transcriptome sequencing* section). In both cases, treated cells had media and drug replaced every 3 days. Untreated cells were expanded completely over the course of the time course (i.e. no cell splitting).

Barcode Sampling: After PLX4720 treatment for 8 days (or no treatment expansion), cells in the first set were pelleted for genomic DNA (gDNA) extraction using the DNeasy Blood and Tissue Kit (Qiagen) per manufacturer's instructions. Barcode sequences were amplified for each replicate by polymerase chain reaction (PCR; 98°C for 30 seconds, followed by 22 cycles of denaturation - 98°C for 10 seconds, annealing - 63°C for 30 seconds, extension - 72°C for 10 seconds, and a final extension of 72°C for 5 minutes) using primers containing flanking regions and Illumina adapter index sequences. 2µg gDNA was used in each PCR reaction, and a combination of 5 distinct pooled forward primers were utilized to minimize sequencing error. Reactions were purified using a 1.8x AMPure XP bead (Beckman Coulter) cleanup. Reaction products were confirmed using agarose gel confirmation (band at ~215bp). The resulting libraries were quantified using a Qubit fluorometer (ThermoFisher), Bioanalyzer 2100 (Agilent) for library profile

assessment, and qPCR (Kapa Biosciences Cat: KK4622) to validate ligated material, according to the manufacturer's instructions. The libraries were sequenced using the NovaSeq 6000 with 150 bp paired end reads as sequencing spike-ins (targeting ~200k reads). RTA (version 2.4.11; Illumina) was used for base calling and MultiQC (version 1.7) for quality control.

Barcode Sampling Analysis: Barcodes were identified from amplified sequence reads by trimming flanking adapter sequences. Barcodes abundances were totaled and normalized to library read depth, resulting in counts per million (CPM). Barcodes less than 100 CPM were removed from the analysis. Numbers of unique barcodes were calculated based on barcodes that exceeded the 100 CPM threshold. Overlaps among experimental replicates were calculated to determine the proportion of barcodes shared across different runs. Total barcode abundance (including low abundance barcodes) were calculated using the $\log_{10}$ of barcode RPM for each replicate. Relative barcode fraction was calculated for each sample across three replicates. Log_2 fold change of the idling to untreated mean barcode fractions was calculated for all barcodes above the CPM threshold.

RNA single-cell transcriptome sequencing

Data Collection: After PLX4720 treatment for 8 days (or no treatment expansion), cells in the second set were prepared targeting ~3000 cells per sample, washed and resuspended in 0.04% bovine serum albumin (BSA) in phosphate-buffered saline (PBS). Cell suspensions were subjected to 10X Genomics single-cell gene expression protocol (version 2, 3' counting) in two separate wells (untreated and idling), according to manufacturer's guidelines. Single-cell mRNA expression libraries were prepared according to manufacturer's instructions. Due to the nature of gRNA barcoding library construction, mRNAs resulting from gRNA barcodes were captured along with other mRNAs. Libraries were cleaned using SPRI beads (Beckman Coulter) and quantified using a Bioanalyzer 2100 (Agilent). The libraries were sequenced using

the NovaSeq 6000 with 150 bp paired-end reads targeting 50M reads per sample for the mRNA library (including barcode library). RTA (version 2.4.11; Illumina) was used for base calling and MultiQC (version 1.7) for quality control. Gene counting, including alignment, filtering, barcode counting, and unique molecular identified (UMI) counting was performed on each library using the *count* function in the 10X Genomics software *Cell Ranger* (version 3.0.2) with the GRCh38 (hg38) reference transcriptome.

Transcriptome Analysis: Cell Ranger output two single-cell gene expression matrices, for untreated and idling cells. Since cells were prepared and processed in parallel, no computational batch correction was performed. Seurat[154] was used to perform gene expression analysis. The *SCTransform* function was used to regress out mitochondrial gene expression (percent.mt), number of features (genes; nFeature_RNA), number of RNA molecules in the cell (nCount_RNA), and cell cycle variables (S.Score and G2M.Score). Feature selection was performed according to Seurat guidelines, using a variance stabilizing transformation of the top 2000 most variable features. Data was normalized and scales according to Seurat guidelines. Data between conditions were combined and visualized using the Uniform Manifold Approximation and Projection[141,142] (UMAP) dimensionality reduction algorithm as implemented in Seurat. Pairwise distances were calculated between cells in each condition, and plotted as a cumulative density function (CDF). An Earth Mover's Distance (EMD) was calculated between 15000 randomly sampled pairwise distances using the *wasserstein1d* function in the *transport* R package. Clustering was performed in the joint UMAP space using the default Seurat implementation, a shared nearest neighbor (SNN) modularity optimization based method. Differential expression was performed using the Seurat *FindMarkers* function to compare treatment conditions and clusters. Differentially expressed genes (DEGs, adjusted-p < 0.05) were input to a gene ontology (GO) over-enrichment analysis using *clusterProfiler*. GO analysis identified cell cycle as a major factor separating major idling clusters, which was not present for untreated cluster separation. Therefore, using cell cycle scores, a cell cycle phase (G1,

G2M, S) was assigned to each cell, which was further simplified into dividing (S, G2M) and nondividing (G1), which we call cell cycle *state*. Cluster proportion was calculated by cell cycle state to quantify the differences between clusters.

Barcode Analysis: After calculation of scRNA-seq gene expression matrices, barcode abundances were incorporated to the matrices by mapping gRNA lineage barcodes to their associated 10X cell barcodes. First, unmapped scRNA-seq BAM files were cleaned to only include the mRNA transcript ID, scRNA-seq cell barcode, and scRNA-seq unique molecular identifier (UMI). Mapped scRNA-seq BAM files (3' heavy) were cleaned to only include the mRNA transcript ID and lineage barcode (from gRNA library). Unmapped and mapped subsets were merged on the mRNA transcript ID to assign a lineage barcode to each cell barcode and UMI. The resulting merged dataset was paired down to a single cell barcode – lineage barcode pair, which was appended to each cell in the gene expression matrix as a metadata tag. Barcode abundances were totaled across all cells in the experiment that captured a barcode, and strongly reflected barcode sampling relative abundances and fold changes upon treatment. Barcodes were overlaid on UMAP projections of scRNA-seq data, and further categorized into the dividing and non-dividing transcriptomic states (see *RNA single-cell transcriptome sequencing: Transcriptome Analysis* subsection above). Total number of cells from each barcode were tallied across each transcriptomic cell cycle state, and a percentage (relative to each barcode) in each state was calculated.

Bulk RNA transcriptome sequencing

Data acquisition: Total RNA was isolated from untreated SKMEL5 single-cell derived sublines, each in triplicate, using Trizol isolation method (Invitrogen) according to the manufacturer's instructions. RNA samples were submitted to Vanderbilt VANTAGE Core services for quality check, where mRNA enrichment and cDNA library preparation were done with Illumina Tru-Seq stranded mRNA sample prep kit.

Sequencing was done at Paired-End 75 bp on the Illumina HiSeq 3000. Reads were aligned to the GRCh38 human reference genome using HISAT[173] and gene counts were obtained using featureCounts[157].

Data analysis: RNA-seq data was analyzed using the DESeq2[158] R package. Cells with less than 18 reads per condition were removed, according to DESeq2 vignette recommendations. Counts were transformed using the regularized logarithm (rlog) normalization algorithm. PCA was performed on the rlog normalized data using the *prcomp* function in R. The path between time series data points was visualized as a line between subline-time point replicate means in the PCA space. Differential expression analysis was performed in DESeq2 using a model design to quantify both changing variables and their interaction (~ subline + treatment time + subline:treatment time). DEGs across sublines between untreated (pre-treatment, day 0) and idling (day 8 post-treatment) were identified (adjusted-p < 0.05, log2 fold change > 2) and input into a GO enrichment analysis (*clusterProfiler*) to identify GO terms associated with biological process (BP), molecular function (MF), and cellular component (CC) GO types. A ferroptosis gene signature was obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG). Differentially expressed genes (q < 0.005, fold change > 1.5, z-score scaled) that overlapped with the KEGG ferroptosis gene signature were plotted using a z-score. Differentially expressed ferroptosis genes were classified according to relationship to glutathione (GSH), polyunsaturated fatty acids (PUFA), and iron regulation.

Bulk ATAC epigenome sequencing

Data acquisition: Data was collected using the omni-ATAC protocol for bulk ATAC sequencing (ATAC-seq). After PLX4720 treatment for 8 days (or no treatment expansion), cells from the first set (in parallel to barcode sampling data collection) were pelleted at 50k cells and resuspended in a cold ATAC-seq resuspension and lysis buffer containing NP40 (0.1%), Tween20 (0.1%), and Digitonin (0.01%) and incubated on ice. A resuspension buffer was added (0.1% Tween20, no NP40 or Digitonin) to wash out the

lysis reaction. Cells were pelleted and resuspended in a transposition mix (5x Tris-DMF, PBS, 1% digitonin, 10% Tween20, nuclease-free H_2O), including transposase Tn5, followed by a 30 minute incubation at 37°C, with shaking to enhance tagmentation. After 30 minutes, the reaction was stopped by adding a DNA binding buffer (Zymo) and purified using a DNA Clean and Concentrate kit (D4004, Zymo). The final product was eluted in nuclease-free H_2O. PCR amplification was performed on the eluate with an NEBNext 2X High Fidelity PCR Mix (NEB, M0541S) N7, and N5 index sequencing primers (extension at 72°C for 5 minutes; denaturation at 90°C for 30 seconds; 12 cycles: denaturation at 98°C for 10 seconds, annealing at 62°C for 30 seconds, extension at 72°C for 30 seconds; final extension at 72°C for 5 minutes). The PCR product was purified with the Zymo DNA Clean and Concentrate kit, and eluted in 22uL nuclease-free H_2O. ATAC-seq PCR libraries were visualized by agarose gel electrophoresis for an initial check of the nucleosome ladder pattern (bands ever ~150 bp). Libraries were also quantified using a Qubit fluorometer (ThermoFisher), Bioanalyzer 2100 (Agilent) for library profile assessment, and qPCR (Kapa Biosciences Cat: KK4622) to validate ligated material, according to the manufacturer's instructions. The libraries were sequenced using the NovaSeq 6000 with 150 bp paired-end reads (untreated: ~160m, idling: ~130m reads). RTA (version 2.4.11; Illumina) was used for base calling and MultiQC (version 1.7) for quality control.

Data Analysis: Reads were trimmed using *cutadapt* (paired-end) to remove primer sequences, and aligned to hg38 reference genome using the *bwa mem* function in Burrows-Wheeler Aligner (BWA, version 0.7.17). Aligned reads were sorted and duplicates were marked using Picard (version 2.17.10). Untreated reads had more detected duplicates (~78% compared to ~32% in idling). Reads were deduplicated, leaving much fewer reads in untreated compared to the idling library. Reads were further cleaned according to sequence quality guidelines. Insert sizes were plotted from the output of *InsertSizeMetrics* after deduplication in Picard. Peaks of open chromatin were called using the *MACS2 callpeak* function according

to recommended guidelines for ATAC-seq data (BAM paired-end method, q-threshold: 0.05, no MACS2 model, shift: -100, extension size: 200). Peaks were subjected to a further round of quality control and cleaning using *ChIPQC* (peak mapping, peak duplication, blacklist peak detection), and blacklisted peaks were removed. Peaks were converted to consensus counts using the *runConsensusCounts* function in *soGGi*. Intersections of and unique cleaned peaks were determined and visualized as a Venn diagram using the *vennDiagram* function in the *limma* package. Unique and intersection peaks were annotated with the nearest neighbor genes using the *annotatePeak* function and hg38 transcriptome in the *ChIPseeker* package. These peaks were also re-aligned to the transcription start site (TSS) for each gene, and average profiles of read subsets across all genes were obtained (nucleosome-free, mono-nucleosome, and di-nucleosome; normalized to the TSS). Peaks were classified based on closeness to the TSS, and assigned to predicted feature (e.g., promoter, UTR, exon, intron, downstream, distal intergenic). Genes associated with unique and intersections of peaks were input into a GO enrichment analysis for BP, MF, and CC GO types. Transcription factor (TF) footprinting in the region around TSSs was performed on untreated and idling unique peaks for key transcription factors TFs.

Calcium flux assays

Untreated and idling cells were seeded on 384 well plates (Greiner) on the day prior to the experiment targeting 1000 cells per well. On the day of the experiment, media is removed from cells, washed with HBSS with Ca^{2+} and Mg^{2+} (Corning), and incubated in assay buffer (HBSS and HEPES (Gibco)) and Fluo-8 dye (1µM, Teflabs) for 1 hour. Cells are then washed with the assay buffer to remove dye and plates are input into the Panoptic experiment (WaveFront Biosciences). Cells were then treated with an agonist (Ionomycin - 8 µM; Thapsigargin - 20 µM; derived from experimental dose response curves in previous experiments). Upon agonist addition, plate images are immediately taken every second for 320 seconds. Images are converted to fluorescence values based on the well intensity (i.e. the green

fluorescence in cells; Fluo-8 is an AM dye, so it is cell impermeable). Well intensity is compiled across technical replicates (12 per condition), and normalized to the time of agonist addition.

Conclusion

Understanding tumor plasticity in the context of treatment evasion is key to targeting recurrent cancers. Chapter III identified a novel 'idling' state in BRAF-mutant melanoma, which was attributed to epigenetic plasticity in response to BRAFi. This chapter provides a means by which cells utilize epigenetic plasticity to enter the idling state. Interestingly, short-term variability in the response to BRAFi is driven by lineage-dependent occupancy of idling transcriptomic states, which are broadly categorized into dividing and nondividing, i.e. more cells that transition into the idling dividing state leads to larger proportions of that barcode. However, all lineages eventually adopt the net-zero DIP rate characteristic of idling, suggesting that they eventually diversify across epigenetic states in dynamic equilibrium. Omics analyses of the SKMEL5 cell line point toward increased ion channel activity in idling cells, suggesting that molecular process is one of the major factors influencing the post-BRAFi landscape. Ion channel flux differences were performed to verify omics molecular findings, and literature suggested these differences opened up the door to cell death by ferroptosis. Ferroptosis inducer RSL3 showed marked increased cell death of idling cells, and rescue by ferroptosis inducer Fer-1 suggests glutathione signaling is a key point of interest for future studies targeting the idling state.

Together, this chapter suggests that a systems-level understanding of cancer plasticity can be utilized to identify potential secondary therapeutics and harness the consequences of that plasticity. Furthermore, although idling may be a reservoir for potential resistance mutations, it may also be a limited time frame where the disease can be targeted for better patient outcomes. We hypothesize ion channel signaling may serve as a network control point for the epigenetic landscape after BRAFi, and targeting that control point (with ferroptosis inducers or otherwise) may again modify the landscape so that all basins

will eventually go extinct, eradicating the tumor. While the analyses here do point to potential hypotheses for killing idling cells, the underlying theory provides a potential mechanism for targeting epigenetic plasticity in general. Namely, sequential therapies may be the best approach for targeting residual disease in tumors, as post-treatment tumors may have unique vulnerabilities associated with their epigenetic landscapes rather than specific single states.

CHAPTER 5

Conclusion

Discussion

Despite significant progress in the treatment of oncogene-addicted cancers in recent decades, single targeted therapies have not developed into the "cure" many had hoped. Although these therapies can increase short-term patient survival, responses are variable and tumors invariably recur[9,11]. Many mechanisms have been proposed to understand tumor recurrence, but a large portion of these cases remain unexplained. Many studies that investigate tumor drug responses fail to grasp the extent of variability present in tumors prior to and in response to treatment. This variability, which comes from a variety of sources[27,28,49,74], is one of the major contributors to recurrence. Inability to combat tumor variability leaves a large gap in our understanding, which results in the continuing poor patient prognoses. In this work, we quantify various aspects of tumor variability in multiple *in vitro* models of tumor heterogeneity and plasticity. We present these findings through the lens of a common heterogeneity framework[23,85], which considers different types of heterogeneity jointly. Throughout this work, we make distinctions between different types of variability, and discuss their implications for drug sensitivity and evasion. The progress made in this work provides a way to interpret the complex array of variability in cancer, which in turn paves a path toward improved therapeutic strategies to treat this recalcitrant disease.

Many modern cancer therapies focus on targeting specific genetic mutations within a tumor. Recent studies have shown that a complex interplay between genetic and non-genetic factors likely plays a key role in the failure of targeted treatments[24,129]. In this Chapter II, we investigated genetic and non-genetic sources of variability in an *in vitro* tumor heterogeneity model comprising multiple versions (VU, MGH, BR1) and single-cell derived sublines of the NSCLC cell line PC9 that exhibit a wide range of different

responses to EGFR inhibition (FIG. 5). Given their histories and how each was derived, we had good reason to believe that the cell line versions were genetically distinct. This was validated using WES and CNV detection, which showed significant mutational differences among them (FIGS. 7-9). Distinct transcriptomic features were also identified by scRNA-seq (FIG. 10) and connections to the underlying genetic states were established by a comparison between GO terms enriched in each data modality (FIGS. 11, 17, and 18). We then isolated seven sublines from PC9-VU that exhibited differential responses to EGFR inhibition (FIG. 6). Clonal drug response assays (FIG. 6B) and scRNA-seq analysis (FIG. 15) showed significant overlap with the PC9-VU parental cell line. WES and CNV detection revealed substantially less genomic variability among six of the seven sublines relative to the cell line versions (FIGS. 12-14) and GO similarity analysis indicated a weak, if any, connection between genomic and transcriptomic states in these sublines (FIGS. 16-18). For the other subline, DS8, the results were dramatically different: DS8 harbors significantly more unique and IMPACT mutations than the other sublines (FIG. 12B-C), there is clearer evidence for copy number variation (FIG. 14), its single-cell transcriptomic state is substantially distinct from the other sublines (FIG. 15), and it displays a much stronger connection between genomic and transcriptomic states (FIGS. 16-18). Finally, stochastic simulations revealed that colony growth dynamics for six of the seven sublines can be explained as a population with a single cell state experiencing probabilistic division/death decisions (FIGS. 19-20). For DS8, a second cell state had to be included in the model to reproduce the bimodal DIP rate distribution observed experimentally (FIG. 21).

In order to interpret our results, we utilize the theoretical framework for tumor heterogeneity discussed previously[23,84,85,92] (FIG. 2). As explained in this view of tumor heterogeneity, tumors may comprise multiple genetic states, each of which has an associated epigenetic landscape with $\geq$1 quasi-potential energy basins corresponding to phenotypic states, across which cells can transition driven by intrinsic (e.g., gene expression) or extrinsic noise sources. Within our *in vitro* tumor model, the PC9 cell line versions (VU, MGH, BR1) correspond to the different genetic states. We assert that four of the PC9-

VU sublines (DS3, DS6, DS7, DS9), based on their genomic similarity (FIGS. 12-14), transcriptomic distinctiveness (FIG. 15), weak genetic-to-transcriptomic correspondence (FIGS. 16-18), and monoclonality (FIGS. 19-20), likely correspond to basins within the epigenetic landscape associated with the PC9-VU genetic state. In contrast, DS8 appears to harbor a distinct genetic state that emerged out of PC9-VU at some point in the past in the absence of selective pressures. We come to this conclusion based on its resistance to EGFRi (FIG. 6), genomic (FIGS. 12-14) and transcriptomic (FIG. 15) distinctiveness from the other sublines and all three cell line versions, strong genomic-to-transcriptomic correspondence (FIGS. 16-18), apparent polyclonality (FIG. 21), and lack of the resistance-conferring mutation (EGFR-T790M) found in PC9-BR1 (FIG. 13B). A schematic of these conclusions is summarized in FIG. 41.

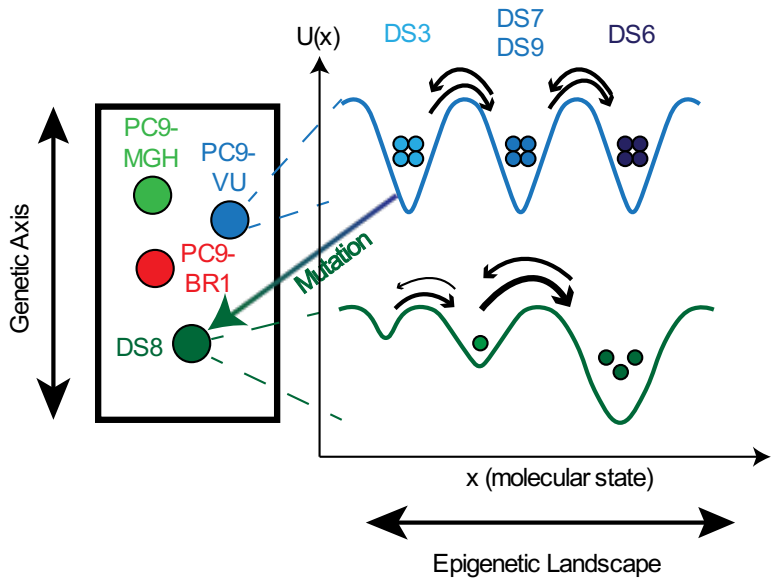

Figure 41 | *Schematic summary of our interpretation of the results of analyses on PC9.*
Cell line versions PC9-MGH (*light green*), PC9-BR1 (*red*), and PC9-VU (*blue*) represent different genetic clones within our *in vitro* tumor model. Each genetic clone has an associated epigenetic landscape, including PC9-VU with at least three distinct basins corresponding to DS3, DS7/DS9, and DS6. Also depicted is the acquisition of a genetic resistance mutation out of the PC9-VU population detected within the DS8 subline (*dark green*). Note that it is unclear at present whether the resistance mutation was acquired before or after the subline was established (see below).

We conclude based on our genomic and transcriptomic analyses that the DS8 subline harbors a novel genetic resistance mutation that emerged from the PC9-VU parental population at some point in the past. However, it remains an open question as to whether this mutation occurred after or before the subline was established (FIG. 42). If it occurred after, it could be the case that DS8 actually harbors two distinct genetic states, the original PC9-VU state and the new emergent state (FIG. 42A). However, if that were the case we would expect to see some DS8 cells in the same region of transcriptomic space as PC9-VU, which we do not (FIG. 15A). Another possibility is that the emergent genetic state in DS8 outgrew the PC9-VU state, leaving behind an isogenic population (i.e., a "selective sweep"). This would explain the lack of DS8 cells in the PC9-VU region of transcriptomic space. However, there does not appear to be a discernable difference between the out-of-drug proliferation rates for DS8 and the other sublines (FIG. 6A), which would seem to discount this possibility. We cannot entirely preclude this prospect, however, as it is possible that the difference in proliferation rates is simply too slight to detect in our current data.

Conversely, if the mutation in DS8 arose prior to the subline being established, we would expect some mutant cells to remain within the PC9-VU parental population. We do, in fact, see a small number of PC9-VU cells in the region of transcriptomic space where DS8 cells reside (compare FIG. 10A and 15A). Interestingly, we also see some DS9 cells in this region (but not DS7, despite the significant overlap between the two populations; FIG. 15A). This could indicate that DS9 actually comprises two epigenetic basins, a deep one where most of the cells reside and a shallow one in the region of transcriptomic space where DS8 resides. It is possible that PC9-VU cells transition frequently between these two basins and that at one point in the past one cell in the shallow basin randomly acquired a genetic mutation that caused it to get "locked into" that basin. A progeny of that cell could have been isolated to start the DS8 subline (FIG. 42B). This scenario would be consistent with the findings of Shaffer et al.[136], although without the selective pressure that drives the mutation in their case. However, if this is the case, given that DS8 proliferates at essentially the same rate as the other sublines in the absence of drug, why are those cells

so rare (~2%) within the PC9-VU parental population? One would expect that population to grow out in the same way that DS8 did. One possible explanation might be that cell-cell interactions[174] destabilize the mutant cells when in culture together with PC9-VU parental cells. This is an intriguing idea that would require additional experimentation to verify, perhaps using DNA barcoding[31–33,175].

Finally, there is also the added complication that DS8 exhibits a bimodal DIP rate distribution under EGFRi, which we conclude from our stochastic simulation analysis (FIG. 21) indicates that DS8 harbors (at least) two distinct cell states. As discussed in Chapter II, the model is agnostic as to whether these two states are genetically identical or not, it merely requires they have distinct DIP rates. Thus, it does not preclude the possibility that DS8 is isogenic, as would be the case in two of the scenarios discussed above (*i.* mutation after subline establishment, followed by selective sweep; *ii.* mutation prior to subline establishment). However, the region of transcriptomic space in which DS8 cells reside is small compared to PC9-VU and there are no obvious subpopulations present (FIG. 15A). It is also difficult to reconcile why the left mode of the DS8 DIP rate distribution overlaps so significantly with the PC9-VU parental and subline distributions (FIG. 6B) despite the lack of overlap in transcriptomic space. These remain open questions that we hope to address in future investigations. We emphasize, however, that none of the issues raised here changes our main conclusion presented that DS8 harbors a novel genetic resistance mutation that emerged from PC9-VU at some point in the past in the absence of selective pressures. We summarize these conclusions in a schematic illustrating the different sources of cell state variability we hypothesize are operating within the PC9 family of cell lines and sublines (FIG. 42).

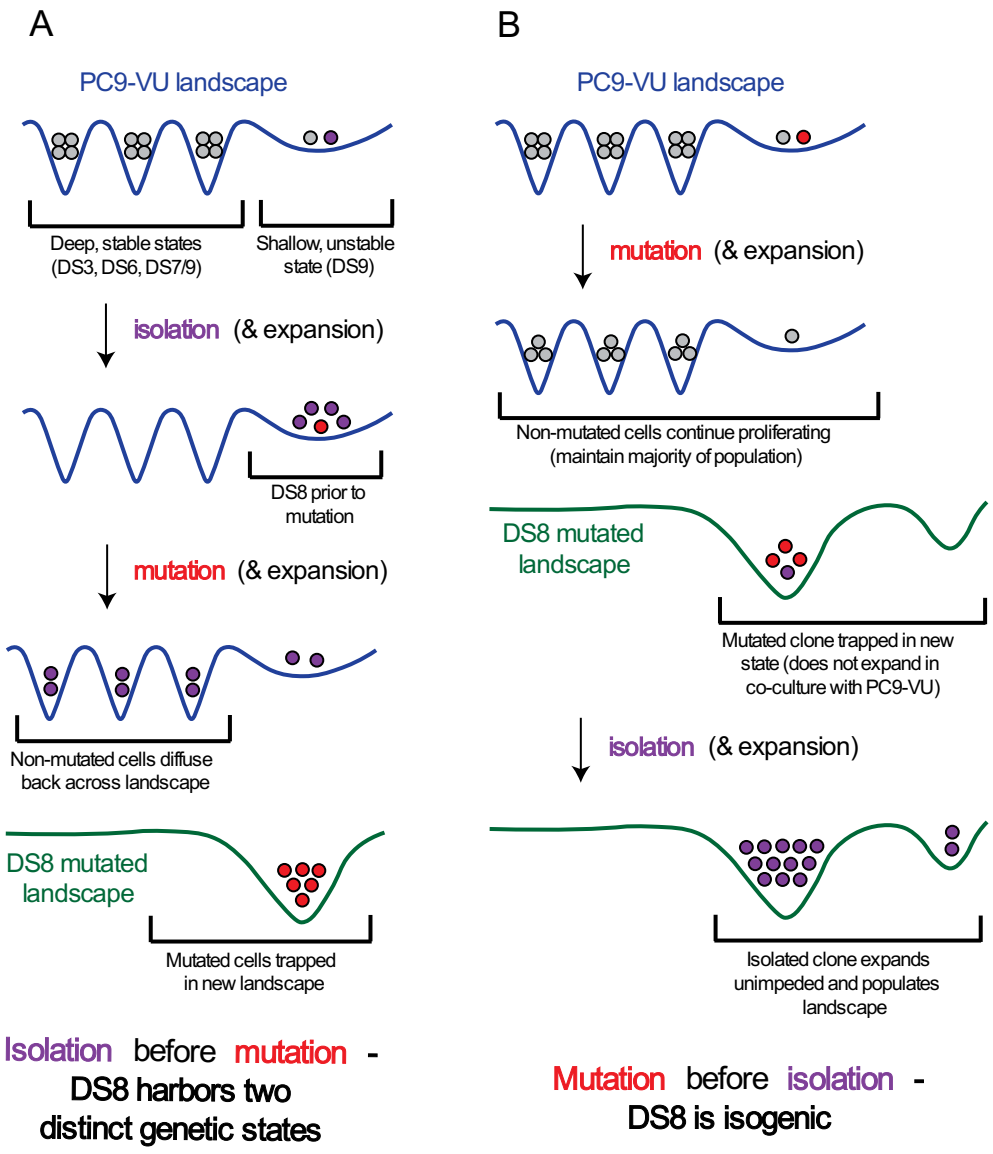

Figure 42 | *Potential explanations for cell state heterogeneity in DS8.*
(A) *Multiple genetic states hypothesis*. In this scenario, a genetic resistance mutation was acquired after the DS8 subline was established. Assuming the mutant state does not outgrow the original genetic state (i.e., a "selective sweep"), both genetic states should co-exist within the subline. (B) *Single genetic state hypothesis*. In this scenario, a genetic resistance mutation emerged within the PC9-VU parental population and a cell containing that mutation was isolated to establish the DS8 subline. To explain our single-cell transcriptomics data, we hypothesize that cell-cell interactions between mutant and PC9-VU cells increases the death rate for mutant cells, making them a small proportion (<2%) of the total PC9-VU population.

This view of tumor heterogeneity, as a three-tiered amalgamation of genetic, epigenetic, and stochastic factors, is not yet broadly accepted within the cancer research community[85] because of the

focus on mutations as the main mechanism of cancer development. This mutation-centric framework has been helpful for many cancers, but can only be taken so far given the proliferation of new knowledge in the field of cancer heterogeneity. A primary goal of Chapter II has been to provide evidence for an alternative framework of cancer heterogeneity that includes genetic mutations, but also epigenetic variation and intrinsic stochasticity (FIG. 2). However, we believe that numerous reports in the literature are also consistent with this view. For example, Ben-David et al.[26] showed that numerous "strains" (comparable to our cell line versions) of human cancer cell lines, obtained from different institutions, display extensive genetic heterogeneity. Moreover, genetically similar strains exhibit similar transcriptomic signatures and drug-response profiles. Thus, they argued that cancer cell lines can drift genetically when kept in culture independently, consistent with our results for the PC9 cell line versions. Our conclusion that the drug-resistant DS8 genetic state emerged spontaneously from PC9-VU in the absence of selective pressure aligns with observations by Ramirez et al.[12] and Hata et al.[13], who independently reported diverse resistance-conferring mutations arising in both untreated and drug-treated PC9-MGH clones. Shaffer et al.[136] described a transient, transcriptionally-encoded pre-resistance state in two BRAF-mutant melanoma cell lines that cells can transition into and out of in the absence of drug. We hypothesize that this pre-resistant state may constitute a basin within a BRAF-mutant melanoma epigenetic landscape, similar to our single cell-derived sublines (DS3, DS6, DS7, DS9) that we allege occupy the PC9-VU epigenetic landscape. The veracity of this hypothesis depends on how long cells reside in the pre-resistant state (its stability) and, correspondingly, whether the state is heritable by progeny cells.

One insight from the study of heterogeneity in the PC9 cell line family was that PC9-VU sublines seem to occupy deep basins in the epigenetic landscape (i.e. the drug-response phenotypes remain independent in continued drug treatment). These deep basins are ideal for the study of heterogeneity, as they are more stable and therefore exhibit less plasticity. BRAF-mutant melanoma, as remarked above, has been shown to display remarkable plasticity in the presence of drug treatment, allowing for drug

evasion[35,132,176]. An early rationale for this plasticity is that BRAF-mutant melanoma is represented by shallow epigenetic landscapes, where transitions happen frequently and allow for cells to find drug tolerant states[29,131]. In Chapter III, we address that very topic. We report that sustained BRAF inhibition (> 100 hours *in vitro*) induces entry of BRAF-mutant melanoma cell populations into a non-quiescent 'idling' state of balanced death and division, characterized by a near-zero proliferation rate[36]. Idling occurs in both parental and clonal populations, independent of differences in initial short-term responses, and is both drug induced and reversible, consistent with nongenetic drug tolerance described in earlier reports. Although a balanced state of cell proliferation and apoptosis has been described as tumor dormancy, the idling population state is distinct because it occurs in the context of drug response. Idling was not previously described, possibly because drug-response assays tend to be performed over short observation times (72-96 hours), and proliferation rates are not usually measured as done here[117,118]. Taken together, our findings are not easily explained within the existing paradigms of drug resistance or tolerance. In particular, cell populations that initially expand but then transition into the idling state (SKMEL5 in FIG. 22A; A375 in FIG. 23A; SC10 in FIG. 25B) can neither be the result of selection of rare, preexisting resistant clones nor of the acquisition of resistance-conferring genetic mutations. Furthermore, idling populations are not due to confluence (FIG. 22C) or quiescence alone (FIG. 22D). This begs the question as to why an apparently thriving cell population would cease expanding and enter a less proliferative state of balanced death and division.

To garner insights into these complex dynamics, we propose a kinetic model (FIG. 26A) in which a cell population is composed of multiple discrete, interconverting subpopulations, each of which is characterized by a DIP rate quantifying its net proliferation in a drug. The model is most easily understood within the framework of epigenetic landscapes, where cell subpopulations are associated with basins of attraction and phenotypic state transitions with traversals of quasi-potential-energy barriers. An implicit assumption of the model is that an epigenetic landscape exists in the absence of drug, defined by the

genetic background of the cell. Over time, cells within an isogenic population (e.g., a cell line) stochastically diffuse across basins in this landscape. This drug-naive "phenotypic drift" sets the initial cellular occupancies of each basin. Upon drug addition, the epigenetic landscape is modified in a drug- and dose-dependent manner. With the cFP assay, we have an experimental platform for quantifying initial cell occupancies in the drug-naive landscape based on measured DIP rate distributions. The central hypothesis of this work is that drug-treated cell populations re-equilibrate over this new drug-modified landscape; the short-term population-level drug response is a reflection of this re-equilibration process, and the idling state constitutes the final equilibrated state of the population. This theoretical framework explains why populations of single-cell-derived clonal sublines respond differently to a drug in the short term (different initial numbers of cells in each basin) but identically in the long term (exposure to the same landscape topography), as we report here. The reversibility of the idling phenotype for parental cell populations (FIG. 23B) is explained by a return to the drug-naive epigenetic landscape upon drug removal and re-equilibration back to the original cell occupancies. This is consistent with recent work showing that intermittent addition and withdrawal of vemurafenib leads to sequential periods of tumor shrinkage and growth, which forestalls development of drug resistance in BRAF-mutant melanoma cell populations[70]. Differential dynamics across cell lines are explained in terms of variations in the topography of drug-modified landscapes (FIG. 27B) that are set by the genetic backgrounds of the cell lines. An important consequence is that each cell line achieves idling in a slightly different way— with varying proportions of regressing, stationary, and expanding subpopulations (FIG. 27B)—despite harboring a common BRAF-activating mutation (i.e., the driving addicting oncogene). Differences in the sizes of idling cell populations may explain the extreme diversity of durability, or lack thereof, in individual patients' clinical responses (i.e., the probability of acquiring resistance mutations depends on the number of cells surviving treatment). For instance, tumors that show significant early regression (e.g., WM88, WM164; FIG. 23A)

would be expected to take longer, perhaps significantly longer, to acquire secondary mutations than those that either show no initial change or expand (e.g., SKMEL5, A375; FIGS. 22A and 23A).

An important aspect to consider when quantifying tumor drug sensitivity is drug-naïve drift, represented by shifts in epigenetic state occupancies, prior to treatment[165]. To better understand how cells diversify across the epigenetic landscape, we isolated a single SKMEL5 clone (SC01) sensitive to BRAFi, and tested its BRAFi response at multiple points along its drug-naïve progression (FIG. 28A). Different responses emerged after prolonged drug-naïve diversification (FIG. 28B), suggesting that the SKMEL5 landscape may be very plastic, even in the absence of selective pressure. Further model predictions (FIG. 28D-E) and experimental validations at the single-cell level (FIG. 29) suggest that, just like cells seem to equilibrate to a drug-treated epigenetic landscape, they too continue to diversify across cellular states in the absence of selective pressure. The implications of this drift are important to determine how the cancer community approaches therapies. If tumors do drift, targeting specific cellular sub-states will not eradicate the tumor, as cells from adjacent basins will repopulate the previous basin (i.e. killing SC10-like cells will not remove the possibility of SC01- or SC07-like cells from transitioning into that basin). This property of cancer adds a significant wrinkle to developing therapeutic approaches that both treat aggressive cells and prevent the plasticity that creates these cells.

The properties predicted by *in silico* models of population dynamics suggested that transitions are rampant across BRAF-mutant epigenetic landscapes, in the presence or absence of drug treatment. We created a model schematic to interpret the effects of drug addition and removal on the underlying landscapes (FIG. 43). In short, we hypothesize that drug treatment modifies the landscape, over which the cells must re-equilibrate. After drug removal, barring any relevant genetic mutations, the landscape will return to its previous form and cells will re-equilibrate to it.

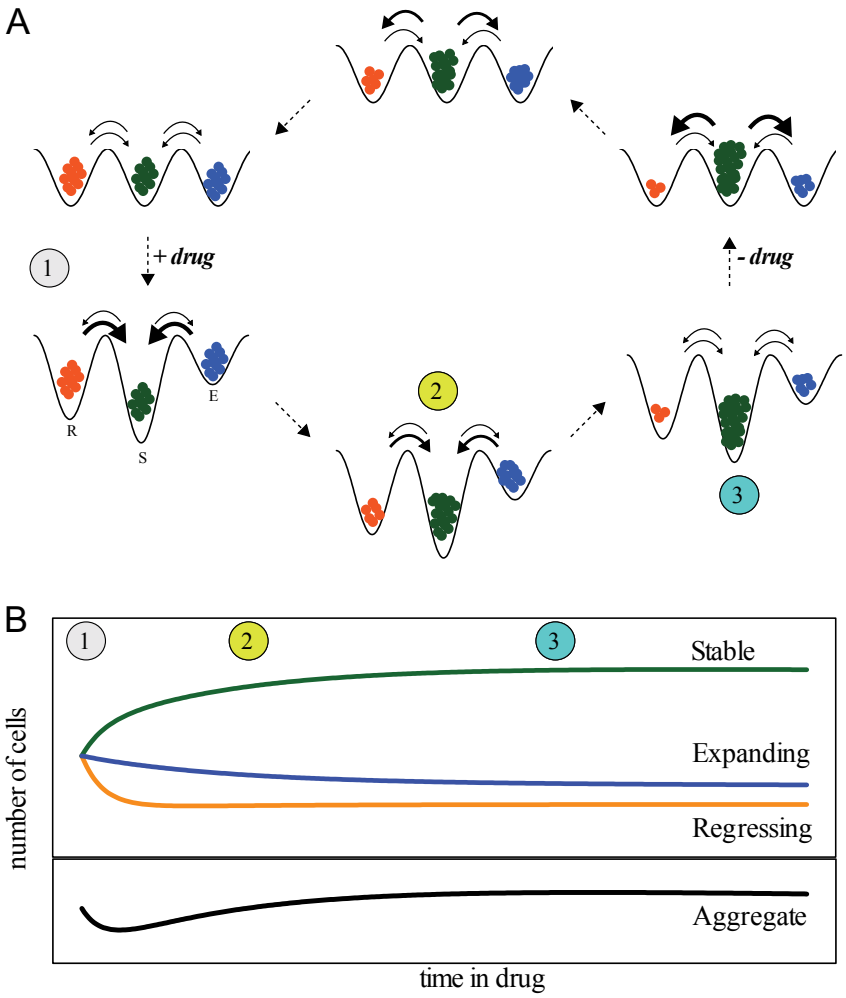

Figure 43 | ***Drug-induced and drug-free population dynamics are explained as re-equilibrations over epigenetic landscapes.***

(A) Schematic representation of a drug-addition and drug-removal cycle for a cancer cell population: (top left) cells begin in complete growth medium and are in a dynamic equilibrium across basins of a drug-free epigenetic landscape; (top left to bottom left) exposure to a drug modifies the landscape, taking the system out of equilibrium; (bottom left to bottom right) the population re-equilibrates over the new drug-induced landscape by reducing cell proportions in the regressing (R) and expanding (E) basins and increasing the proportion in the stationary (S) basin; (bottom right) the idling state corresponds to the newly achieved dynamic equilibrium; (bottom right to top right) removal of the drug reestablishes the drug-free epigenetic landscape and, again, takes the system out of equilibrium; (top right to top left) the population re-equilibrates over the drug-free landscape, returning the system to the original dynamic equilibrium. Arrow thicknesses represent relative transition rates at each stage of the re-equilibration process. (B) Clonal (top) and parental (bottom) drug-response population dynamics illustrating the connection to the different stages of the drug-induced re-equilibration process. Numbers and colors (orange, regressing; green, stationary; blue, expanding) correspond to those in the schematic above.

However, this hypothesis left a series of gaps: (1) how different are the pre- and post-treated landscapes? (2) how do cells re-equilibrate to the post-treated landscape? (3) what are the molecular properties of the landscapes? In Chapter IV, we used an *in vitro* model of melanoma tumor plasticity, BRAF-mutant melanoma cell line SKMEL5, to clarify the mechanism by which cells adopt a population "idling" state that is tolerant to BRAFi. Using cellular barcoding, we showed that idling cells result from an overwhelming majority of untreated clones, rather than clonal selection of a special idling clone (FIG. 30). Distinct transcriptomic signatures were identified that differentiate untreated and idling cells, with the idling cells represented in a more restrained transcriptomic space (FIG. 32). Importantly, both untreated and idling cells both occupied two separate transcriptomic states, one large and one small each (FIG. 33). These two states were defined by different features in each treatment condition (i.e., untreated and idling), with the idling cells separated into dividing and non-dividing cell states, as defined by each cell's location in the cell cycle (FIG. 33). Barcoded clonal lineages were distributed across both transcriptomic states in each treatment condition, but relative barcode abundances after BRAFi were correlated to the relative proportion in each idling transcriptomic state, i.e., lineages that have a larger proportion of cells in the dividing than nondividing idling state results in a larger proportion of that barcode after treatment (FIGS. 31 and 34). Single-cell derived clonal subline transcriptomics show a similar behavior, as treatment causes clones to move to a more convergent transcriptomic space as they enter idling (FIG. 35A).

Interestingly, subline differential expression analysis pointed towards ion channel activity as a characteristic of entrance into idling across sublines (FIG. 35B). Distinct regions of open chromatin were also found to differentiate idling from untreated cells, which also points toward ion channel activity (FIG. 37C). Gene ontology similarity suggest each modality shows a reflection of the underlying epigenetic landscape. Validations of ion channel flux differences (FIG. 38) led to the identification of ferroptosis as a potential avenue for idling cell death (FIG. 39). Sequential addition of ferroptosis inducer RSL3 (after BRAFi) was shown to increase drug potency for idling cells (FIG. 40C), providing potential future avenues

for drug treatment regimens. Together, these data provide evidence for a view of tumor plasticity where cells fall into basins across an epigenetic landscape, which can change in response to perturbations (e.g., drug treatment), and over which cells have to re-equilibrate. Differential short-term behaviors associated with the location where cells land in the treated landscape, and the idling phenotype represents cells equilibrated across the drug-treated landscape.

Many studies have remarked on the epigenetic plasticity as a way to understand decreased drug sensitivity[27,139], and others postulated that the epigenetic landscape as a mechanism to understand how cancer cells transition between states[36,68,132]. However, little direct experimental evidence exists showing how epigenetic state transitions lead to drug tolerance and eventual resistance. Chapter IV puts forth data that suggests a timeline for epigenetic plasticity prior to and after treatment. Tumor initiation from a single clone creates a population with the same genetic background. The genetic clone emanates an epigenetic landscape, comprised of several basins of attraction over which cells populate to create multiple cell types, each with different molecular phenotypes (FIG. 2). This "bet-hedging" strategy of epigenetic diversification in the absence of perturbations has been observed in bacteria[40] and proposed as a survival strategy in cancer[21,177]. Over time, cells in the landscape reach a dynamic equilibrium, i.e., cells can still transition between basins but the population is in a state of balance. It seems our untreated population had nearly equilibrated prior to treatment, as most barcodes had a similar proportional split between transcriptomic clusters to the overall population (FIG. 34). The introduction of a perturbation, such as an anticancer drug treatment, upends the equilibrated landscape and drops cells into a new landscape. Cells re-equilibrate to the new landscape, and adopt cell fates corresponding to the state in which they now reside. In the case of a drug with good efficacy, most cells will fall into a state of the drug-treated landscape that results in death. However, if the new landscape includes a state where cells have a positive proliferation rate in drug, the population will invariably rebound. In our case, most of the cells matriculate into the large non-dividing state upon treatment with BRAFi, but some end up in a smaller

dividing state (FIG. 34). However, some barcoded clonal lineages disproportionally fall into one of the two states (FIG. 34), leading some lineages to have increased short-term drug fitness but eventually adopt a drug-tolerant phenotype after lineages fully equilibrate to the new landscape. This occurrence is largely consistent with other oncogene-addicted cancers treated with targeted therapies[4,12,21,36,71,135].

Overall, this dissertation puts forth a framework to understand tumor variability, supported by multiple types of data, that could have a profound impact of how patients are treated in the clinic. Our quantitative approach to study tumor variability, which included drug response experiments, an array of omics technologies, lineage tracing, and mathematical modeling, captured high-level rationales for how tumors tolerate, evade, and become resistant to anticancer tumor therapies. By integrating these data across modalities and scales, we identify the properties of tumor populations at different points along disease progression that can lead to recurrence. These analyses provided a general mechanism of cancer cells: (1) prior to treatment, become heterogeneous (genetic, epigenetic, etc.) in a bet-hedging strategy for future perturbations; (2) in response to treatment, utilize plasticity between epigenetic basins to evade treatment in drug-tolerant states; (3) while evading treatment, allow some cells to continue dividing in order to acquire mutations that could provide resistance (i.e. genetic plasticity, via the baseline mutation rate). While this is an interesting mechanism, about which there remains much debate, it provides a simple and transferable approach for analyzing drug response variability in a variety of cancer systems.

Future Directions

It is now abundantly clear that a focus on single data modalities alone cannot solve the complex problems of tumor progression, metastasis, and treatment failure that continue to plague clinical oncology[178,179]. The view of tumor heterogeneity advocated in this work offers an alternative to the traditional gene-centric view and may transform how we understand and treat the disease. For example, that each genetic state has an associated epigenetic landscape with potentially numerous accessible

phenotypic states may explain why targeted drug treatments eventually fail in almost all cases[180]: some of these phenotypes may have molecular compositions that enable their survival under drug treatment. Cells pre-existing in these states (e.g., the pre-resistance state of Shaffer et al.[136]), and those that escape into them upon drug addition, may act as a refuge from which genetic resistance mutations can arise[21,36]. Alternative treatments based on targeting cancer stem cells[181] (CSCs) have also been proposed but have so far proven unsuccessful[182]. This could be because CSCs correspond to shallow basins within an epigenetic landscape; killing cells in this basin does not eradicate the basin, hence leaving it available to be repopulated by cells from "adjacent" basins. Limited success can be attributed to understanding epithelial-mesenchymal transition (EMT), which is similar in principle to CSCs in that it employs epigenetic plasticity[183,184]. In a similar fashion, transitioned mesenchymal cells, targeted by drugs, could easily be repopulated by the epithelial state. The nature of this strategy suggests that targeting a single tumor subpopulation (involved in tolerance, resistance, etc.) will not eradicate the tumor and delay inevitable recurrence, as seen in the clinic[36,185].

Consistent with these reports, we believe that the idling population presented here, and the relatively simple theoretical framework describing it, has potentially far-reaching implications for patient therapies. In particular, even in tumors with high therapeutic sensitivity, a minority of cells often survive and can persist for months or even years. This ''residual disease'' is suspected to act as a reservoir from which resistance-conferring genetic mutations, and ultimately tumor recurrence, arise[186]. We speculate that idling cancer populations may, in fact, constitute the bulk of the residual disease. Indeed, by continuing active progression through the cell cycle, idling populations are more prone to accumulate deleterious mutations and, hence, are a more fertile ground for acquiring resistance mutations than quiescent or senescent populations. Recently described ''drug-addicted'' cells[187], which can arise by either genetic or epigenetic mechanisms, may also emerge from idling cell populations. These cells are dependent upon a drug for continued proliferation such that drug withdrawal leads to initial tumor

shrinkage followed by regrowth. Within our modeling framework, drug addiction due to genetic changes would correspond to a change in the epigenetic landscape relative to that for drug-sensitive cells. Alternatively, if drug addiction is nongenetic in nature, this implies that additional basins exist within the drug-modified epigenetic landscape that are not easily accessible. In either case, idling cell populations are clearly distinct from drug-addicted populations. Applying the methods presented here to drug-addicted cells is a possible area of future investigation.

We have shown that the idling state is not a property of individual cells but rather a property of a BRAF-mutant melanoma population as a whole under prolonged BRAF-inhibition. As such, idling populations cannot be eradicated by targeting one particular subpopulation (i.e., a basin). Rather, the landscape itself must be altered (e.g., using drugs) to favor basins for regressing states over stationary and expanding states. Drug-naïve phenotypic drift has to also be accounted, as repopulation of drug insensitive cancer cells further complicates the problem. This is a significant departure from recent approaches that aimed to identify and eliminate rare cell subtypes (e.g., cancer stem cells[182,188], drug-tolerant persisters[12,21]) thought to be responsible for tumor progression and recurrence. This type of cellular reprogramming will require deep knowledge of the molecular factors that shape and define the epigenetic landscapes cancer cells inhabit. It has been suggested that a better approach, termed "targeted landscaping"[36,185], is to use drugs in combination or in sequence to alter the topography of a landscape to favor drug-sensitive states over drug-tolerant states[168]. The feasibility of such an approach is supported by multiple studies showing that resistance to one drug can confer sensitivity to another, known as "collateral sensitivity"[189–191]. However, there remains the problem of how to administer treatment. Combination therapies have been shown to have clinical efficacy, but most tumors still recur and are recalcitrant. This is likely because combinations create an entirely new epigenetic landscape that has new molecular properties with unknown vulnerabilities, making the new tumor difficult to treat. An alternative approach is sequential therapy, in which cells are able to equilibrate to the new drug-treated landscape

with the hope that it is more sensitive as a whole. In fact, it has been shown that sequential drug applications can often be more effective than up-front drug combination treatments[192,193]. Molecular analyses in our study show that the drug-treated landscape in BRAF-mutant melanoma cell line SKMEL5 is more homogeneous after treatment and has a common thread in ion channel activity across transcriptomics and epigenomics data (Chapter IV). By targeting the idling state with a sequential therapy that reduces ion channel activity, we could create another epigenetic landscape where all of the basins would have a negative proliferation rate in drug treatment and result in tumor eradication. Our attempts with ferroptosis inducers were a first step toward that goal (FIG. 40). However, future sequential drugs would need to improve efficacy in addition to the improved potency of ferroptosis inducers in idling. Although identification of such a target is a difficult endeavor, it is invariably a better option than the mutation-centric approach of targeting single states that eventually leads to tumor recurrence. Therefore, potential future studies could be aimed at identifying vulnerabilities of epigenetic landscapes that result after primary drug treatment, and finding patient-specific secondary drugs used for sequential drug treatment regimens that eradicate the entire tumor.

Although this work reveals that idling is a property of nearly all BRAF-mutant melanoma lineages upon BRAFi treatment, it would be interesting to perform matched multi-platform studies on lineages over the course of treatment to determine patterns (or lack thereof) of acquired resistance. The framework provided here states that different types of heterogeneity are relevant at different times, so an ideal study would include genomic, transcriptomic, and epigenomic profiling on lineage-resolved cancer cell populations with a common genetic background in response to long-term BRAFi treatment at various intervals (e.g. every five days for one month). Inherent in this view is that epigenetic differences would dominate in the short-term response to treatment, while genetic mutations, driven by the baseline mutation rate while cells are idling, will initiate genetic states that may become resistant in treatment. An experiment could yield the following result: (1) Exome sequencing shows no genetic difference between

lineages prior to and early in the response to treatment, while ATAC-seq and scRNA-seq shows the pre-treatment epigenetic variability indicative of differential short-term dynamics; (2) As cells enter idling, exome sequencing still shows no difference between lineages while ATAC-seq and scRNA-seq converge and lineages do not shift relative abundances; (3) Exome sequencing identifies rare and distinct genetic resistance mutations in a small minority of lineages during long-term treatment, establishing new and independent scRNA-seq and ATAC-seq epigenetic profiles associated with resistant genetic states. Together, this type of experiment would provide an explanation for how the different levels of heterogeneity enact changes across characteristic timescales. However, this type of experiment is expensive and lacks the impact of cancer research studies that utilize more clinically relevant systems, such as patient-derived material and multiple cell lines. But, these types of in-depth, multi-platform experiments could mimic a best-case scenario for personalized patient therapy, where we focus on a single patient (i.e. genetic background) and truly understand the dynamics of cancer progression, advancing mathematical models of drug-response dynamics and informing future therapeutic strategies.

The study of systems biology encourages a repeating "loop" of modeling, experimentation, and informatics. In this study, we used a top-down approach to understand cancer cell population responses to treatment, substantiating a view to consider genetic, epigenetic, and stochastic variability jointly. By design, we only completed a couple of rounds through the systems biology "loop," and therefore our models and analyses touched on different scales of biological variability but did not focus on any specific mechanisms. Further rounds through this loop could drill down into biological interpretations at the level of cellular processes, and eventually biochemical mechanism. Our GO analyses lends themselves to mathematical models of cellular processes (e.g. cancer signaling, stress response, ion channel activity) as a way to further understand phenotypic differences in cancer cell populations. From there, cellular processes could be modeled at a higher resolution as modules of protein signaling kinetics[162]. While the sequencing datasets collected here could be used calibrate these models, sensitive and specific time-

series molecular assays, such as time-series single-cell protein measurements, would be ideal to fit these models and identify the signaling dynamics that lead to drug-response phenotypes. Future modeling studies should be aimed at understanding these molecular networks at a high-level, then drilling down to biochemical signaling to understand the network control points that can be targeted to improve existing therapies.

An implicit limitation to all of the sequencing studies performed in this work is that they were performed using next-generation sequencing assays. While sequencing provides the large datasets that detect variation across the entire genome, there can be issues with the robustness of these methods, leading to false positives and targets with unknown clinical significance[194]. Other methods not utilized here, such as proteomics and metabolomics, use a different technology that has less coverage but is more sensitive and potentially more clinically relevant[195]. Additionally, the types of molecules these methods assess (i.e. proteins, metabolites) are end-products and signaling molecules that are more likely to have a functional effect than individual mutations or gene expression. Furthermore, these methods can measure cell activity through modifications like phosphoproteomics[196] and activity-based protein profiling[197]. Therefore, future multi-omics studies should consider supplementing sequencing studies with proteomic and metabolomic profiling of key molecules (predicted from sequencing analyses) to confirm findings and elucidate the effects of upstream (of protein production; DNA, RNA) changes. These additional studies could bolster clinical relevance and improve consistency in the field.

Cellular resistance to anticancer therapies is a complex, multifaceted problem. Rather than continuing to chase the newest drug to remove molecular sub-states of cancer, this study implores the cancer research community to adopt a systems-level approach to the treatment of tumors, where clinicians consider multiple types of variability (including, but not limited to, genetic). Looking forward, one can envision a future cancer treatment regimen involving genetic profiling of a tumor to identify dominant genetic states, followed by characterization of the associated epigenetic landscapes using

single-cell experimentation and computational modeling[88,139,198,199] and then large-scale *in vitro* and *in silico* drug screens to devise personalized treatments for patients that can be tested *in vivo* before being administered clinically (FIG. 44). By leveraging state-of-the-art technologies and currently available drugs to tackle tumor heterogeneity at the genetic and non-genetic levels, this approach may finally give us a leg up in the longstanding War on Cancer.

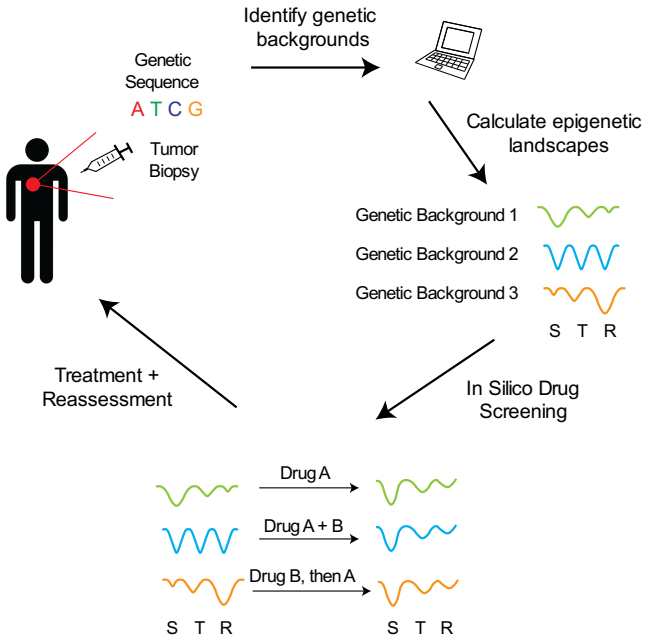

Figure 44 | *Future strategies for treatment targeting the epigenetic landscape.*
Upon identifying and biopsying a tumor, genomic profiles will be taken. From these profiles, patient genetic data will be assigned to one of many relevant genetic backgrounds. *In vitro* data would provide information on the types of epigenetic states occurring in each tumor genetic background. To target tumor genetic backgrounds, batteries of *in silico* drug screens would be performed to identify the ideal treatment method (i.e. single agent, combination, sequential) in order to convert each epigenetic landscape into one where cells are in dying drugged cell states. Patients will then be provided personalized treatments (at the level of genetic backgrounds) according to the ideal method, and the cycle continues for as long as tumors remain.

REFERENCES

1. Fitzmaurice C, Dicker D, Pain A, et al. The Global Burden of Cancer 2013. *JAMA Oncol*. 1(4):505-527. doi:10.1001/jamaoncol.2015.0735
2. DeVita VT, Chu E. A history of cancer chemotherapy. *Cancer Res*. 2008;68(21):8643-8653. doi:10.1158/0008-5472.CAN-07-6611
3. Torti D, Trusolino L. Oncogene addiction as a foundational rationale for targeted anti-cancer therapy: Promises and perils. *EMBO Mol Med*. 2011;3(11):623-636. doi:10.1002/emmm.201100176
4. Sharma S V, Settleman J. Oncogene addiction : setting the stage for molecularly targeted cancer therapy. 2007:3214-3231. doi:10.1101/gad.1609907.
5. Lynch TJ, Bell DW, Sordella R, et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *N Engl J Med*. 2004;350(21):2129-2139. doi:10.1056/NEJMoa040938
6. Bollag G, Tsai J, Zhang J, et al. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov*. 2012;11(11):873-886. doi:10.1038/nrd3847
7. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026
8. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol*. 2011;12(8):735-742. doi:10.1016/S1470-2045(11)70184-X
9. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011;364(26):2507-2516. doi:10.1056/NEJMoa1103782
10. Geyer CE, Forster J, Lindquist D, Chan S RC, Al E. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med*. 2006;355:2733-2743.
11. Sosman JA, Kim KB, Schuchter L, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med*. 2012;366(8):707-714. doi:10.1056/NEJMoa1112302
12. Ramirez M, Rajaram S, Steininger RJ, et al. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun*. 2016;7:10690. doi:10.1038/ncomms10690
13. Hata AN, Niederst MJ, Archibald HL, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med*. 2016;(August 2015). doi:10.1038/nm.4040
14. Nowell PC. The clonal evolution of tumor cell populations. *Science (80-)*. 1976;194(4260):23-28. doi:10.1126/science.959840
15. Soria J-C, Ohe Y, Vansteenkiste J, et al. Osimertinib in Untreated *EGFR* -Mutated Advanced Non–Small-Cell Lung Cancer. *N Engl J Med*. 2017:NEJMoa1713137. doi:10.1056/NEJMoa1713137
16. Subbiah V, Baik C, Kirkwood JM. Clinical Development of BRAF plus MEK Inhibitor Combinations. *Trends in Cancer*. 2020;6(9):797-810. doi:10.1016/j.trecan.2020.05.009

17. Long G V., Stroyakovskiy D, Gogas H, et al. Combined BRAF and MEK Inhibition versus BRAF Inhibition Alone in Melanoma. *N Engl J Med*. 2014;371(20):1877-1888. doi:10.1056/nejmoa1406037
18. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013;501(7467):328-337. doi:10.1038/nature12624
19. Tang J, Salama R, Gadgeel SM, Sarkar FH, Ahmad A. Erlotinib resistance in lung cancer: Current progress and future perspectives. *Front Pharmacol*. 2013;4 FEB(February):1-9. doi:10.3389/fphar.2013.00015
20. Luebker SA, Koepsell SA. Diverse mechanisms of BRAF inhibitor resistance in melanoma identified in clinical and preclinical studies. *Front Oncol*. 2019;9(MAR):1-8. doi:10.3389/fonc.2019.00268
21. Sharma S V., Lee DY, Li B, et al. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell*. 2010;141(1):69-80. doi:10.1016/j.cell.2010.02.027
22. Xu J, Wang J, Zhang S. Mechanisms of resistance to irreversible epidermal growth factor receptor tyrosine kinase inhibitors and therapeutic strategies in non-small cell lung cancer. *Oncotarget*. 2017;8(52):90557-90578. doi:10.18632/oncotarget.21164
23. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012;12(5):323-334. doi:10.1038/nrc3261
24. Marusyk A, Polyak K. Tumor heterogeneity: Causes and consequences. *Biochim Biophys Acta - Rev Cancer*. 2010;1805(1):105-117. doi:10.1016/j.bbcan.2009.11.002
25. Pao W, Chmielecki J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nat Rev Cancer*. 2010;10(11):760-774. doi:10.1038/nrc2947
26. Ben-David U, Siranosian B, Ha G, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*. 2018;560(7718):325–330. doi:10.1038/s41586-018-0409-3
27. Brock A, Chang H, Huang S. Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. *Nat Rev Genet*. 2009;10(5):336-342. doi:10.1038/nrg2556
28. Niepel M, Spencer SL, Sorger PK. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Curr Opin Chem Biol*. 2009;13(5-6):556-561. doi:10.1016/j.cbpa.2009.09.015
29. Pisco AO, Huang S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: “What does not kill me strengthens me”. *Br J Cancer*. 2015;112(11):1725-1732. doi:10.1038/bjc.2015.146
30. Jia D, Jolly MK, Kulkarni P, Levine H. Phenotypic plasticity and cell fate decisions in cancer: Insights from dynamical systems theory. *Cancers (Basel)*. 2017;9(7):1-19. doi:10.3390/cancers9070070
31. Bhang HC, Ruddy DA, Krishnamurthy Radhakrishna V, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med*. 2015;21(5):440-448. doi:10.1038/nm.3841
32. Al’Khafaji AM, Deatherage D, Brock A. Control of Lineage-Specific Gene Expression by Functionalized gRNA Barcodes. *ACS Synth Biol*. 2018;7(10):2468-2474. doi:10.1021/acssynbio.8b00105
33. Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE. Simultaneous single-cell profiling of

lineage and identity in direct reprogramming. *Nature*. 2018. doi:10.1038/s41586-018-0744-4
34. Mojtahedi M, Skupin A, Zhou J, et al. Cell Fate Decision as High-Dimensional Critical State Transition. 2016:1-28. doi:10.1371/journal.pbio.2000640
35. Su Y, Bintz M, Yang Y, et al. Phenotypic heterogeneity and evolution of melanoma cells associated with targeted therapy resistance. Tanay A, ed. *PLOS Comput Biol*. 2019;15(6):e1007034. doi:10.1371/journal.pcbi.1007034
36. Paudel BB, Harris LA, Hardeman KN, et al. A Nonquiescent "Idling" Population State in Drug-Treated, BRAF-Mutated Melanoma. *Biophys J*. 2018;114(6):1499-1511. doi:10.1016/j.bpj.2018.01.016
37. Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. *Nature*. 2009;461(7269):90-93. doi:10.1038/nature08504
38. Thattai M, Van Oudenaarden A. Stochastic gene expression in fluctuating environments. *Genetics*. 2004;167(1):523-530. doi:10.1534/genetics.167.1.523
39. Flusberg DA, Sorger PK. Surviving apoptosis : life – death signaling in single cells. 2015;25(8).
40. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial Persistence as a Phenotypic Switch. *Science (80-)*. 2010;305(2004):1622-1625. doi:10.1126/science.1099390
41. Balkwill FR, Capasso M, Hagemann T. The tumor microenvironment at a glance. *J Cell Sci*. 2012;125(23):5591-5596. doi:10.1242/jcs.116392
42. Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-)*. 2016;352(6282):189-196. doi:10.1126/science.aad0501
43. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. *Oncogene*. 2008;27(45):5904-5912. doi:10.1038/onc.2008.271
44. O'Connell MP, Marchbank K, Webster MR, et al. Hypoxia induces phenotypic plasticity and therapy resistance in melanoma via the tyrosine kinase receptors ROR1 and ROR2. *Cancer Discov*. 2013;3(12):1378-1393. doi:10.1158/2159-8290.CD-13-0005
45. Hillen F, Griffioen AW. Tumour vascularization: Sprouting angiogenesis and beyond. *Cancer Metastasis Rev*. 2007;26(3-4):489-502. doi:10.1007/s10555-007-9094-7
46. Seidel JA, Otsuka A, Kabashima K. Anti-PD-1 and anti-CTLA-4 therapies in cancer: Mechanisms of action, efficacy, and limitations. *Front Oncol*. 2018;8(MAR):1-14. doi:10.3389/fonc.2018.00086
47. Rini BI. Vascular endothelial growth factor-targeted therapy in renal cell carcinoma: Current status and future directions. *Clin Cancer Res*. 2007;13(4):1098-1106. doi:10.1158/1078-0432.CCR-06-1989
48. Lawrence MS, Sougnez C, Lichtenstein L, et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517(7536):576-582. doi:10.1038/nature14129
49. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-421. doi:10.1038/nature12477
50. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-313. doi:10.1038/nature10762
51. Burrell R a, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic

heterogeneity in cancer evolution. *Nature*. 2013;501(7467):338-345. doi:10.1038/nature12625
52. Andor N, Graham TA, Jansen M, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med*. 2015;22(1):105-113. doi:10.1038/nm.3984
53. Hanahan D, Weinberg RA. Hallmarks of Cancer : The Next Generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
54. Hanahan D, Weinberg RA, Francisco S. The Hallmarks of Cancer. *Cell*. 2000;100:57-70.
55. De Bruin EC, McGranahan N, Mitter R, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science (80-)*. 2014;346(6206):251-256. doi:10.1126/science.1253462
56. Campbell PJ, Yachida S, Mudie LJ, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*. 2010;467(7319):1109-1113. doi:10.1038/nature09460
57. Russo A, Franchina T, Rita Ricciardi GR, et al. A decade of EGFR inhibition in EGFR-mutated non small cell lung cancer (NSCLC): Old successes and future perspectives. *Oncotarget*. 2015;6(29):26814-26825. doi:10.18632/oncotarget.4254
58. Hinohara K, Polyak K. Intratumoral Heterogeneity: More Than Just Mutations. *Trends Cell Biol*. 2019;29(7):569-579. doi:10.1016/j.tcb.2019.03.003
59. Letai A. Functional precision cancer medicine-moving beyond pure genomics. *Nat Med*. 2017;23(9):1028-1035. doi:10.1038/nm.4389
60. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. *Development*. 2009;136(23):3853-3862. doi:10.1242/dev.035139
61. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*. 2008;135(2):216-226. doi:10.1016/j.cell.2008.09.050
62. Samoilov MS, Price G, Arkin AP. From fluctuations to phenotypes: the physiology of noise. *Sci STKE*. 2006;2006(366):1-10. doi:10.1126/stke.3662006re17
63. Raser JM, O'Shea EK. Noise in Gene Expression: Origins, Consequences, and Control. *Science (80-)*. 2005;309(5743):2010-2013. doi:10.1126/science.1105891
64. Ceccarelli M, Barthel FP, Malta TM, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell*. 2016;164(3):550-563. doi:10.1016/j.cell.2015.12.028
65. Sanchez A, Choubey S, Kondev J. Regulation of noise in gene expression. *Annu Rev Biophys*. 2013;42:469-491. doi:10.1146/annurev-biophys-083012-130401
66. Thomas P, Terradot G, Danos V, Weiße AY. Sources, propagation and consequences of stochasticity in cellular growth. *Nat Commun*. 2018;9(1):1-11. doi:10.1038/s41467-018-06912-9
67. Huh D, Paulsson J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet*. 2011;43(2):95-100. doi:10.1038/ng.729
68. Risom T, Langer EM, Chapman MP, et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-05729-w
69. Hangauer MJ, Viswanathan VS, Ryan MJ, et al. Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature*. 2017;551(7679):247-250.

doi:10.1038/nature24297
70. Das Thakur M, Salangsang F, Landman AS, et al. Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance. *Nature*. 2013;494(7436):251-255. doi:10.1038/nature11814
71. Hugo W, Shi H, Sun L, et al. Non-genomic and Immune Evolution of Melanoma Acquiring MAPKi Resistance. *Cell*. 2015;162(6):1271-1285. doi:10.1016/j.cell.2015.07.061
72. Song C, Piva M, Sun L, et al. Recurrent tumor cell–intrinsic and –extrinsic alterations during mapki-induced melanoma regression and early adaptation. *Cancer Discov*. 2017;7(11):1248-1265. doi:10.1158/2159-8290.CD-17-0401
73. Jackson AL, Loeb LA. The mutation rate and cancer. *Genetics*. 1998;148(4):1483-1490. doi:10.1073/pnas.93.25.14800
74. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214-218. doi:10.1038/nature12213
75. Sasai M, Kawabata Y, Makishi K, Itoh K, Terada TP. Time Scales in Epigenetic Dynamics and Phenotypic Heterogeneity of Embryonic Stem Cells. *PLoS Comput Biol*. 2013;9(12). doi:10.1371/journal.pcbi.1003380
76. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science (80-)*. 2002;297(5584):1183-1186. doi:10.1126/science.1070919
77. 't Hooft G, Vandoren S, 't Hooft S. *Time in Powers of Ten : Natural Phenomena and Their Timescales.* World Scientific Publishing Company; 2014.
78. Gunawardena J. Time-scale separation - Michaelis and Menten's old idea, still bearing fruit. *FEBS J*. 2014;281(2):473-488. doi:10.1111/febs.12532
79. Weitzel JN, Blazer KR, MacDonald DJ, Culver JO, Offit K. Genetics, genomics, and cancer risk assessment: State of the Art and Future Directions in the Era of Personalized Medicine. *CA Cancer J Clin*. 2011;61(5):327-359. doi:10.3322/caac.20128
80. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet*. 2003;33(3S):238-244. doi:10.1038/ng1107
81. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell*. 2010;141(2):210-217. doi:10.1016/j.cell.2010.03.032
82. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):1-14. doi:10.1186/s13059-016-0974-4
83. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 2012;6(2):80-92. doi:10.4161/fly.19695
84. Waddington CH. Canalization of development and the inheritance of acquired characters. *Nature*. 1942;150(3811):563-565. doi:10.1038/150563a0
85. Huang S. Genetic and non-genetic instability in tumor progression: Link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev*. 2013;32(3-4):423-448. doi:10.1007/s10555-013-9435-7
86. Kar S, Baumann WT, Paul MR, Tyson JJ. Exploring the roles of noise in the eukaryotic cell cycle. *Proc Natl Acad Sci U S A*. 2009;106(16):6471-6476. doi:10.1073/pnas.0810034106
87. Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet*. 2016;17(11):693-703. doi:10.1038/nrg.2016.98

88. Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology? *BioEssays*. 2012;34(2):149-157. doi:10.1002/bies.201100031
89. Zhou JX, Pisco AO, Qian H, Huang S. Nonequilibrium population dynamics of phenotype conversion of cancer cells. *PLoS One*. 2014;9(12):1-19. doi:10.1371/journal.pone.0110714
90. Britton N. *Essential Mathematical Biology*. London: Springer-Verlag; 2005.
91. Kaufman PD, Rando OJ. Chromatin as a potential carrier of heritable information. *Curr Opin Cell Biol*. 2010;22(3):284-290. doi:10.1016/j.ceb.2010.02.002
92. Huang S, Ernberg I, Kauffman S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol*. 2009;20(7):869-876. doi:10.1016/j.semcdb.2009.07.003
93. Shaffer SM, Emert BL, Reyes Hueros RA, et al. Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. *Cell*. 2020;182(4):947-959.e17. doi:10.1016/j.cell.2020.07.003
94. Bošković A, Rando OJ. Transgenerational epigenetic inheritance. *Annu Rev Genet*. 2018;52:21-41. doi:10.1146/annurev-genet-120417-031404
95. Ptashne M. On the use of the word "epigenetic." *Curr Biol*. 2007. doi:10.1016/j.cub.2007.02.030
96. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. *Genes Dev*. 2009;23(7):781-783. doi:10.1101/gad.1787609
97. Muzzey D, Gomez-Uribe CA, Mettetal JT, van Oudenaarden A. A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation. *Cell*. 2009;138(1):160-171. doi:10.1016/j.cell.2009.04.047
98. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, Oudenaarden A Van. Systematic Identification of Signal-Activated Stochastic Gene Regulation. *Science (80-)*. 2013;339(6119):584-587. doi:10.1126/science.1231456
99. Gupta PBB, Fillmore CMM, Jiang G, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*. 2011;146(4):633-644. doi:10.1016/j.cell.2011.07.026
100. Magklara A, Lomvardas S. Stochastic gene expression in mammals: Lessons from olfaction. *Trends Cell Biol*. 2013;23(9):449-456. doi:10.1016/j.tcb.2013.04.005
101. Gillespie DT. Stochastic Simulation of Chemical Kinetics. *Annu Rev Phys Chem*. 2009:1-23. doi:10.1146/annurev.physchem.58.032806.104637
102. Hilfinger A, Paulsson J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A*. 2011;108(29):12167-12172. doi:10.1073/pnas.1018832108
103. Friedman N, Cai L, Xie XS. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett*. 2006;97(16):1-4. doi:10.1103/PhysRevLett.97.168302
104. McAdams HH, Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet*. 1999;15(2):65-69. doi:10.1016/s0168-9525(98)01659-x
105. Sigal A, Milo R, Cohen A, et al. Variability and memory of protein levels in human cells. *Nature*. 2006;444(7119):643-646. doi:10.1038/nature05316

106. El-Samad H, Khammash M. Regulated degradation is a mechanism for suppressing stochastic fluctuations in gene regulatory networks. *Biophys J*. 2006;90(10):3749-3761. doi:10.1529/biophysj.105.060491
107. Alarcón T, Page KM. Stochastic models of receptor oligomerization by bivalent ligand. *J R Soc Interface*. 2006;3(9):545-559. doi:10.1098/rsif.2006.0116
108. Ladbury JE, Arold ST. Noise in cellular signaling pathways: Causes and effects. *Trends Biochem Sci*. 2012;37(5):173-178. doi:10.1016/j.tibs.2012.01.001
109. Huh D, Paulsson J. Random partitioning of molecules at cell division. *Proc Natl Acad Sci U S A*. 2011;108(36):15004-15009. doi:10.1073/pnas.1013171108
110. Hardeman KN, Peng C, Paudel BB, et al. Dependence on glycolysis sensitizes BRAF-mutated melanomas for increased response to targeted BRAF inhibition. *Sci Rep*. 2017;7:42604. doi:10.1038/srep42604
111. Charlebois DA, Hauser K, Marshall S, Balázsi G. Multiscale effects of heating and cooling on genes and gene networks. *Proc Natl Acad Sci U S A*. 2018;115(45):E10797-E10806. doi:10.1073/pnas.1810858115
112. Webb BA, Chimenti M, Jacobson MP, Barber DL. Dysregulated pH: A perfect storm for cancer progression. *Nat Rev Cancer*. 2011;11(9):671-677. doi:10.1038/nrc3110
113. Lee CF, Brangwynne CP, Gharakhani J, Hyman AA, Jülicher F. Spatial organization of the cell cytoplasm by position-dependent phase separation. *Phys Rev Lett*. 2013;111(8):1-5. doi:10.1103/PhysRevLett.111.088101
114. Mcadams HH, Arkin A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*. 1997;94(3):814-819. doi:10.1073/pnas.94.3.814
115. Losick R, Desplan C. Stochasticity and cell fate. *Science (80-)*. 2008;320(5872):65-68. doi:10.1126/science.1147888
116. Rao C V., Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. *Nature*. 2002;420(6912):231-237. doi:10.1038/nature01258
117. Tyson DR, Garbett SP, Frick PL, Quaranta V. Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data. *Nat Methods*. 2012;9(9):923-928. doi:10.1038/nmeth.2138
118. Harris LA, Frick PL, Garbett SP, et al. An unbiased metric of antiproliferative drug effect in vitro. *Nat Methods*. 2016;13(October 2015):1-6. doi:10.1038/nmeth.3852
119. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell*. 2019;24(1):65-78. doi:10.1016/j.stem.2018.11.011
120. McQuarrie DA. Stochastic approach to chemical kinetics. *J Appl Probab*. 1967;4(3):413-478. doi:10.2307/3212214
121. Gillespie DT. A rigorous derivation of the chemical master equation. *Phys A Stat Mech its Appl*. 1992;188(1-3):404-425. doi:10.1016/0378-4371(92)90283-V
122. Li H, Cao Y, Petzold LR, Gillespie DT. Algorithms and software for stochastic simulation of biochemical reacting systems. In: *Biotechnology Progress*. Vol 24. NIH Public Access; 2008:56-61. doi:10.1021/bp070255h
123. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091-1098. doi:10.1038/ng.3367

124. Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. *Nat Rev Genet*. 2015;16(4):213-223. doi:10.1038/nrg3890
125. Marks DC, Belov L, Davey MW, Davey RA, Kidman AD. The MTT cell viability assay for cytotoxicity testing in multidrug-resistant human leukemic cells. *Leuk Res*. 1992;16(12):1165-1173. doi:10.1016/0145-2126(92)90114-M
126. Franken NAP, Rodermond HM, Stap J, Haveman J, van Bree C. Clonogenic assay of cells in vitro. *Nat Protoc*. 2006;1(5):2315-2319. doi:10.1038/nprot.2006.339
127. Frick PL, Paudel BB, Tyson DR, Quaranta V. Quantifying heterogeneity and dynamics of clonal fitness in response to perturbation. *J Cell Physiol*. 2015;230(7):1403-1412. doi:10.1002/jcp.24888
128. Koizumi F, Shimoyama T, Taguchi F, Saijo N, Nishio K. Establishment of a human non-small cell lung cancer cell line resistant to gefitinib. *Int J Cancer*. 2005;116(1):36-44. doi:10.1002/ijc.20985
129. Tan MC, Quinlan MP, Singh A, et al. Reduced erlotinib sensitivity of epidermal growth factor receptor-mutant non-small cell lung cancer following cisplatin exposure: A cell culture model of second-line erlotinib treatment. *Clin Cancer Res*. 2008;14(21):6867-6876. doi:10.1158/1078-0432.CCR-08-0093
130. Jia P, Jin H, Meador CB, et al. Next-generation sequencing of paired tyrosine kinase inhibitor-sensitive and -resistant EGFR mutant lung cancer cell lines identifies spectrum of DNA changes associated with drug resistance. *Genome Res*. 2013;23(9):1434-1445. doi:10.1101/gr.152322.112
131. Roesch A. Tumor heterogeneity and plasticity as elusive drivers for resistance to MAPK pathway inhibition in melanoma. *Oncogene*. 2014;34(April):1-7. doi:10.1038/onc.2014.249
132. Su Y, Ko ME, Cheng H, et al. Multi-omic single-cell snapshots reveal multiple independent trajectories to drug tolerance in a melanoma cell line. *Nat Commun*. 2020;11(1):2345. doi:10.1038/s41467-020-15956-9
133. Paudel BB, Lewis JE, Hardeman KN, et al. An Integrative Gene Expression and Mathematical Flux Balance Analysis Identifies Targetable Redox Vulnerabilities in Melanoma Cells. *Cancer Res*. 2020;80(20):4565-4577. doi:10.1158/0008-5472.can-19-3588
134. Jia D, Paudel BB, Hayford CE, et al. Drug-Tolerant Idling Melanoma Cells Exhibit Theory-Predicted Metabolic Low-Low Phenotype. *Front Oncol*. 2020;10(August). doi:10.3389/fonc.2020.01426
135. Pisco AO, Brock A, Zhou J, et al. Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat Commun*. 2013;4:2467. doi:10.1038/ncomms3467
136. Shaffer SM, Dunagin MC, Torborg SR, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. 2017. doi:10.1038/nature22794
137. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys*. 1976;22(4):403-434. doi:10.1016/0021-9991(76)90041-3
138. Landry JJM, Pyl PT, Rausch T, et al. The Genomic and Transcriptomic Landscape of a HeLa Cell Line. *G3: Genes|Genomes|Genetics*. 2013;3(8):1213-1224.

doi:10.1534/g3.113.005777
139. Wang J, Zhang K, Xu L, Wang E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci U S A*. 2011;108(20):8257-8262. doi:10.1073/pnas.1017017108
140. Govindan R, Ding L, Griffith M, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell*. 2012;150(6):1121-1134. doi:10.1016/j.cell.2012.08.024
141. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*. 2018. http://arxiv.org/abs/1802.03426.
142. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2019;37(1):38-47. doi:10.1038/nbt.4314
143. Carbon S, Douglass E, Dunn N, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330-D338. doi:10.1093/nar/gky1055
144. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556
145. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. *Bioinformatics*. 2007;23(10):1274-1281. doi:10.1093/bioinformatics/btm087
146. Meyer CT, Wooten DJ, Paudel BB, et al. Quantifying Drug Combination Synergy along Potency and Efficacy Axes. *Cell Syst*. 2019;8(2):97-108.e16. doi:10.1016/j.cels.2019.01.003
147. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013.
148. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158. doi:10.1093/bioinformatics/btr330
149. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour*. 2017;17(1):44-53. doi:10.1111/1755-0998.12549
150. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998;8(3):175-185. doi:10.1101/gr.8.3.175
151. Martincorena I, Raine KM, Gerstung M, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. 2017;171(5):1029-1041.e21. doi:10.1016/j.cell.2017.09.042
152. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:1-12. doi:10.1038/ncomms14049
153. Stoeckius M, Zheng S, Houck-Loomis B, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol*. 2018;19(1):1-12. doi:10.1186/s13059-018-1603-1
154. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411-420. doi:10.1038/nbt.4096
155. DeTomaso D, Jones MG, Subramaniam M, Ashuach T, Ye CJ, Yosef N. Functional interpretation of single cell similarity maps. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-12235-0
156. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
157. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for

assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656
158. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1-21. doi:10.1186/s13059-014-0550-8
159. Gerstung M, Pellagatti A, Malcovati L, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun*. 2015;6:1-11. doi:10.1038/ncomms6901
160. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14(1):128. doi:10.1186/1471-2105-14-128
161. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. 2010;26(7):976-978. doi:10.1093/bioinformatics/btq064
162. Lopez CF, Muhlich JL, Bachman JA, Sorger PK. Programming biological models in Python using PySB. *Mol Syst Biol*. 2014;9(1):646-646. doi:10.1038/msb.2013.1
163. Harris LA, Hogg JS, Tapia JJ, et al. BioNetGen 2.2: Advances in rule-based modeling. *Bioinformatics*. 2016;32(21):3366-3368. doi:10.1093/bioinformatics/btw469
164. Sottoriva A, Kang H, Ma Z, et al. A Big Bang model of human colorectal tumor growth. *Nat Genet*. 2015;47(3):209-216. doi:10.1038/ng.3214
165. Lynch M. Phylogenetic divergence of cell biological features. 2018:1-17. doi:https://doi.org/10.7554/eLife.34820
166. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*. 2007;3(10):1871-1878. doi:10.1371/journal.pcbi.0030189
167. Eydgahi H, Chen WW, Muhlich JL, Vitkup D, Tsitsiklis JN, Sorger PK. Properties of cell death models calibrated and compared using Bayesian approaches. *Mol Syst Biol*. 2013;9(644):644. doi:10.1038/msb.2012.69
168. Huang S, Kauffman S. How to escape the cancer attractor: Rationale and limitations of multi-target drugs. *Semin Cancer Biol*. 2013;23(4):270-278. doi:10.1016/j.semcancer.2013.06.003
169. Lock JT, Parker I, Smith IF. A comparison of fluorescent Ca2+ indicators for imaging local Ca2+ signals in cultured cells. *Cell Calcium*. 2015;58(6):638-648. doi:10.1016/j.ceca.2015.10.003
170. Marciel MP, Khadka VS, Deng Y, et al. Correction: Selenoprotein K deficiency inhibits melanoma by reducing calcium flux required for tumor growth and metastasis [Oncotarget, 9, (2018) (13407-13422)] DOI: 10.18632/oncotarget.24388. *Oncotarget*. 2018;9(56):30937. doi:10.18632/oncotarget.25857
171. Xie Y, Hou W, Song X, et al. Ferroptosis: Process and function. *Cell Death Differ*. 2016;23(3):369-379. doi:10.1038/cdd.2015.158
172. Abdalkader M, Lampinen R, Kanninen KM, Malm TM, Liddell JR. Targeting Nrf2 to suppress ferroptosis and mitochondrial dysfunction in neurodegeneration. *Front Neurosci*. 2018;12(JUL):1-9. doi:10.3389/fnins.2018.00466
173. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory

requirements. *Nat Methods*. 2015;12(4):357-360. doi:10.1038/nmeth.3317
174. Harris LA, Beik S, Ozawa PMM, Jimenez L, Weaver AM. Modeling heterogeneous tumor growth dynamics and cell-cell interactions at single-cell and cell-population resolution. *Curr Opin Syst Biol*. In Review.
175. Weinreb C, Briggs JA, Wagner DE, Collins ZM, Klein AM, Megason SG. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science (80-)*. 2018;360(6392):981-987. doi:10.1126/science.aar4362
176. Roesch A, Fukunaga-kalabis M, Schmidt EC, et al. A Temporarily Distinct Subpopulation of Slow-Cycling Melanoma Cells Is Required for Continuous Tumor Growth. *Cell*. 2010;141:583-594. doi:10.1016/j.cell.2010.04.020
177. Kreso A, O'Brien CA, van Galen P, et al. Variable Clonal Repopulation Dynamics Influence Chemotherapy Response in Colorectal Cancer. *Science (80-)*. 2013;339(February):543-548.
178. Yaffe MB. Why geneticists stole cancer research even though cancer is primarily a signaling disease. *Sci Signal*. 2019;12(565):eaaw3483. doi:10.1126/scisignal.aaw3483
179. Cagan R, Meyer P. Rethinking cancer: Current challenges and opportunities in cancer research. *Dis Model Mech*. 2017;10(4):349-352. doi:10.1242/dmm.030007
180. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*. 2018;15(2):81-94. doi:10.1038/nrclinonc.2017.166
181. Kaiser J. The cancer stem cell gamble. *Science (80-)*. 2015;347(6219):226-229. doi:10.1126/science.347.6219.226
182. Annett S, Robson T. Targeting cancer stem cells in the clinic: Current status and perspectives. *Pharmacol Ther*. 2018;187:13-30. doi:10.1016/j.pharmthera.2018.02.001
183. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: An emerging axis of evil in the war on cancer. *Oncogene*. 2010;29(34):4741-4751. doi:10.1038/onc.2010.215
184. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: Acquisition of malignant and stem cell traits. *Nat Rev Cancer*. 2009;9(4):265-273. doi:10.1038/nrc2620
185. Aranda-Anzaldo A, Dent MAR. Landscaping the epigenetic landscape of cancer. *Prog Biophys Mol Biol*. 2018;140:155-174. doi:10.1016/j.pbiomolbio.2018.06.005
186. Blatter S, Rottenberg S. Minimal residual disease in cancer therapy - Small things make all the difference. *Drug Resist Updat*. 2015;21-22:1-10. doi:10.1016/j.drup.2015.08.003
187. Sale MJ, Balmanno K, Saxena J, et al. MEK1/2 inhibitor withdrawal reverses acquired resistance driven by BRAF V600E amplification whereas KRAS G13D amplification promotes EMT-chemoresistance. *Nat Commun*. 2019;10(1):1-22. doi:10.1038/s41467-019-09438-w
188. Kaiser C a., Schekman R. Distinct sets of SEC genes govern transport vesicle formation and fusion early in the secretory pathway. *Cell*. 1990;61(4):723-733. doi:10.1016/0092-8674(90)90483-U
189. Zhao B, Sedlak JC, Srinivas R, et al. Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution. *Cell*. 2016;165(1):234-246. doi:10.1016/j.cell.2016.01.045
190. Imamovic L, Sommer MOA. Use of Collateral Sensitivity Networks to Design Drug Cycling Protocols That Avoid Resistance Development. *Sci Transl Med*. 2013;5(204):204ra132-204ra132. doi:10.1126/scitranslmed.3006609

191. Imamovic L, Ellabaan MMH, Dantas Machado AM, et al. Drug-Driven Phenotypic Convergence Supports Rational Treatment Strategies of Chronic Infections. *Cell*. 2018;172(1-2):121-134.e14. doi:10.1016/j.cell.2017.12.012
192. Basanta D, Gatenby RA, Anderson ARA. Exploiting evolution to treat drug resistance: Combination therapy and the double bind. *Mol Pharm*. 2012;9(4):914-921. doi:10.1021/mp200458e
193. Lee MJ, Ye AS, Gardino AK, et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*. 2012;149(4):780-794. doi:10.1016/j.cell.2012.03.031
194. Cayer DM, Nazor KL, Schork NJ. Mission critical: The need for proteomics in the era of next-generation sequencing and precision medicine. *Hum Mol Genet*. 2016;25(R2):R182-R189. doi:10.1093/hmg/ddw214
195. Petricoin EF, Liotta LA. Clinical Applications of Proteomics. *J Nutr*. 2003;133(7):2476S-2484S. doi:10.1093/jn/133.7.2476S
196. Mumby M, Brekken D. Phosphoproteomics: New insights into cellular signaling. *Genome Biol*. 2005;6(9). doi:10.1186/gb-2005-6-9-230
197. Cravatt BF, Wright AT, Kozarich JW. Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. *Annu Rev Biochem*. 2008;77:383-414. doi:10.1146/annurev.biochem.75.101304.124125
198. Chylek LA, Harris LA, Faeder JR, Hlavacek WS. Modeling for (physical) biologists: an introduction to the rule-based approach. *Phys Biol*. 2015;12(4):045007. doi:10.1088/1478-3975/12/4/045007
199. Wu F, Su RQ, Lai YC, Wang X. Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. Shou W, ed. *Elife*. 2017;6:e23702. doi:10.7554/eLife.23702